The Dissertation Committee for Xing Lan Certifies that this is the approved version of the following Dissertation:

Two Essays on Mobile Marketing

Committee:

Jason Duan, Co-Supervisor

Vijay Mahajan, Co-Supervisor

Tianshu Sun

Raghunath Rao

Two Essays on Mobile Marketing

by

Xing Lan

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2023

Acknowledgements

I am incredibly indebted to both of my advisors, Dr. Jason Duan and Dr. Vijay Mahajan, for their wisdom and guidance all throughout my PhD program. I not only benefited immensely from their research advice, but also learned a great deal from them about how to be a better person. I'm deeply grateful for Dr. Raghunath Rao who led me into the field of quantitative marketing. I also cannot thank Dr. Tianshu Sun enough for his continual support that kept me looking forward.

I would like to thank all my friends, especially Wen Wang, Haris Krijestorac, Tushmit Hasan, Skylar DeTure, Chandra Srivastava, Zhuping Liu, Gunes Biliciler, Xinying Hao, Jianqing Zheng, Max Alberhasky, Xiao Han, Yixuan Liu, Qifei Zhu, Shuo Liu, Xinchen Gu, and Xiang Liu. Your friendships have been valuable experiences for me throughout different stages of my PhD journey, and your presence in my life has helped me see the better version of myself that I could become, even only in hindsight.

I thank my parents for their unconditional love and support all throughout my life.

Abstract

Two Essays on Mobile Marketing

Xing Lan, Ph.D.

The University of Texas at Austin, 2023

Supervisors: Jason Duan, Vijay Mahajan

Consumers' dependence on mobile technology for their daily personal and business dealings has created many new business touchpoints, which in turn bring about novel opportunities for marketers to understand consumers' behavior and identify ways in which more effective marketing strategies can be utilized. In two essays, this dissertation explores two domains in mobile marketing where new insights can be gained.

In the first essay, we leverage data collected from IoT sensors and combine with scanner transaction data to construct and estimate the entire offline conversion funnel for brick-and-mortar retail stores. We uncover the multi-stage effect of the marketing mix, and demonstrate the value of IoT data in enabling customized marketing strategies.

In the second essay, we examine what makes mobile marketing communications effective by studying the design elements of app push notifications. We construct a theory-based framework and separate out the treatment effects of proposed design elements using a causal inference machine learning method. Our findings provide guidance for marketing managers on design choice that improves the effectiveness of future push messages.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

As usage of mobile devices becomes ubiquitous in consumers' day-to-day lives, businesses also gain access to more touchpoints with consumers by way of mobile contact and communications. The interactions between businesses and consumers that are made available by these mobile touchpoints not only generate new information that could provide valuable and previously unknown business insights, but also enable businesses to evaluate and improve the effectiveness of their marketing strategies.

Although consumers most often utilize their mobile devices for long-range virtual communications, these portable gadgets also follow their owners' movement in the real world, and can nonetheless leave behind digital footprints that reveal the behavioral patterns of consumers in a physical environment. With the advancement of IoT technologies, detecting and capturing consumers' location information can be achieved through the tracking of their mobile devices. This kind of locational data is highly desirable for brick-and-mortar retail stores, because it contains information that originates beyond the physical boundaries of the stores themselves and provides the opportunity to uncover consumers' behaviors prior to a store visit. Consequently, an offline business can map out a more complete conversion funnel which includes stages pertaining to consumers' behavior prior to a store visit. Combined with traditional scanner data, we propose a holistic framework to model consumers' entire offline journey in the context of brick-and-mortar retail stores, and identify a new margin – visit effect, as a significant contributing factor to the success of marketing strategies. We also demonstrate the business value of the IoT technology by showing the improvement that it brings to the design of optimal marketing strategies.

Imagine you are the manager of a small convenience store in a major metropolitan area, and you are in charge of designing marketing strategies at the store. The convenience store has a considerable amount of foot traffic, but the sales performance is not as ideal. Therefore, you have decided to run a price promotion campaign across multiple product categories. You place promotion posters outside the store and also have on-sale tags posted on the shelves. Luckily, you see an increase in sales across the promoted product categories during the promotion period, but now you are facing a new question: is the increase in sales driven by a higher number of store visits from passersby ("visit effect"), or is it driven by an increase in purchase probability from consumers who have visited the store ("purchase effect")? Or maybe it's only by sheer luck of having more foot traffic during the same period? Differentiating the cause is crucial, as each cause has very different implications for how to design your business strategy next: more in-store up-selling if visit effect dominates, more out-of-store events and promotions to attract visits if purchase effect dominates, and re-designing the whole promotion campaign if it is due to luck. You pull out the store scanner data to try to solve the puzzle, but you are not able to gain any real insights by merely looking at these transaction records: neither a consumer's pass-by nor her store visit is recorded; moreover, no common unique identifier exists to help you link a consumer's behaviors across different stages even if consumer pass-by and store visit information are available. You, as the store manager, are not able to "see" through (even a part of) consumers' offline journey and can thus only throw the next dart in the dark.

The above example illustrates a general challenge faced by many offline businesses, including brick-and-mortar retailers as well as service providers (e.g. restaurants). The consumer's complete offline journey still largely remains a mystery to these businesses: neither consumers' actual offline behaviors nor the nuanced impacts of marketing strategies on their whole journey are well understood. It has long been a dream of offline businesses to observe and understand consumers' behaviors across multiple stages of the offline conversion funnel (pass-by, visit, purchase), so that they can better design their marketing strategies accordingly (e.g. price promotions, outdoor campaigns).

Fortunately, the Internet-of-Things (IoT) technology offers a solution. The power of the IoT technology originates not only from the new data that it captures, which consists of crucial behavioral information through the early stages of a consumer's offline journey, but also from the unique identification system that it creates. As illustrated in Figure 1.1 below, with the deployment of IoT technology (e.g. in a convenience store), the manager can now observe potential consumers both near and inside the store, and use this information to understand their pass-by behavior and store visit decisions. In addition, the manager now can not only link consumers’ behaviors across multiple stages using the unique mobile identifier, but also infer two key decisions of each individual consumer over time, i.e. whether to visit the store and whether to make a purchase. Just as clickstream data has helped marketers gain in-depth understanding of consumers’ path to purchase online (Moe 2003; Moe and Fader 2004; Montgomery et al. 2004, Li and Kannan 2014; Xu et al. 2014), the location data captured by IoT also provides the missing link that helps managers develop better working knowledge of the consumers’ offline conversion funnel.

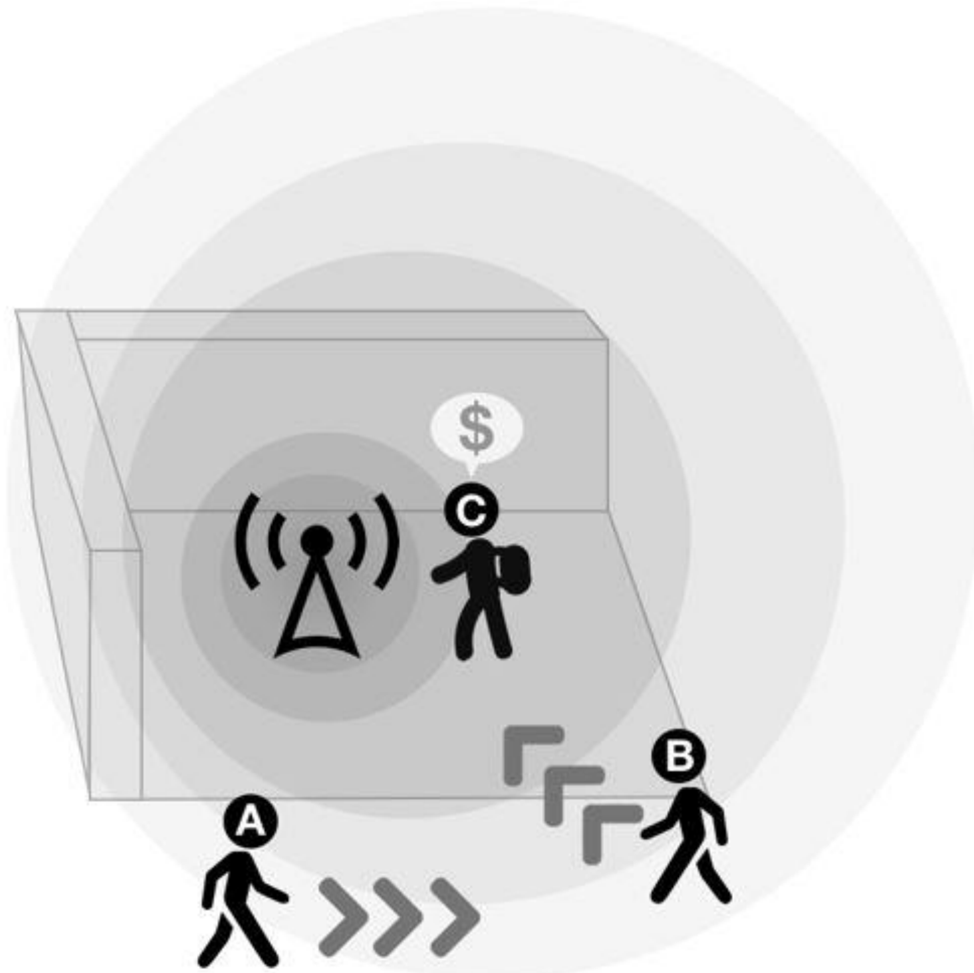

Figure 1.1: Illustration of IoT device used for Wi-Fi tracking

We believe the new data from the IoT technology complements traditional scanner transaction data in three fundamentally important ways. First, offline businesses can now observe a more complete set of potential consumers at the store (such as the passerby, Individual A, and a store visitor like Individual B in Figure 1.1), instead of only those who have chosen to make a purchase (Individual C, who is the only type of individuals captured by traditional scanner data). Second, the detailed signal strength information captured by the IoT sensor allows businesses to differentiate consumers' location status (in-store vs.

out-of-store), thus enabling inferences about their store visit decisions as well as in-store purchase decisions, thereby completing the construction of the entire offline conversion funnel. Third, the IoT sensor captures the unique mobile device ID associated with each consumer and makes it possible to observe repeated pass-bys as well as return visits from the same consumer over time, thus allowing businesses to exploit information on consumers' historical behavioral patterns (rather than only historical transaction records) to better understand each consumer.

Recognizing the large potential of said IoT technology, we aim to demonstrate how firms may 1) leverage the information captured by this novel technology to develop deeper insights into consumers' offline journey and 2) build upon this new knowledge to improve the design of marketing strategies. To achieve these goals, we collect and construct a large and proprietary dataset combining both IoT and scanner transaction data for over 840,000 consumers. We propose a three-stage model to capture consumers' decisions across the entire offline conversion funnel (pass-by, visit and purchase, Figure 2.1). Employing a Bayesian hierarchical modeling framework, we build our model to fully take advantage of the IoT data. We find that price promotion indeed has a strong and positive effect on consumers' likelihood of visiting the store (visit effect), in addition to its well-understood effect of increasing purchase when consumers do visit the store (purchase effect). More interestingly, we find the visit effect is immensely important vis-à-vis the purchase effect through simulation studies, which use the estimated coefficients in the store visit model and purchase incidence model and take into account unobserved heterogeneity across different stages of the conversion funnel.

Based on the model estimates, we also conduct simulations and find that whereas a model with IoT data can offer recommendations for price promotions tailored to each product category and store, the model without IoT data consistently fails to produce non-zero price promotion recommendations. Consequently, we demonstrate that implementing IoT technologies and leveraging their information in the design of marketing strategies can bring about significant business value for retailers operating in the offline setting.

In the digital realm, mobile marketing communications to consumers in the form of app push notifications have become a prevalent business practice. As business apps frequently employ this touchpoint to encourage user engagement, there emerges the need to better understand what kind of messages will perform well. This is because not only do mobile users often face a large volume of push notifications from different apps on a daily basis, but also that each push notification presents only a small opportunity for user engagement due to its limited display space and easy dismissal by the user. As a result, the design of the message plays a critical role in effecting a push notification's ability to prompt user engagement with the app. We construct a theory-based framework to investigate what design elements of app push notifications causally drive user engagement. Applying the causal forest machine learning method, we disentangle the treatment effects of multiple concurrent design elements using observational data. Our findings point to ways in which managers can improve the message design of push notifications to bring about better user engagement.

Retailers, merchants, and brands have long utilized mobile marketing communications to engage with their customers, because highly engaged customers are more likely to make purchases, promote the brand, and demonstrate loyalty.[1] With mobile devices becoming the primary platform for digital media consumption, mobile push notifications ("push") have turned into a common method for mobile applications ("apps")

[1] Source: https://www.qualtrics.com/experience-management/customer/customer-engagement/?vid=clarabridge_redirect

to connect with users. In the United States, the average mobile user spends 54.8% of their digital media time on mobile devices[2] and receives 46 app push notifications per day.[3]

Push notifications are an integral part of an app's engagement strategy, as effective push notifications can increase app engagement by prompting users to open the app. When a user opens the app in response to a push notification, they may engage in activities that provide business value to the app, such as browsing, clicking on outbound links, and making purchases. However, the user has the power to decide whether or not to open the push and engage with the app. Excessive push notifications can lead to users uninstalling the app altogether.[4] Therefore, it is important for marketers to understand how to design push notifications effectively to engage users without overwhelming them.

Many real-world examples exist and show that well-designed push notifications effectively lead to user engagement with an app. For instance, Figure 1.2 below shows a push notification that prompted a user to engage with the app due to the offer of "free pizza". It is unclear if the app in fact handed out free pizzas, but the jocular language of the push notification alone could have been effective at capturing the user's attention. This observation then leads us to question what factors can contribute to the effectiveness of push notifications and what are their individual impacts on the overall effectiveness of the push.

From a general point of view, app push notifications are one of numerous formats of mobile marketing communications (e.g. short message service, phone calls, email) that marketers can utilize to communicate marketing offers to their customers on mobile.

[2] Source: https://www.insiderintelligence.com/insights/mobile-users-smartphone-usage/

[3] Source: https://www.businessofapps.com/marketplace/push-notifications/research/push-notifications-statistics/

[4] Source: https://www.businessofapps.com/marketplace/push-notifications/research/push-notifications-statistics/

However, research on mobile marketing has mostly focused on studying the effectiveness of the marketing strategies delivered by these mobile marketing communications. In other words, the various methods of mobile marketing communications have been treated as a vehicle of delivery, and academic research has yet to pay more attention to understanding the message design aspect of these communication methods themselves.

Figure 1.2: Example of an engaging app push notification

We study the design aspect of mobile marketing communications by focusing on app push notifications. There are four main advantages that make app push notifications the ideal format for this research perspective. First, push notifications take little time to deliver to the recipients on their mobile devices, which means new message designs can be rapidly implemented and tested. Second, the audience reach of push notifications is sizeable compared to other forms of mobile marketing communications, because most popular mobile apps boast a large user base. Third, thanks to the on-screen display feature, users are unlikely to miss a push notification once it’s delivered, and so the time span between users’ receipt of a push and their taking an action in response to it is usually short; this makes it possible for marketers to measure the performance of a push in a timely manner. Fourth, users receiving a push can be characterized in terms of their observed individual features based on their past engagement and interactions with the app, and such observed user heterogeneity can be helpful in the researcher’s analysis of the effects of message design on user engagement, because different users may behave differently in response to the same marketing communication.

We adopt a theory-based framework to first codify push notification messages into a total of 11 design elements, and apply the causal inference machine learning method causal forest to uncover the average treatment effect for each design element on user engagement, with implementation adapted to our empirical context. We find that 5 of the 11 language design elements play a significant role in driving user engagement with push notifications, and these results offer practical guidance for marketing managers in their choice of message design for future push notifications.

Chapter 2: Uncovering the Offline Conversion Funnel with Internet-of-Things: The Case of Wi-Fi Tracking in Retail Industry

2.1 BACKGROUND

In this chapter, we collaborate with a leading IoT analytics company in China and collect a large-scale panel dataset from state-of-the-art mobile Wi-Fi tracking sensors, which are deployed by a successful national retail chain in some of its stores. Second, we further work with the retail chain to acquire all transaction records at those stores. Finally, we obtain information on the price promotion campaigns carried out by the retailer during the study period. By merging the unique and large datasets on individual Wi-Fi tracking records and micro-level transaction data, we uncover the entire offline conversion funnel for hundreds of thousands of consumers at multiple stores of the retail chain, over the period of seven weeks. The compiled datasets provide a comprehensive record of consumers' offline behaviors across all stages of their offline journey and enable us to investigate the impact of a commonly used business strategy – price promotions – in influencing consumers' decisions through the conversion funnel. We use price promotions as a concrete example to illustrate how firms may utilize information gathered by the IoT technology to derive new managerial insights. However, our conceptual framework and modeling approach can readily be extended to other forms of marketing strategies that are often used in offline settings.

We propose a three-stage model to capture consumers' decisions across the entire offline conversion funnel (pass-by, visit and purchase, Figure 2.1). Employing a Bayesian hierarchical modeling framework, we build our model to fully take advantage of the IoT data. In particular, consumers' decisions may be dependent on their individual propensity as well as their past experience with the store, thus our model controls for both observed and unobserved consumer heterogeneity in different stages of the conversion funnel. Furthermore, our model not only allows us to estimate the overall effect of price promotions on sales in different product categories, but also decompose the overall effect into two components: 1) the “visit effect”, which arises from price promotions increasing consumers’ likelihood of visiting the store given a pass-by and 2) the “purchase effect”, which arises from price promotions increasing consumers’ purchase likelihood in a given product category during a store visit.

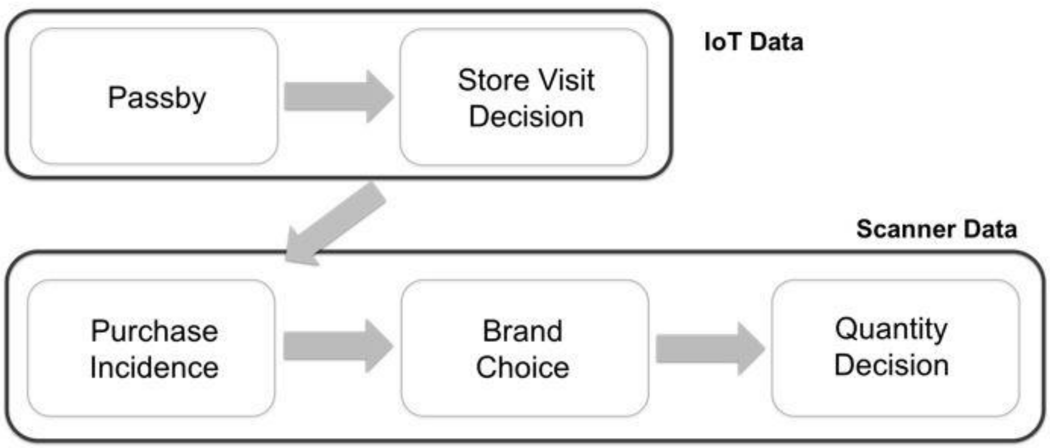

Figure 2.1: Role of the IoT data in completing the offline conversion funnel

We leverage the exogenous variation in marketing mix (price promotions) and estimate our model on data from multiple stores. The exogeneity of price promotions is due to the differences in store product inventory and location, as well as to the temporal variations in the promotion schedule. Not surprisingly, we find that price promotions indeed have a strong and positive effect on consumers' likelihood of visiting the store (visit effect), in addition to the well-understood effect of increasing purchase when consumers do visit the store (purchase effect). More interestingly, we find the visit effect is immensely important vis-à-vis the purchase effect through simulation studies, which use the estimated coefficients in the store visit model and purchase incidence model and take into account unobserved heterogeneity across different stages of the conversion funnel. We define visit effect as the percentage increase in purchases that can be attributed to additional store visits driven by price promotions, and purchase effect as the percentage increase in purchases that can be attributed to increased purchase amounts given consumers' visits. We find that visit effect may account for as much as 25% of the overall increase in category sales for the three product categories we consider. Furthermore, the individual-level data on consumers' offline behaviors (pass-by, visit and purchase) also reveals several key behavioral insights that can be utilized by offline businesses. For instance, our results show that consumers' store visit decisions are influenced by how often they have passed by and how often they have visited the store in the past. The model estimates also reveal that a substantial amount of heterogeneity exists in consumers' baseline propensity to pass by, visit, and purchase.

Finally, we use simulations to show that retailers can more effectively design price promotions at the store level by incorporating IoT data. In particular, we leverage the IoT data at each store to capture its consumers' individual behaviors and derive the optimal price promotion strategy for the store (e.g. customized promotion breadth for each category). We then calculate optimized price promotions across product categories for each store and compare the results with those derived from a model without utilizing individual consumers' location information contained in the IoT data. We find that, whereas a model with IoT data can offer recommendations for price promotions tailored to each product category and store, the model without IoT data consistently fails to produce non-zero price promotion recommendations. Consequently, we demonstrate that implementing IoT technologies and leveraging their information in the design of marketing strategies can bring about significant business value for retailers operating in the offline setting.

In summary, our contribution in this paper is threefold. First, to the best of our knowledge, our study is the first to demonstrate the business value of integrating emerging IoT technologies into the design of marketing strategies for offline businesses. We combine the unique IoT (mobile Wi-Fi tracking) data with commonly used scanner transaction data to fully uncover consumers' offline conversion funnel. Accounting for both observed and unobserved consumer heterogeneity across different stages in the conversion funnel, we use simulations to show that designing optimized price promotions can hardly be achieved without incorporating the individual consumer information contained in the IoT data. Second, by leveraging the IoT data and using the concrete example of price promotions, we identify a new competitive edge – store visits, through which the marketing mix can impact the overall sales performance of an offline business. The discussion of the importance and the quantification of this edge has been scarce in the extant literature, largely owing to the lack of granular data on consumers' offline behaviors prior to purchase. Finally, we propose a holistic approach in modeling consumer decisions across multiple stages in the offline conversion funnel, which can be extended to other similar settings. Within this framework, we construct a correlated model that not only deals with the potential endogeneity in price promotions in the store visit stage, but also accounts for multi-category product purchases.

Overall, our study serves as a first step towards understanding the potential of IoT technology in business applications. Our conceptual framework, modeling approach and managerial insights provide concrete guidance on how firms can leverage the information from the novel IoT technology to better understand consumer behaviors across the entire offline conversion funnel, and then use this knowledge to enhance their bottom lines by designing more effective marketing campaigns.

2.2 Literature Review

IoT is an emerging topic in recent years and has attracted a lot of research interest in various disciplines (Ruiz-Garcia et al. 2009; Almedr and Ersoy 2010; Xu et al. 2011, Yoo et al. 2012). As discussed in two review papers (Hoffman and Novak 2015; Ng and Wakenshaw 2017), IoT could fundamentally impact marketing practice and research. Specifically, IoT may allow for the "liquification" of information about consumers in the physical world (Ng and Wakenshaw 2017, e.g. location and environment) and provide new opportunities for marketers to improve their practice. However, until now, no empirical study has demonstrated the value of IoT technologies to offline retailers using a large-scale dataset based on real-world application of the technology. We intend for our study to fill this gap by quantifying the business value of IoT information, meanwhile also providing a model-based method for designing more effective marketing strategies. Toward this goal, our study is closely related to and build on three streams of literature in marketing.

First, recent technological innovations help businesses better understand and reach out to consumers in offline settings. The use of offline tracking technologies has become increasingly prevalent, and enabled retailers to capture offline consumer shopping behaviors that were previously unobserved and unavailable to the researcher. For example, Hui et al. (2013a) deployed in-store video tracking technology to study important drivers of consumers' unplanned purchase in-store. Hui et al. (2013b) applied in-store radio frequency identification tracking in experimental studies to understand how the length of in-store travel distance affects unplanned spending, and they used simulation to investigate the effectiveness of on-the-spot promotion based on estimates from their first study.

However, until recently, most studies using mobile and offline tracking data have focused on understanding consumers' in-store behavior such as search or purchase (Seiler and Pinna 2017), that is, after they already chose to visit the store. Yet much remains to be known about consumers' offline travel and visit behavior from this emerging literature, and to date, no study has explicitly investigated consumers' offline decision process as a whole to offer a complete understanding of consumers' behavioral patterns throughout the conversion funnel. Furthermore, the potential multi-stage effect of marketing strategies on consumer behaviors across the entire conversion funnel remains largely unexplored. As marketers become enabled to carry out mobile targeting campaigns tailored for each individual consumer in real time[5], there is a commanding need for better knowledge of consumers' behaviors prior to their purchases. The additional insights generated from understanding consumer behavior and how consumers react to marketing interventions across different stages in the offline conversion funnel, will provide valuable input into the design of more effective marketing strategies.

Our study also extends the large stream of literature based on scanner datasets. Scanner data has been widely used by both marketers and economists alike to investigate various aspects of consumer purchase behavior, such as brand choice decisions (Guadagni and Little 1984) and consumer learning (Erdem and Keane 1996) among many others.

[5] IS and Marketing researchers have actively explored how to use digital interventions, especially mobile messaging, to influence user behavior (Ghose et al. 2013; Luo et al. 2014; Ghose et al. 2018a). The rise to prominence of mobile device usage has allowed marketers to reach consumers in real time based on their location, context and behaviors. For example, a series of studies have examined the moderating effect of user geographic location (Ghose et al. 2013; Fang et al. 2015; Fong et al. 2015), local environment (Andrews et al. 2016; Ghose et al. 2018b; Ho et al. 2018), shopping path (Ghose et al. 2018a), timing (Luo et al. 2014), and weather (Li et al. 2017) on consumers' responsiveness to mobile marketing messages. A study by Aral and Nicolaides (2017) also finds that digital notifications with information about peers' exercise behavior have a positive and causal effect on a subject's exercise behavior. This stream of research has established the effectiveness of mobile messages in influencing individual behavior, such as clicking on ads (Andrews et al. 2016) and purchasing tickets (Luo et al. 2014). Moreover, individuals' offline trajectory within a physical space can further help increase the effectiveness of mobile targeting (Ghose et al. 2018a).

Most of the research stream based on scanner data however, is focused on uncovering factors that underlie consumers' decision making when they do make purchases within a store (Guadagni and Little 1984; Gupta et al. 1998; van-Nierop et al. 2010), and is inconclusive about consumers' decision making before a potential purchase occasion arises. By combining data on consumers' offline journey and scanner data, we observe the intermediate outcomes in the early stages of the offline conversion funnel (pass-by and store visit). As a result, we can improve the efficacy of marketing campaigns by examining the effects of marketing mix at different stages of the funnel. Which factors affect consumers' store visit decisions? Do marketing strategies only affect consumers' decisions in store, or does their influence extend to earlier stages of the conversion funnel? Can each individual store design its own optimal marketing strategies, and if yes, how? These questions are hard to answer without directly observing consumers' behaviors prior to purchase over time, and our study helps answer these questions by utilizing the unique IoT data that we previously introduced.

Finally, the insights we gain from the specific example of price promotions, also contribute to the literature on the effectiveness of price promotions. In an offline retail setting, when the consumer has already visited the store, in-store price promotions provide incentives for the consumer to accelerate purchase in a given product category, increase purchase quantity, or switch to different brands (Gupta 1988; Bell et al. 1999). Although price promotions have been consistently documented to generate favorable consumer response during a store visit, very little work has been done to examine the effect of price promotions on consumers' store visit decisions. Managers remain agnostic of the intricate mechanisms through which price promotions can influence sales: is the increase in sales due solely to price promotions affecting in-store purchase behavior or is there more to the story? The reason for this lack of research is two-fold. First, empirical studies of price promotions have predominantly relied on scanner transaction data, which is devoid of information on consumer behavior outside of their recorded shopping trips. Second, the same studies are also generally focused on grocery shopping settings, where consumer shopping trips consist largely of either planned store visits or short fill-in trips (Kahn and Schmittlein 1989). In the broader retailing scene however, consumers’ store visits can be more spontaneous. Consider a consumer passing by a convenience store, who is on her way to a meeting and who did not intend to visit a convenience store, but because she sees the posters for price promotions outside the store, she decides to visit the store. In this manner, price promotions have the potential to convert passersby into store visitors, which in turn may lead to more sales. Second, there was a lack of technology and corresponding data to understand consumers' store visit decisions in the past. By merging scanner data with IoT tracking data, we find that a significant portion of the increase in category sales due to price promotions can be attributed to the additional store visits that price promotions induce. Such insight sheds light on the previously overlooked benefit of price promotions, which may also exist in other types of marketing strategies, such as new product launch and product assortment changes. This finding also implies that retailers may design complementary marketing strategies in conjunction with price promotions to further monetize the increased store visits.

2.3 DATA

We collaborate with a leading IoT analytics company in China and collect data on consumers' offline journey in the retailing sector. By coordinating with a national convenience store chain with thousands of store locations, the IoT analytics company installed its Wi-Fi sensors at a number of the chain’s stores in a coastal province of China. We further work with the chain and obtain all the transaction data as well as price promotion campaign information during the same study period. We construct the data for analysis by combining three data sources: (a) a Wi-Fi dataset that contains mobile Wi-Fi signals for 8 randomly selected store locations of the retail chain between July 15 and September 5 of 2015, (b) a transaction dataset that contains transaction records for the same 8 store locations in the same time frame, and (c) a dataset on marketing mix including multiple bi-weekly or monthly price promotion campaigns implemented by the retail chain across all of its stores during the same time period.

2.3.1 Wi-Fi signal data

A unique aspect of our study lies in the Wi-Fi data collected from the IoT technology, which records all mobile device signals detected by specially designed Wi-Fi signal sensors. The signals detected by the sensors are coded into the following variables: the time (accurate to the second) when a signal is detected, the signal strength, the unique MAC ID of the mobile device which transmits the signal, and the brand (if any) of the mobile device. Such data collection through Wi-Fi sensor does not require consumers to actively connect to an available Wi-Fi source in store, and only records a privacy-preserved anonymous identifier of the mobile device (MAC ID). The Wi-Fi signal sensor is able to detect signals from active devices in a fairly wide range, encompassing the whole area inside the store and the area in close vicinity outside the store (e.g. entire pedestrian path in front of the store). The sensor has been deployed and tested by the analytics firm across hundreds of retailers and tens of thousands of retail stores within the last five years before its deployment to the chain, thus the recorded information of device ID and signal strength is highly reliable and can accurately capture offline traffic around and within a store at the identifiable individual level.

Figure 2.2 shows the distribution of Wi-Fi signal strengths for one store, and it is a representative distribution for all other stores. The distribution of detected signal strength shows two peaks: the mass around the left peak corresponds to weaker signals that are detected from the periphery of the store location within the detection range, and constitutes the large majority (>85%) of all signal readings; the mass around the right peak corresponds to strong signals that are picked up in close proximity to the sensor, which is considered inside the store. Because the signals detected inside the store also contain those coming from mobile devices owned by store clerks, we scrutinize the data and filter out those MAC ID's which may be associated to these individuals. Specifically, this involves examining the number of signals observed for each MAC ID and identifying the ones with anomalously large numbers of in-store signals compared to the rest of the population, then these signals are treated as coming from store clerks' mobile devices and dropped from the dataset.[6] We also examine the distribution of Wi-Fi signals by each brand of devices and find no evidence that the distribution of Wi-Fi signal strength differs by brand, thus we treat signals from differently-branded devices as equivalent.

[6] This leads to dropping between 6 to 10 MAC ID's for each of the 8 stores in our data, which is a reasonable number for the staff size of convenient stores.

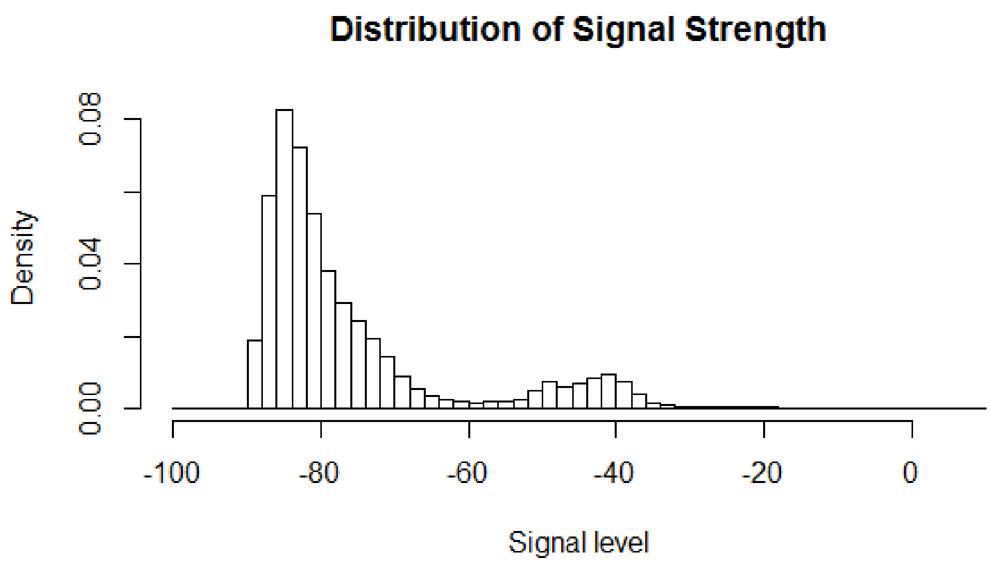

Figure 2.2: Sample distribution of signal strength

Notice that the signal strength distribution displays a rather flat trough in the range of -62 to -56, and this raises the question of what level of signal strength should be used as the threshold to separate in-store signals from outside ones. We do not seek to find the “optimal” threshold according to some metric; instead, we follow the analytics company's suggestion and use -56 as the cutoff, which has been tested extensively and found to be accurate in determining in-store signals. With this threshold in place, we can draw a clear-cut distinction between whether a consumer is outside the store or in-store based on the signal strength of her mobile device. This enables us to model a consumer's store visit decision, which is not feasible with only traditional scanner datasets.

Figure 2.3 below further depicts the traffic pattern by breaking down the number of detected signals by time of the day and different days of the week (weekdays versus weekends). We find that foot traffic around the store varies by different times of the day, and different days of the week. Consequently, in the first stage of our model, we allow consumers' propensity to pass by a store to depend on the time of the day and the day of the week.

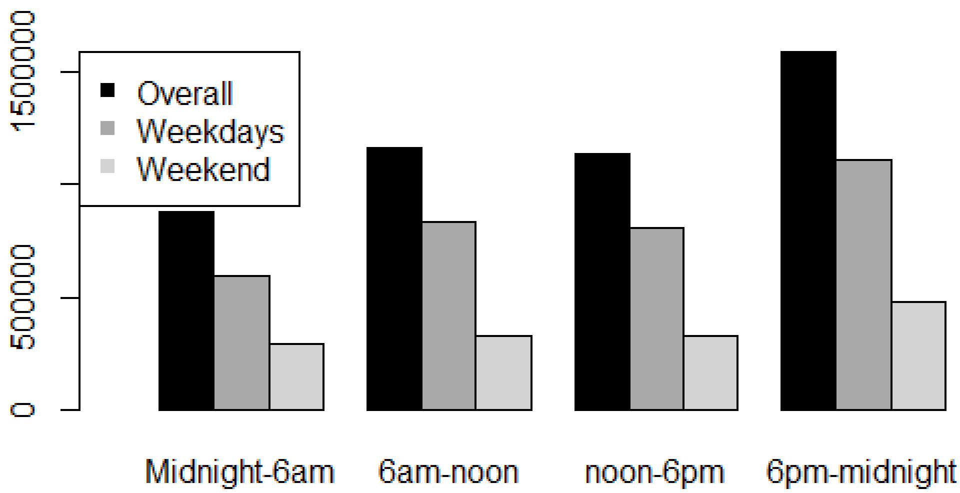

Figure 2.3: Distribution of signal counts by time

2.3.2 Transaction data

The transaction dataset contains a comprehensive record of all scanned transactions that occurred during the data period, and it comprises the following variables: the timestamp (accurate to the second) for when a transaction is processed, a unique transaction code that identifies each transaction, name of the products purchased in that transaction, product code, category code, number of items purchased for each product, price paid, and discounts received.

Depending on the store, the total number of unique products sold varies from 1,000 to 2,363, covering 44 to 51 unique product categories. Common product categories include instant food, snacks, soda, cigarettes, beer and liquor, and household products. We consolidate the common product categories into three aggregated categories: food, beverage and miscellaneous. The food category encompasses all the products in the snacks, instant food and candies categories; the beverage category includes all products that belong to one of the following categories: soda, water, tea and juice, flavored drinks, beer and liquor; the miscellaneous category is composed of household products in all other product categories that do not fall into either the food or beverage category. Certain product categories such as cigarettes, did not have any products on price promotion throughout the entire data period, and thus we exclude them from our analysis.

More than 20% of the transactions involve the consumer purchasing from different categories, and the products most frequently purchased together come from the beverage and food categories. Because of the prevalent co-occurrence of different product categories in transactions, we allow consumers' purchase decisions to be correlated across the three product categories in the second stage of our model through a common covariance structure. The average amount of money spent per transactions ranges from the equivalent of $5 to $10, depending on the store. Furthermore, all transactions are anonymous, meaning there is no exact identification of the consumer who conducted the purchase, and this calls for associating a given transaction to detected mobile devices that may be in-store around the time of transaction, as we explain in Appendix A.

2.3.3 Marketing mix data

For marketing mix variables, we observe all periodical promotion information, which consists of the following variables: the beginning and ending time of each promotion period, the detailed information about each product on promotion, including product names and codes, category names and codes, original prices, promotion prices and discount percentages. There are four promotion periods in our data: July 1 - July 31, August 1 - August 15, August 16 - August 31, and September 1 - September 15. The items on promotion, if available, have the same discount percentages across different chain stores, but the variety of items on promotion and the discount percentages vary across these four promotion periods. It is important to emphasize that the design of price promotions across multiple periods is exogenous to the eight stores that we study, for three reasons: first, promotions are determined at the level of the retailer headquarters based on the sales of and central inventory for hundreds of stores across the region. The same promotion applies to all chain stores at any given time. Therefore, it is unlikely that a given promotion scheme at any time is endogenously designed to capture unobserved demand characteristics specific to any single store in our sample. Second, the price promotion campaign shifts at least every half month and the selection of products on promotion changes drastically. The centralized planning of the price promotion campaign for next period takes place before the end of the current period (to print out and distribute promotion tag and posters), so the promotion is unlikely to be correlated with sales in the current period. Finally, the exogeneity of price promotion to each store is further corroborated by the fact that at any particular point in time, the availability of products that undergo price promotions in the same product category differs across store locations, giving rise to additional variations in actual price promotions in the same product category across stores.

The price promotion scheme is designed by the chain's corporate office and rotated on either a biweekly or monthly basis. The price promotions cover a wide range of different product categories, including candies, instant food, snacks, liquor, carbonated drinks, frozen products etc. We consolidate and construct a promotion breadth variable to capture the promotional activities both throughout the entire store and within each consolidated product category. Category-level promotion breadth is defined to be the proportion of products that are on price promotion within a given category, and the store-wide aggregate promotion breadth is defined to be the proportion of products that are on price promotion among all products sold in the store. Across all categories, promotion periods and stores, the category-level promotion breadth ranges from as low as 1% to as high as 70% given that there is at least one product being on price promotion in the category.

We expend further efforts to integrate the WiFi tracking data, the transaction data and the price promotion data through a deliberate process, based on the data structure and the suggestions from the IoT analytics firm. We document the integration approach in Appendix A. The constructed dataset records each consumer's pass-by, store visit and purchase actions over a 7-week period, as well as the price and discount of each product in the transactions. Such detailed data allows us to investigate consumers' entire offline journey and build a holistic model to understand the nuanced effect of marketing mix across different stages.

2.4 MODELLING FRAMEWORK

In line with a consumer's decision process along the offline conversion funnel, we propose a three-stage model that captures the consumer's pass-by, store visit and category-level purchase decisions in an integrated framework.

2.4.1 Pass-by sessions

In the first stage of the offline funnel, we model the number of pass-by sessions that a consumer makes to the vicinity of a given store location. We assume that a consumer's decision to pass by a given store location is exogenously determined and is independent of the store's marketing activities, but it can depend on time-varying factors such as the time of the day and the day of the week. We divide a day into four windows: [midnight, 6am], [6am, noon], [noon, 6pm] and [6pm, midnight], and denote them as window 0, window 1, window 2 and window 3 respectively. We capture consumers' arrival to the vicinity of the store with a Poisson process as follows:

$$\ln \lambda_{iq} = \alpha_{i0} + Q_q \cdot \alpha_Q + W_q \cdot \alpha_W + Q_q \cdot W_q \cdot \alpha_{QW}$$

where λ_{iw} is the Poisson parameter characterizing consumer i's arrival process at the store in window q: $n_{iq} \sim$Poisson(λ_{iq}), where n_{iq} is the number of pass-by sessions made by the consumer to the store during time window q. Q_q comprises three dummy variables for the four windows of the day, with window 0 as the baseline. W_q is a dummy variable for whether the pass-by sessions fall on a weekend, and we also include the interactions between the window dummies and the weekend dummy. Lastly, α_{i0} is an individual-specific intercept that follows a common prior normal distribution which is fixed at the focal store: $\alpha_{i0} \sim N(\mu_\alpha, \sigma_\alpha)$. In essence, α_{i0} captures an individual consumer's baseline propensity to pass by the given store location, which reflects the consumer's habitual pattern of presence in the surrounding area. The mean of the prior distribution of the random intercept, μ_α, is store-specific, which is equivalent to a store-level fixed effect that captures any systematic differences in the baseline pass-by propensities across stores.

2.4.2 Store visit decisions

Given consumer i's pass-by session t to the vicinity of the store, we construct a latent-utility model for the consumer's decision to visit the store. We first model the latent utility of a store visit by the following specification that corresponds with a probit model:

$$UV_{it} = \beta_{io} + X_{it,1\sim 8} \cdot \beta_1 + X_{it,9} \cdot \beta_2 + \varepsilon_{it}$$

where we let the latent utility of store visit, UV_{it}, be a function of consumer i's historical and current-session behavioral characteristics, the consumer's phone brand, and the price promotion that the consumer faces. In particular, the set of variables in $X_{it,1\sim 8}$ includes consumer i's cumulative number of pass-by sessions and cumulative number of store visits up to the current session t, the time duration that the consumer spent outside of the store, whether or not the consumer had a prior session, whether or not the consumer made a prior visit to the store,), time passed since the consumer's last session (in hours and set to 0 if there has been no prior session), time passed since the consumer's last visit (in hours and set to 0 if there has been no prior visit, and the consumer's phone brand indicators. $X_{it,9}$ is the total effective breadth of price promotions across all products that have been sold in the store, taking into account the popularity of each product by giving more weights to products with higher sales volumes in the current promotion period. We use promotion breadth as our main variable of interest for marketing mix variables, because studies in existing literature have included regular price, promotion breadth and promotion depth as covariates in the analysis of sales outcomes (Ailawadi et al. 2010; Bezawada and Pauwels 2013). An individual-specific random intercept, β_{io}, captures unobserved consumer heterogeneity in terms of their baseline benefit of visiting the store. And ε_{it} is a normally distributed error term with unit variance. Similar to specifications that have been used in existing literature, we include consumer *i*'s cumulative number of pass-by sessions and cumulative number of store visits to reflect the consumer's cost of visiting the store, since they can capture the consumer's experience and familiarity with the store and its environment. For example, Li and Kannan (2014) allow a consumer's cost of visiting an online channel to depend on the consumer's cumulative experience with the channel.

The consumer visits the store if $UV_{it} > 0$ and does not visit the store otherwise. We set β_{io} to follow a common prior normal distribution that is fixed at the focal store: $\beta_{io} \sim N(\mu_\beta, \sigma_\beta)$. The mean of the prior distribution, μ_β, is store-specific and allows us to tease out unobserved systematic differences in store environment that drives store visits.

Intuitively, since the promotion posters outside the stores present a conglomerate of price promotion information for a multitude of products, a consumer's decision to visit is more likely to be influenced by her perception of the overall promotion activities in the store than the promotional information about any specific product or category. On the other hand, instead of using a large array of promotion variables for each product or category, we construct an aggregate measure of price promotions that also allows for differential effect on each individual consumer. Specifically, $X_{it,9}$ is constructed as an interaction between the store-wide promotion breadth weighted by each product's sales volume in the current promotion period, and the consumer's number of cumulative visits up to the current session *t* during the current promotion period.[7] In essence, this term captures the idea that price promotions have a larger impact on frequent shoppers' store visit decision than less frequent ones, and it is consistent with findings in extant literature which suggest that shoppers who make more shopping trips will benefit more from price variations (Bell and Latin 1998).

The store visit stage is central to our model, since it allows us to directly capture the impact of price promotions on a consumer's store visit decisions. With the consumer traffic data, this feature of our model enables us to quantify and attribute consumers' in-store purchases to the effects of the marketing mix variables which either lead to an increased number of store visits, or give rise to more purchases once the consumer visits the store. This component of our model also warrants a discussion with respect to identification. There are two major concerns that the price promotion variable may be endogenous: 1) price promotions can be designed by the corporate office to temporally match unobserved (to the researcher) demand patterns that give rise to overstock of certain products, i.e., $X_{it,9}$ may be contemporaneously correlated with ε_{it}; 2) price promotions may be designed by the corporate office to capture unobserved demand characteristics that are specific to the geographical locations of the stores, i.e. $X_{it,9}$ may be correlated with ε_{it}

[7] Sales-volume weighted promotion breadth is given by $\sum I_k \cdot V_k$, where I_k is an indicator for whether product k is on sale, and $V_k = {S_k}/{\sum S_j}$ is the sales volume of product k divided by the sales volume of all products sold in the store.

it through store-specific factors shared by the stores in question. The first concern is alleviated by the fact that the stores in our data are all franchisees, so the corporate office does not have real-time knowledge about each of these store's inventory level, preventing it from being able to utilize inventory information in its design of price promotion schedules that are specifically targeted at reducing overstocked products (at least for the stores in question). The second endogeneity concern is addressed by observing that the stores in our data are all located in very different geographical areas, with thousands more store locations belonging to the retail chain scattered all throughout the province. Therefore, it is highly unlikely that the corporate office designs its price promotion schedules according to any unobserved demand characteristics that are particular to the 8 store locations in our data. Moreover, we use a store-specific mean for the prior distribution of unobserved heterogeneous intercept, $\beta_{io} \sim N(\mu_\beta, \sigma_\beta)$, which similar to a store-level fixed effect, further mitigates any correlation that may exist between the error and the price promotion schedule. Regardless, in our empirical analysis, we include a specification that allows for the correction of potential endogeneity in the promotion variable, and find little evidence that endogeneity poses a serious issue.

2.4.3 In-store purchase decisions

We model a consumer’s purchase decisions in terms of the monetary value spent on purchases in each of the three product categories. A large body of prior literature has thoroughly investigated the brand choice phenomenon (Guadagni and Little 1984; Kamakura and Russell 1989; Bucklin and Gupta 1992; Chintagunta 1993) in various retail product categories, hence we focus only on the aggregate outcome of marketing strategies at the category level, rather than the sales outcome of any particular brand or product.

Given consumer i's visit to the store, we model her latent utility of purchasing from category c during store visit t by the following specification that corresponds to a tobit model:

$$UP_{it}^{c} = \delta_{i0}^{c} + Z_{it,1\sim 9}^{c} \cdot \delta_{1}^{c} + Z_{it,10}^{c} \cdot \delta_{2}^{c} + \tau_{it}^{c}$$

where the latent utility of purchase in category c, UP_{it}^{c}, is a function of the consumer's historical and current-session behavioral characteristics, the consumer's phone brand, and the breadth of price promotion that the consumer faces for products in the category. In particular, the set of variables in $Z_{it,1\sim 9}^{c}$ includes consumer i's cumulative number of pass-by sessions and cumulative number of store visits up to the current session t, the time duration that the consumer spent outside the store, the time duration that the consumer spent inside the store, whether or not the consumer had a prior session, whether or not the consumer made a prior visit, time passed since the consumer's last visit (in hours and set to 0 if there has been no prior visit), time passed since the consumer's last session (in hours and set to 0 if there has been no prior session), and the consumer’s phone brand. $Z_{it,10}^{c}$ is the breadth of price promotions across all products in category c that have been sold in the store, weighted by each product’s sales volume in the current promotion period. An individual-specific random intercept, δ_{i0}^{c}, captures unobserved consumer heterogeneity in terms of their baseline benefit of purchasing from category c. τ_{it}^{c} is a normally distributed error term with variance σ_{c}^{2} that is to be estimated.

The consumer makes a purchase of monetary value UP_{it}^{c} if $UP_{it}^{c} > 0$, and does not make a purchase otherwise. The individual random intercept, δ_{i0}^{c}, follows a common prior normal distribution that is specific to the store and category: $\delta_{i0}^{c} \sim N(\mu_{\delta}^{c}, \sigma_{\delta}^{c})$. We again note that, by allowing for a store- and category-specific prior distribution for the heterogeneous intercepts, we also control for the store-specific unobserved demand characteristics that may affect a consumer's decision to purchase in the category. Similar to the store visit decision model, this specification is equivalent to having store and category-level fixed effects, which mitigate any potential concern of endogenous price promotions in the purchase decision model.

2.4.4 Correlation between visit and purchase decisions

Because there may exist unobserved factors that simultaneously impact a consumer's store visit and category purchase decisions, we allow our second stage visit model and third stage purchase models to be correlated through a common covariance structure. Recall that we have a probit model for the second stage visit decision, and tobit models for the third stage purchase decisions, letting y_{it}^{v} and y_{it}^{c} denote consumer i's observed store visit and category purchase outcomes, we have that:

$$y_{it}^{v} = \begin{cases} 1 & UV_{it} > 0 \\ 0 & UV_{it} \leq 0 \end{cases}$$

and

$$(y_{it}^{c} | y_{it}^{v} = 1) = \begin{cases} UP_{it}^{c} & UP_{it}^{c} > 0 \\ 0 & UP_{it}^{c} \leq 0 \end{cases}$$

for $c = 1,2,3$. Thus, we let the errors, ε_{it}, τ_{it}^{1}, τ_{it}^{2}, τ_{it}^{3} follow a multinomial normal distribution:

$$(\varepsilon_{it}, \tau_{it}^{1}, \tau_{it}^{2}, \tau_{it}^{3})' \sim MVN(0, \Sigma)$$

where Σ is a covariance matrix with non-zero off-diagonal elements to account for correlations between the store visit and category purchase decisions.

2.4.5 Latent instrumental variable for potential endogeneity

To account for potential endogeneity of price promotion in the store visit model, we also develop an extension of the correlated modeling framework that includes a latent instrument for the price promotion variable in the store visit decision. In particular, recall that consumer i's latent utility for store visit is given by: $UV_{it} = \beta_{io} + X_{it,1\sim 8} \cdot \beta_1 + X_{it,9} \cdot \beta_2 + \varepsilon_{it}$, where $X_{it,9}$ is the price promotion variable that could potentially be endogenous by way of correlation with the error term ε_{it}. Hence we introduce a latent instrumental variable, γ_i, to separate out the endogenous and exogenous variation in price promotion:

$$X_{it,9} = \omega \cdot \gamma_{it} + \xi_{it}$$

where γ_{it} is a latent categorial variable which follows a multinuolli (generalized Bernoulli) distribution $\gamma_{it} \sim multinulli(\pi_1, \dots, \pi_k)$, with $\pi_1, \dots, \pi_k \sim Dirichlet(r_1, \dots, r_k)$, and ω is the corresponding category mean for a given latent category. The extra error term from the latent instrument variable equation now joins with the errors from both store visit and in-store purchase equations to capture the potential endogeneity of price promotions:

$$(\varepsilon_{it}, \tau_{it}^1, \tau_{it}^2, \tau_{it}^3, \xi_{it})' \sim MVN(0, \Sigma)$$

where again Σ is a covariance matrix with non-zero off-diagonal elements.

2.5 Estimation and Simulation Results

2.5.1 Estimation

We carry out the model estimation under Bayesian inference with a Markov Chain Monte Carlo (MCMC) sampler implemented in R. We run the sampling chain on the model parameters for 20,000 iterations for each store, where the first 10,000 iterations are treated as burn-in, and we keep the last 10,000 posterior draws for conducting statistical inferences on the parameters as well as carrying out simulation studies. The posterior trajectories of the MCMC sampling chains are inspected to determine that the sampler has converged to stationary draws from the posterior distribution of the model parameters. We present the details of the sampling chain in Appendix B.

2.5.2 Estimation results

Table 2.1, 2.2, 2.3 report the estimation results for the pass-by, store visit, and purchase stage for 3 random stores in our data.

Variable	Store 1		Store 2		Store 3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Weekend	**0.160**	0.011	**0.052**	0.020	**0.096**	0.045
Window 1	**0.551**	0.008	**1.425**	0.011	**2.009**	0.032
Window 2	**0.539**	0.009	**1.482**	0.011	**2.234**	0.031
Window 3	**0.811**	0.008	**1.601**	0.011	**2.314**	0.031
Weekend× Window1	**-0.213**	0.015	**-0.080**	0.022	**-0.076**	0.049
Weekend× Window2	**-0.104**	0.014	**-0.082**	0.023	**-0.126**	0.047
Weekend× Window3	**-0.103**	0.013	**-0.258**	0.023	**-0.173**	0.048
Intercept Mean	**-4.690**	0.008	**-5.324**	0.010	**-0.611**	0.031
Intercept SD	**1.265**	0.008	**1.261**	0.008	**1.285**	0.015

Estimates whose 95% credible intervals do not include 0 are in bold

Table 2.1: Estimates for pass-by model

In the pass-by stage of the model, we find that consumer traffic in terms of the number of pass-by sessions indeed varies by different times of the day and different days of the week. For example, for Store 1, a day on the weekend is associated with significantly more consumer pass-bys than on a weekday, and windows 1-3 (6am-midnight) all witness significantly more consumer traffic compared to the baseline of Period 0 (midnight-6am). Furthermore, there is an interaction effect between weekend and the daily time windows, which shows a decrease in consumers' propensity to pass by during Periods 1-3 (6am-midnight) on weekends, and this suggests that the increase in weekend pass-by activities around this particular store is due only to an increase in the number of pass-bys during Period 0 (midnight-6am) on weekends. The estimates for the hyperparameters of the random intercept suggest that there is substantial heterogeneity in consumers' baseline propensity to pass by in the vicinity of the store in any given time window. While some estimates are not significantly different from 0, estimates for Store 2 and Store 3 suggest that similar qualitative conclusions can be drawn about consumers' pass-by behavior.

In the store visit stage of the model, depending on the store, a consumer's cumulative sessions and cumulative visits may be associated with their probability of visiting the store. This implies that a consumer with repeat pass-bys or repeat visits behave differently when it comes to their decisions to visit the store. Consumers who spend more time outside the store are found to be associated with a higher probability of visiting the store, as indicated by the positive coefficient estimates on the variable $Out\ Duration$. Both having prior sessions and having prior visits are associated with a lower probability of visiting, which implies that consumers with more experience with the store and its physical environment have a lesser need to visit the store on any given pass-by, because they may choose to easily return for a visit in the future. Interestingly though, the longer the time gap between a consumer's current session and their last session, the higher is their probability of visiting the store in the current session. However, no consistent pattern exists between the time gap from a consumer's last visit to the current session, and their probability of visiting the store during the current session. Most importantly, the estimates for price promotion breadth

Variable	Store 1		Store 2		Store 3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Cumulative Sessions	**4.7×10^{-4}**	1.3×10^{-4}	**-4.8×10^{-4}**	1.6×10^{-4}	4.1×10^{-4}	2.7×10^{-4}
Cumulative Visits	**-1.9×10^{-5}**	3.7×10^{-6}	9.4×10^{-6}	9.9×10^{-6}	1.9×10^{-5}	1.6×10^{-5}
Out Duration	**9.0×10^{-5}**	5.6×10^{-6}	**1.7×10^{-4}**	6.4×10^{-6}	**1.8×10^{-4}**	9.5×10^{-6}
Prior Session	**-0.351**	0.032	**-0.305**	0.031	**-0.345**	0.037
Prior Visit	**-0.310**	0.025	**-0.062**	0.025	**-0.224**	0.032
Time Since Last Session	**1.1×10^{-3}**	7.4×10^{-5}	**8.0×10^{-4}**	6.7×10^{-5}	**1.3×10^{-3}**	1.1×10^{-4}
Time Since Last Visit	-4.8×10^{-5}	6.3×10^{-5}	**-5.3×10^{-4}**	6.1×10^{-5}	**2.8×10^{-4}**	8.8×10^{-5}
Promotion Breadth	**1.376**	0.045	**0.697**	0.034	**0.051**	0.009
Intercept Mean	**-1.279**	0.032	**-1.412**	0.032	0.058	0.032
Intercept SD	**0.426**	0.015	**0.396**	0.015	**0.570**	0.018

Estimates whose 95% credible intervals do not include 0 are in bold

Table 2.2: Estimates for store visit model

are found to be consistently positive across stores, which suggests that price promotion indeed positively influences consumers' store visit decisions. The estimated hyperparameters of the random intercept suggest that there is also a considerable amount of heterogeneity in consumers' baseline propensity to visit the store, after controlling for all other observable factors. The phone brand fixed effects are omitted because they are peripheral to our analysis.

Results for the in-store purchase model are reported in Table 2.3, and we present the estimates for all three product categories for each store. Most consistently across stores, a higher time duration spent in-store is associated with higher purchase values in all product categories. $Time\ Since\ Last\ Session$ and $Time\ Since\ Last\ Visit$ both demonstrate consistent correlations with purchase values: while time since last session is positively correlated with purchase values, times since last visit is negatively correlated with purchase values. Having prior sessions is also associated with lower purchase values across product categories and across stores in general, which suggests that consumers who visit the store during their first session tend to make more purchases. The estimates for $Cumulative\ Sessions$, $Cumulative\ Visits$, $Out\ Duration$, $Prior\ Visit$ all show mixed results depending on the product category and the store. This implies that each store can be differently characterized by its consumers' behavioral patterns, and such information would be helpful in tailoring the marketing strategy for each individual store. Most importantly among the results, is that the estimates on $Promotion\ Breadth$ are consistently significantly positive across products and stores, confirming the effectiveness of price promotion at inducing more purchases from consumers. The estimated hyperparameters of the random intercept suggest there is also substantial heterogeneity in consumers' baseline willingness to purchase in any given product category across the stores. We again omit the phone brand fixed effects from the reported results as they are not central to the insights we offer.

	Store 1			Store 2			Store 3		
Variable	Food	Bev.	Misc.	Food	Bev.	Misc.	Food	Bev.	Misc.
Cumulative Sessions	**0.015**	-0.003	-0.009	**-0.108**	-0.062	**-0.061**	**-0.026**	**-0.046**	**-0.087**
Cumulative Visits	**-5.1×10^{-4}**	**-6.2×10^{-4}**	-8.2×10^{-4}	**9.7×10^{-4}**	**1.1×10^{-3}**	6.0×10^{-4}	2.6×10^{-4}	**0.001**	**0.002**
Out Duration	**-5.2×10^{-4}**	**-2.4×10^{-4}**	**-5.4×10^{-4}**	**1.2×10^{-3}**	**1.6×10^{-3}**	7.5×10^{-4}	**2.6×10^{-3}**	**1.9×10^{-3}**	**2.8×10^{-3}**
In Duration	**0.105**	**0.076**	**0.117**	**0.030**	**0.035**	**0.023**	**0.011**	**7.2×10^{-3}**	**8.8×10^{-3}**
Prior Session	1.785	**-5.717**	**-10.940**	**-13.995**	**-11.273**	**-16.192**	**-3.9969**	**-6.092**	**-9.041**
Prior Visit	**-7.021**	**-5.843**	**-7.918**	**5.125**	0.879	**8.911**	-1.111	0.656	-1.815
Time Since Last Session	**8.4×10^{-3}**	**0.016**	0.014	**0.016**	**0.022**	**0.027**	**0.017**	**0.019**	**0.032**
Time Since Last Visit	1.5×10^{-4}	**-8.9×10^{-3}**	2.1×10^{-3}	**-0.024**	**-0.024**	**-0.050**	**-0.011**	**-0.010**	-0.009
Promotion Breadth	**13.934**	**7.200**	**12.561**	**1.422**	0.556	0.344	**1.210**	**0.681**	**3.851**
Intercept Mean	**-40.595**	**-39.433**	**-114.422**	**-54.329**	**-58.291**	**-67.666**	**-20.369**	**-23.971**	**-63.836**
Intercept SD	**22.415**	**16.226**	**33.198**	**0.861**	**12.993**	**1.356**	**10.531**	**9.016**	**14.404**

Estimates whose 95% credible intervals do not include 0 are in bold

Table 2.3: Estimates for purchase model

2.5.3 Results with latent instrument variable

We also estimated the model with a latent instrument variable approach described in Section 2.4.5. We tested the approach with 2 and 3 latent classes, and find that the correlation between the errors in the visit equation and the latent instrument equation is generally quite low (~0.1). Thus, we conclude that endogeneity of price promotion in the visit equation does not pose a serious econometric issue in our empirical analysis.

2.5.4 Decomposition of the price promotion effects

To decompose the effect of price promotions on purchases that is attributable to the promotion attracting more store visits versus the promotion inducing more in-store purchases given store visits, we may begin by examining a single consumer's unconditional purchase value equation in category c:

$$Purchase\ Value_i^c = (Purchase\ Value_i^c|Visit) \cdot Pr(Visit)$$

and consider the promotion elasticity of purchase value with respect to promotion breadth:

$$\begin{aligned}
\epsilon_i^c = \frac{\partial(Purchase\ Value_i^c)}{\partial Promotion\ Breadth^c} \cdot \frac{Promotion\ Breadth^c}{Purchase\ Value_i^c} \\
&= \left[\frac{\partial(Purchase\ Value_i^c|Visit)}{\partial Promotion\ Breadth^c} \cdot Pr(Visit) + \frac{\partial Pr(Visit)}{\partial Promotion\ Breadth} \right. \\
&\left. \cdot \frac{\partial Promotion\ Breadth}{\partial Promotion\ Breadth^c} \cdot (Purchase\ Value_i^c|Visit)\right] \\
&\cdot \frac{Promotion\ Breadth^c}{Purchase\ Value_i^c} \\
&= \frac{\partial(Purchase\ Value_i^c|Visit)}{\partial Promotion\ Breadth^c} \cdot \frac{Promotion\ Breadth^c}{Purchase\ Value_i^c|Visit} \\
&+ \frac{\partial Pr(Visit)}{\partial Promotion\ Breadth} \cdot \frac{\partial Promotion\ Breadth}{\partial Promotion\ Breadth^c} \\
&\cdot \frac{Promotion\ Breadth^c}{Pr(Visit)} = \epsilon_i^c(Purchase\ Value_i^c|Visit) + \epsilon_i^c(Visit)
\end{aligned}$$

where we decompose the elasticity into 2 parts: one corresponding to the promotion elasticity of purchase value given visit, and another corresponding to the promotion elasticity of visit, both with respect to price promotion. The term $\frac{\partial Promotion\ Breadth}{\partial Promotion\ Breadth^c}$ is the derivative of overall promotion breadth with respect to the promotion breadth in category c, and it will be 1 if we change promotion breadths for all categories by the same percentage at the same time.

Then we can define

$$VE = \frac{\epsilon_i^c(Visit)}{\epsilon_i^c\left(Purchase\ Value_i^c|Visit\right)+\epsilon_i^c(Visit)},$$

$$PE = \frac{\epsilon_i^c\left(Purchase\ Value_i^c|Visit\right)}{\epsilon_i^c\left(Purchase\ Value_i^c|Visit\right)+\epsilon_i^c(Visit)}$$

as the visit effect and purchase effect, which capture the percentage contributions to increased purchase values that could be attributed to either more store visits or increased purchases given a store visit respectively. However, because of the presence of unobserved heterogeneity in individual consumers' propensity to pass by, visit the store and purchase from different product categories, and the timing of a consumer's pass-by sessions is also uncertain, these terms are analytically intractable.

As an alternative approach, we use simulations to gauge the magnitudes of visit effect versus purchase effect and investigate their percentage contributions to the total change in purchase values across product categories. Toward this goal, we consider each consumer's pass-by propensity and forward simulate their pass-by sessions over the time period of one day, immediately following their last observed day of appearance in the data. As a consumer can have potentially multiple pass-by sessions (a consumer may appear multiple times around the store on the same day), we shall consider the overall increase in store visits and sales from each consumer during that day. Specifically, we use the following identity:

$$
\begin{aligned}
&Visit' \cdot Purchase' - Visit \cdot Purchase \\
&\qquad = \underbrace{(Visit' \cdot Purchase - Visit \cdot Purchase)}_{Visit\ effect} \\
&\qquad + \underbrace{Visit' \cdot Purchase' - Visit' \cdot Purchase}_{Purchase\ effect}
\end{aligned}
$$

where $Visit'$ and $Purchase'$ denote the overall number of visits and value of products purchased in the category after an increase in price promotion, while $Visit$ and $Purchase$ are the ones before.[8] The first term on the right hand side of the equation, which corresponds to the visit effect, reflects the change in purchase value in the product category that is solely driven by the change in store visits through an increase in price promotion.

[8] Alternatively, we could also decompose the difference as $Visit' \cdot Purchase' - Visit \cdot Purchase = (Visit' \cdot Purchase' - Visit \cdot Purchase') + (Visit \cdot Purchase' - Visit \cdot Purchase)$, where the first difference on the right is the visit effect, and the second difference would be the purchase effect. However, because $Purchase'$ is expected to be larger than $Purchase$ after an increase in price promotion, this will tend to produce a larger estimate for the visit effect. In essence, our chosen method of decomposition provides a conservative estimate for the visit effect of price promotion.

The second term represents the purchase effect, and it captures the change in purchase value in the product category that is driven only by a change in the value of products purchased given that a store visit is made.

We simulate the effect of an increase in price promotions at different magnitudes and decompose its overall effect on category purchase value into the visit effect and purchase effect components, as described above. Table 2.4 through 2.6 report the decomposition of effects for each of the three product categories aggregated across all stores. The first column in each table shows the % of relative increase in price promotions, e.g., a 10% increase means that if the original promotion breadth is x, the new promotion breadth level at which sales outcomes are simulated is $(1 + 10\%)x$.

% Promotion Increase	Visit Effect	Purchase Effect	% Contribution of Visit Effect
10%	7.383	20.442	26.53%
20%	14.517	44.485	24.60%
30%	21.494	72.135	22.96%
40%	28.154	103.365	21.41%
50%	34.993	138.026	20.22%
60%	41.617	176.984	19.04%
70%	49.061	219.107	18.29%
80%	54.566	264.334	17.11%
90%	60.624	314.853	16.15%
100%	68.233	368.648	15.62%

Table 2.4: Promotion effect decomposition for food category

% Promotion Increase	Visit Effect	Purchase Effect	% Contribution of Visit Effect
10%	4.637	13.491	25.58%
20%	9.231	29.792	23.66%
30%	13.374	49.393	21.31%
40%	18.039	71.534	20.14%
50%	21.909	96.994	18.43%
60%	26.063	125.307	17.22%
70%	29.993	157.641	15.98%
80%	33.859	192.377	14.97%
90%	37.718	230.11	14.08%
100%	41.896	270.936	13.40%

Table 2.5: Promotion effect decomposition for beverage category

Based on the simulation results, we find that for all three product categories, a significant proportion (more than 20%) of the increase in the value of products sold can be attributed to the visit effect. The exact percentage varies depending on the magnitude of promotion increase, and generally decreases as the percentage of promotion increase becomes larger. This is because the effect of higher price promotions on store visits is comparatively smaller than the effect on purchases given visits; i.e., very high price promotions will induce almost all visitors to purchase more but not proportionally as many passersby to visit. Despite this, the percentage contribution of the visit effect to the overall increase in purchase values across all three product categories remains well over 20% for small perturbations (10%-20%) in the promotion breadth.

% Promotion Increase	Visit Effect	Purchase Effect	% Contribution of Visit Effect
10%	1.626	5.782	21.94%
20%	3.459	13.218	20.74%
30%	5.219	22.322	18.95%
40%	7.045	33.174	17.52%
50%	8.706	46.856	15.67%
60%	10.389	62.379	14.28%
70%	12.107	80.272	13.11%
80%	13.338	99.941	11.77%
90%	14.87	123.704	10.73%
100%	17.492	148.968	10.51%

Table 2.6: Promotion effect decomposition for miscellaneous category

2.5.5 Optimization of price promotions and value of IoT data

Although the stores all belong to the same retail chain, they vary in geographical locations, foot traffic, sales revenue and behavioral characteristics of consumers, as summarized in Table 2.7. These observed differences across stores afford us an opportunity to examine the potential of leveraging IoT data in improving the design of marketing strategies to achieve better business bottom lines. Specifically, we can optimize the price promotions for each product category in each store, taking as given the idiosyncratic traffic pattern, visit and purchase conversion rates of the consumers at each store.

Store	Location Type	Traffic (Sessions)	Sales Revenue	Store Visit	Repeat Passersby	Repeat Visitors	Repeat Purchasers
1	Urban Village[9]	304,832	244,018	2.08%	38.75%	23.02%	17.85%
2	Unknown	315,591	269,337	2.54%	36.34%	20.63%	19.36%
3	Residential	401,702	223,588	2.79%	27.44%	22.61%	20.14%
4	Hotel	539,363	159,531	2.35%	32.40%	20.16%	16.24%
5	Marketplace	436,651	142,947	1.83%	31.98%	20.94%	13.36%
6	Office Building	216,754	232,703	3.23%	31.22%	18.38%	15.19%
7	Residential	114,745	276,486	8.14%	32.67%	28.08%	26.76%
8	Residential	149,278	215,001	6.11%	27.19%	21.88%	21.12%

Table 2.7: Differences in store characteristics

We allow the breadth of price promotions in each category to vary as a free parameter, and search for the optimal level of promotion breadth that maximizes the expected profit from sales in a given category. Mathematically, we consider the following expected profit equation for category c:

$$E\Pi_c(\rho_c|D) = E_{N(D)}\left\{\sum_{N(D)} I(Visit) \cdot E[y^c \cdot (m_c - \rho_c)|Visit]\right\}$$

where D denotes the time duration over which we aggregate profit, $N(D)$ is the number of pass-by sessions during this time period, m_c is the profit margin for the product category, y^c is the value of products purchased in the category, and ρ_c is the breadth

[9] This type of area is usually associated with low-income populations in the context of urban planning in China

of price promotions that we optimize over. Thus, we search for an optimal value of ρ_c, ρ_c^*, such that

$$\rho_c^* = argmax\ E_{N(D)}\left\{\sum_{N(D)} I(Visit) \cdot E[y^c \cdot (m_c - \rho_c)|Visit]\right\}$$

Since the number of pass-by sessions over the given time period D, $N(D)$, is a random count variable (non-homogeneous Poisson process) that depends on both the length of the time duration and unobserved heterogeneity across consumers, the expected profit equation does not have a closed-form expression. Furthermore, whether a given pass-by session turns into a store visit, and whether a store visit converts into purchases both depend on consumer-specific parameters and their observed behavioral characteristics. Hence, we use a simulation approach to search for an approximate solution to the problem of optimal price promotions, by calculating the simulated expected profit over a grid of candidate values for ρ_c, and choosing the ρ_c^* that gives rise to the highest expected profit.

In practice, we fixed the time duration D to be one day immediately after each consumer's last observed session, and first simulate the number of pass-by sessions made by each consumer within that day. Then for each simulated pass-by session, we draw random errors to compute the latent utility of a store visit. And last, for each pass-by session where the latent utility of store visit is positive, we simulate the latent utility of purchase in a given product category over the candidate breadth of price promotion values. Since we do not have data on the retail chain's costs or gross profit margins, we consult industry reports and refer to available profit margin information on convenience stores.[10] This leads

[10] See https://www.statista.com/statistics/887938/convenience-store-profit-margin-us-product-category/ for a breakdown of gross profit margin by product category for convenience stores in the US and https://retailowner.com/Benchmarks/Food-and-Beverage-Stores/Convenience-Stores#2898437-gross-margin for a summary of the overall gross profit margin.

us to assume an average gross profit margin of 30% for each product category. We also allow the possible ρ values to vary between 0% and 25%, with a grid size of 1%.

As a baseline for comparison, we also simulate the expected profit and search for maximizing levels of price promotions from an aggregate model without IoT traffic data. Specifically, the model now encompasses two stages: a Poisson model for the total number of store visits (as determined from IoT data) within each time window of a day, and a tobit model for the overall category purchase values as a function of the number of store visits and the price promotion breadth within each time window of a day. Table 2.8 compares the optimal price promotion breadths from the model with IoT data and the model without.

	With IoT Data			Without IoT Data		
	Food	Beverage	Misc.	Food	Beverage	Misc.
Store 1	20%	16%	17%	12%	0%	13%
Store 2	12%	10%	7%	0%	0%	0%
Store 3	17%	13%	22%	10%	0%	10%
Store 4	19%	20%	15%	0%	5%	0%
Store 5	8%	10%	2%	13%	9%	24%
Store 6	13%	15%	2%	9%	0%	0%
Store 7	9%	11%	13%	0%	2%	8%
Store 8	12%	5%	9%	0%	0%	0%

Table 2.8: Comparison of maximizing levels of price promotion breadths

Consistently across categories and stores, we see that the model with IoT traffic data produces optimal levels of price promotion breadths that are more aggressive than their counterparts from the model without IoT data. In fact, in many of the cases for the model without IoT traffic data, the optimal price promotion breadth is set to zero, which even goes counter to the retail chain's actual practice. The reason for this difference is twofold: first, the aggregate model does not account for the effect of price promotion on consumers' store visit decisions; second, the aggregate model also does not capture the effect of price promotion on the probability of *individual consumers*' purchase, hence inadequately accounting for the extra extensive margin brought about by higher levels of price promotions, and allowing the intensive margin to dominate which forces price promotions to be as low as possible.

2.6 DISCUSSIONS

With the advancement of IoT technologies and widespread adoption of mobile phone tracking devices, this study demonstrates the business value of IoT technologies and addresses the important question of how to leverage information captured by IoT technologies to improve marketing strategies. In particular, mobile Wi-Fi tracking operationalized by IoT sensors enables retailers to construct a unique dataset which combines consumers' traffic information with the traditional scanner data to form a complete profile of the offline purchase conversion funnel. As we have demonstrated, retailers are now able to investigate the effects of the marketing mix variables on consumers' behaviors at different stages of the offline conversion funnel, including both the store visit decision and the in-store purchase decision. We separate the effect of price promotions on sales into its two components: the visit effect and the purchase effect. We find that the visit effect accounts for a significant proportion of the overall increase in sales value across the three product categories we consider. In particular, for small perturbations in the breadth of price promotions at the category level, the visit effect accounts for as much as 25% of the overall increase in sales. To our best knowledge, such "visit effect" has not been discovered and reported in the marketing literature on the traditional brick-and-mortar businesses, chiefly due to the lack of granular data on consumer's offline pass-by and visit behaviors. Our results suggest that when designing marketing strategies, retailers should be aware of the “visit effect” and commit more efforts to increasing consumers' visit likelihood, especially when the volume of pass-by traffic is high and in-store conversion is effective.

Our discovery of the visit effect also sheds light on previous findings on price elasticity of purchase in existing literature. Without considering the effect of price promotion on consumers’ early-stage behaviors such as store visit in the offline conversion funnel, a model on consumers’ in-store purchase will tend to exaggerate the effect of price promotion at the purchase stage and lead to overestimated price elasticity of purchase. This is because what positive effect that price promotion has on consumers’ early-stage behavior is absorbed into its effect on their purchase decisions instead, thus inflating the purchase-stage effect of the price promotion.

To fully characterize consumers' complete decision-making process throughout the offline conversion funnel, we build a multi-stage model that encompasses consumers' pass-by, store visit and purchase decisions. This modeling framework enables us to estimate the impact of marketing strategies in each stage of a consumer's decision process that leads up to the completion of purchases. Based on this model, we use simulation studies to first demonstrate the multi-stage impact of marketing strategies as a result of the insights gained from the IoT data. We find that visit effect from price promotions may account for over 20% of the overall increase in category sales for the three product categories we consider. Furthermore, we show that exploiting the information in the IoT data also enables an offline business to design customized marketing strategies which are otherwise infeasible due to a lack of knowledge of individual consumers’ in-store response to the marketing mix. Our study not only deepens the understanding of the marketing mix effects in different stages of the conversion funnel, but also provides a useful set of tools for practitioners to design and implement marketing strategies customized for each store.

For future research, opportunities abound in the flourishing area related to using the IoT technologies to deepen understanding of consumers' behaviors and incorporating new information generated by IoT in business decision making. First, the IoT will generate data from tracking consumers' day-to-day journeys, shopping decisions and consumption patterns. These enormous amounts of data, which tend to be noisy, unbalanced and unstructured, will require researchers to develop new data integration and statistical methods for data analysis. Second, to fully exploit the business value of the IoT data, marketing researchers need to develop integrated models for consumers' traveling, purchasing and consuming behaviors. These models can further provide insights to facilitate the design of more sophisticated marketing strategies, which not only generate more revenue and profits for businesses, but also avail consumers the benefits of better services, lower transaction costs and greater value. Third, the IoT will facilitate better ways to test the effectiveness of marketing interventions in different stages of the consumption process. For example, marketing researchers can design novel field experiments using the IoT to minimize the experiment's intrusiveness and collect more precise data in order to discern the causal effects of various forms of marketing interventions. Lastly, the IoT can help integrate traditional marketing channels with modern channels which rely on the Internet, mobile devices and social networks, by facilitating information sharing among these channels and providing instant feedback based on consumer experience. We believe our study is the beginning of many promising directions in future research.

Chapter 3: Design Elements of Mobile Marketing Communications: Evaluating Effectiveness of App Push Notifications

3.1 BACKGROUND

In this chapter, we seek to discover what design elements contribute to the effectiveness of app push notifications, and quantify the causal impact of each design element on user engagement. Existing research has examined how design elements on product web pages can impact the online customer experience (Bleier, Harmeling, and Palmatier 2019). In line with this approach, we propose a theory-based framework for analyzing the message information contained in push notifications and identifying the design elements that compose them. Our focus is on how these design elements influence user engagement with the app that is engendered by push notifications. To accomplish this, we analyze a unique data set of approximately 1600 push notifications sent by a prominent mobile coupon publishing app. Methodologically, we use the causal forest method (Wager and Athey 2018; Athey and Wager 2019; Athey, Tibshirani, and Wager 2019) to estimate the average treatment effects of the various design elements on users' opening of push notifications to measure the effectiveness of each design element.

Because design elements are not randomly assigned in the messages of different push notifications, we combine the causal forest method with linear regression to isolate the individual treatment effects of the design elements. This is achieved by first combining design elements into composite "treatments" and estimating the overall treatment effects over a chosen "control", and then exploiting the variation in design elements across different push notifications to identify the individual treatment effects of each element. The results of this analysis indicate that the inclusion of visual elements like emojis leads to increased effectiveness of push notifications. Additionally, the use of directive language that calls for action on the part of the user brings about significantly higher effectiveness of the push. While the inclusion of detailed price information decreases the open rate of the push, a higher discount percentage gives rise to a higher open rate, thus mitigating the negative effect of including price information. Furthermore, the mention of offer restrictions leads to a decrease in the effectiveness of push notifications.

This study contributes to the literature on mobile marketing by establishing a causal relationship between the design of message content in mobile push notifications and user engagement with mobile apps. Previous research on mobile marketing has focused on various utilitarian variables such as time, distance, discount depth, location, etc. as factors that may impact the effectiveness of a mobile marketing communication, while holding the design of the message content fixed. In contrast, this study approaches the entire message of a mobile marketing communication as an "artifact" that is composed of various movable parts, all of which can be modified and pieced together in different combinations to achieve different levels of effectiveness from the push notification. Under this light, mobile marketing communications are not simply delivery vehicles for conveying information to the intended audience, but rather become a flexible tool which can be easily modified by marketers to improve their effectiveness in engaging customers.

This study also contributes to the literature on marketing message design by demonstrating the importance of considering design elements in the effectiveness of mobile marketing messages. Companies use various forms of marketing messages to communicate to their customers about their brand, product, and service, and well-designed marketing messages are more likely to be well received by their intended audience. Extant research has examined the design of marketing messages in a variety of contexts, including product web pages, online service communications, and social media posts. Our study extends this research by showing that the effectiveness of mobile marketing communications can also be understood in terms of their composing elements from a design perspective.

Additionally, our findings provide practical guidance for marketers on the selection of effective marketing message design elements to help them achieve their desired customer engagement goals.

3.2 LITERATURE

3.2.1 Literature Review

This research lies at the intersection of and adds to two streams of research: mobile marketing and marketing message design.

The mobile marketing literature has largely focused on the influence of utilitarian and contextual factors on consumer engagement with mobile marketing communications. Temporal and geographical distances of the mobile marketing communications both impact the redemption rate of mobile coupon offers sent via text messages (Danaher et al. 2015; Luo et al. 2014). The time at which mobile marketing offers are delivered also plays an important role in their redemption rate (Danaher et al. 2015). Physical crowdedness has been found to increase consumers' conversion rate to targeted mobile ads they received while on subway trains (Andrews et al. 2016). Competitive geographically targeted mobile coupons sent to consumers at a competitor's location are shown to bring about profit gains for the focal business (Fong, Fang, and Luo 2015; Dube et al. 2017). Targeted promotions sent by push notifications on a mobile e-book app have been found to increase sales of books in the same genre as the promoted book, but decrease sales for those in the nontargeted genres (Fong et al. 2018).

In the marketing message design literature, researchers have examined the effectiveness of verbal, visual, and emotional elements in generating user engagement in social media and website contexts. Previous research has found that both verbal and visual design elements of a product web page can influence purchase by affecting the customer experience (Bleier, Harmeling, and Palmatier, 2019). The use of emoticons by service employees in online service communications has been found to be associated with increased warmth and decreased competence (Li, Chan, and Kim, 2019). Emotional, humorous, and directly informative content in companies' posts on Facebook Pages has been linked to higher levels of user engagement (Lee, Hosanagar, and Nair, 2017). The scheduling attributes of a company's Facebook posts, such as the time of day, content type, and targeted advertising, have been found to influence the number of consumer clicks on the post link (Kanuri, Chen, and Sridhar, 2018). Additionally, speech acts, image acts, rhetoric style, and cross-message dynamics have been shown to affect consumer sharing of brand messages on Twitter and Facebook (Ordenes et al., 2019).

We draw upon speech act theory as the theoretical foundation for constructing and coding the textual design elements. According to speech act theory (Searle, 1969), any communication is an implicit action intended to elicit a certain behavior in the recipient. Because a mobile marketing communication holds the promise of encouraging its recipients to take certain downstream actions with respect to the marketing offers contained in the communication, the speech act theory is well-suited to be the foundation on which we build and extend our conceptual framework. In fact, previous research in marketing has also built on speech act theory to examine the drivers of consumer sharing of brand messages on social media (Ordenes et al., 2019). Three forms of the speech act taxonomy are relevant in our empirical context: assertive, expressive/affective, and directive. Thus, we encode the textual information according to these categories.

3.2.2 Variables and definitions

In this study, we examine the textual, visual, and emotional elements of mobile marketing communications. A total of five design categories and eleven design elements are identified for analysis. Table 3.1 below presents a breakdown of these categories and the definition for each design element.

Design Category	Design Variable	Variable Definition
Compositional	Message Length	Length of the push message in number of words
Visual	Number of Emojis	Number of emojis used in the message
	Face Emoji	Indicator for the presence of face emoji(s)
Affective	Emotional Valence	Percentage of words in the message that are emotionally positive according to LIWC
Assertive	Price Information	Indicator for whether price information is present in the message
	Discount Amount	Discount offered in dollar amount terms
	Discount Percentage	Discount offered in percentage terms
	Restrictions	Indicator for when there are any conditions or restrictions to redeeming the offer
	Multiple Benefits	Indicator for when the message mentions multiple benefits in the offering
Directive	Urgency	Indicator for when the message highlights the time urgency of the offer (e.g. "Today only!")
	Call to Action	Indicator for when the message uses language that calls for user's taking action (e.g. "Swipe right to take a look.")

Table 3.1: Definition of design elements

The design categories chosen include the compositional, visual, affective, assertive, and directive aspects of the push message content. Since push notifications have limited display space on the screen of a mobile device, the compositional feature of the message can have a direct effect on how it will be received by the user. Visual elements have been shown to be effective components of marketing communications that boost user response (Bleier, Harmeling, and Palmatier, 2019; Li, Chan, and Kim, 2019; Ordenes et al., 2019; McShane et al., 2021), hence we include this dimension in considering design elements of push notifications. The affective or emotional content of a marketing communication has been a near-constant factor of examination in marketing research (Berger and Milkman 2012; Lee, Hosanagar, and Nair, 2017), thus we also consider its role in influencing the effectiveness of push notifications. The informative content of push notifications provides straightforward information that can influence the user's decision making, so it also takes a place among the dimension of design elements to be considered. Finally, usage of language that either implies the need for or directly calls for taking action in marketing communications may also generate favorable response from the recipient (Danaher et al. 2015; Ordenes et al., 2019).

3.3 Empirical Context and Data

The data used in this study was obtained from a major mobile coupon publisher that operates a mobile app with a large user base on both the iOS and Android platforms. The publisher works with retailers to send targeted push notifications to selected users, with each push notification featuring a specific retailer carrying out a marketing campaign. These retailers are well-known brands that span a wide range of product categories, including clothing, cosmetics, food, electronics, personal care, sporting goods and more. We focus our analysis on five product categories: clothing, cosmetics, department store, electronics, and food. These categories were chosen because they are the most frequently featured and use relatively similar language in their corresponding push messages. It is worth noting that while department stores often have a clothing section, not all department store campaigns exclusively relate to clothing offers. Only department-store campaigns which make it clear that the department store is exclusively promoting clothing offers are classified in the clothing category.

3.1.1 Message data

The final dataset for this study contains push notifications sent by the publisher between 2015 and 2018, from a total of 1,589 push campaigns. Table 3.2 below provides examples of messages used in some of the push notifications, along with their sponsoring retailers.

Retailer Name	Push Message
Best Buy	Up to 50% off Best Buy clearance, open-box & more!
Sonic	🌭 Swipe To See How To Receive Your Exclusive $1 6" Hot Dog Offer From SONIC. Today Only!
Reebok	👟 40% Off Reebok Running Footwear & Apparel!
Michael Kors	👜 ❤️ Up to $200 Off Michael Kors Valentine's Day Deal!
Ulta	💄 Buy 2 Mix & Match Items, Get 2 FREE at ULTA!
Jamba Juice	🍌🍓 $2 Off Any Smoothie, Juice, Energy Bowl or Oatmeal at Jamba Juice!

Table 3.2: Examples of message

The publisher often includes emojis at the beginning of push notifications, which serve as a brief and interesting visual addition to the plain textual information in the message. In contrast, the textual content of push notification can include detailed information regarding the retailer’s offer, including the discount amount, discount percentage, any restrictions to offer redemption, and so on. A user can only receive one push notification every 24 hours, and given the publisher's large user base, the average number of users targeted by each push is substantial. Therefore, we select a random sample of 10,000 users targeted by each push notification for our empirical analysis.

3.3.2 Design elements summary statistics

Table 3.3 provides the summary statistics for the design elements of push notifications. On average, a push notification message is around 9 words in length. 1.2 emojis are used in push notifications, while only 6% of push notifications have a face emoji in them. The average emotional valence of the push notifications is 10.29, which is the result of the message containing emotionally positive words as identified by LIWC. 90% of the push notifications include detailed price information, which specifies the discount offered either in dollar amount or in percentage terms. 37% of the push notifications include mentions of offer restrictions, while 12% of them presented information about multiple benefits contained in the offer. Lastly, 9% of the push notifications highlight the time urgency of the offer, and 3% of them use language that calls the users to take certain actions with the app.

Design Category	Design Element	Mean	Standard Deviation
Compositional	Message Length	8.98	2.96
Visual	Number of Emojis	1.2	0.8
	Face Emoji	0.06	0.23
Affective	Emotional Valence	10.29	4.01
Assertive	Price Information	0.9	0.3
	Discount Amount	29.35	51.89
	Discount Percentage	34.04	18.97
	Restrictions	0.37	0.48
	Multiple Benefits	0.12	0.33
Directive	Urgency	0.09	0.28
	Call to Action	0.03	0.16

Table 3.3: Design elements summary statistics

3.3.3 Model free analysis

To examine the relationships between design elements and user engagement, we present visual evidence of their correlations at the campaign level.

Figure 3.1 plots the average open rate against the message length of push notifications by category. While there does not seem to be a linear correlation between the average open rate and the message length in either direction for the clothing and cosmetics categories, it appears that a somewhat positive linear relationship exists between them in both the department store and electronics categories. In the food category, however, a slightly negative linear correlation seems to be trending between the average open rate and the message length.

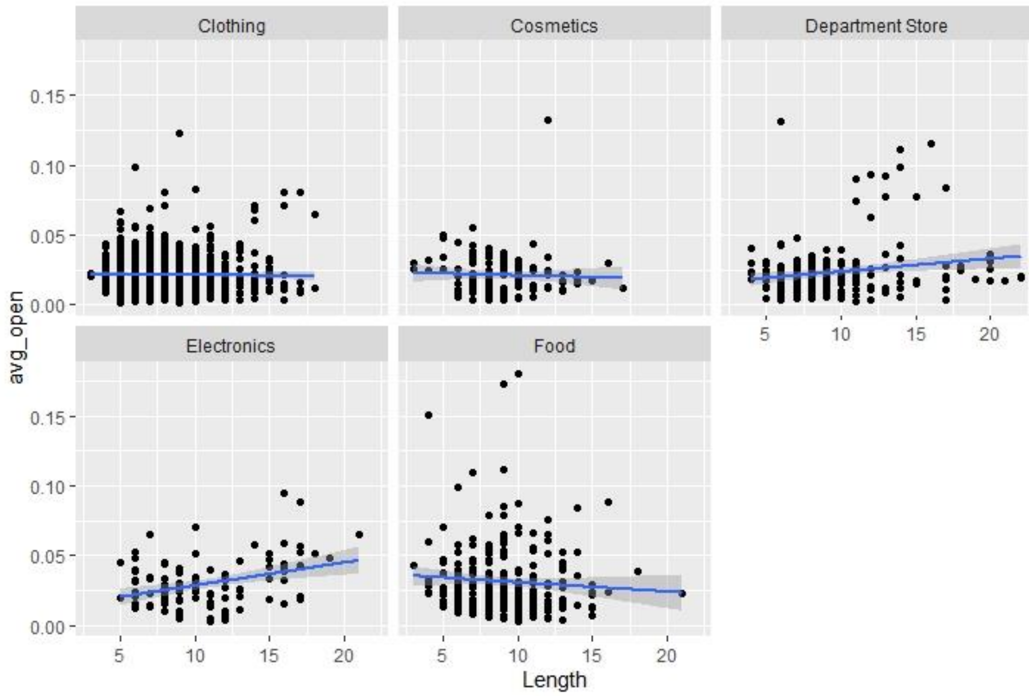

Figure 3.1: Average open rate against message length by category

For visual elements, we examine the boxplots of the average open rate at different numbers of emojis, and for whether a face emoji is present in the push in Figure 3.2. We can see that as a push notification starts to include emojis, there seems to be an initial slight dip in the average open rate. However, as the number of emojis increases from 1 to 3 (maximum number of emojis observed in the data), the average open rate also increases. On the other hand, the average open rate stays constant regardless of whether a face emoji is present in the push.

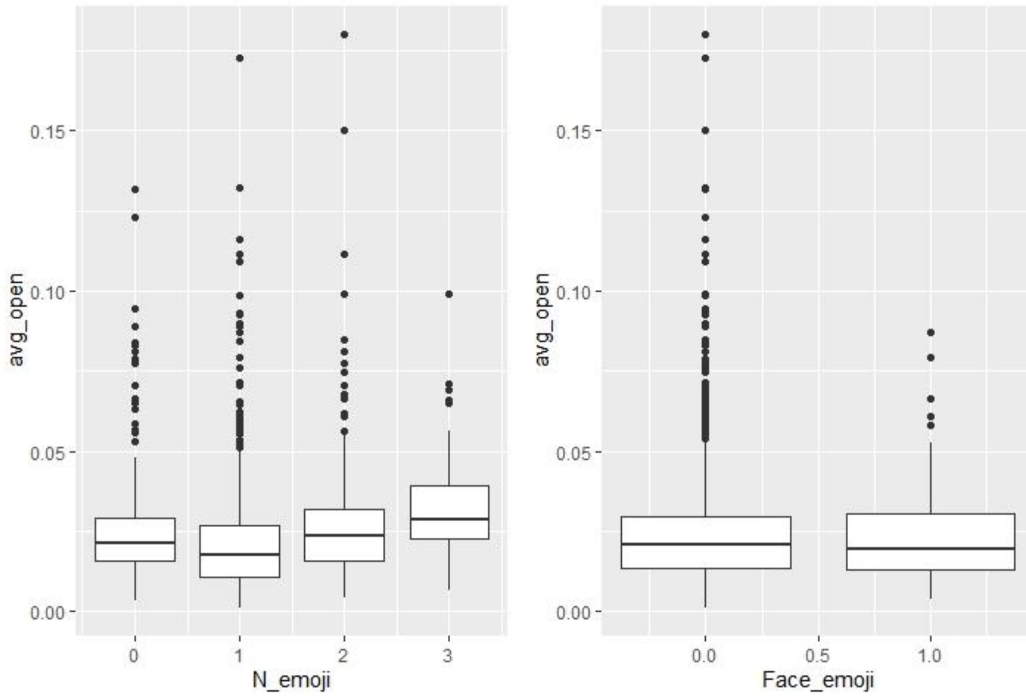

Figure 3.2: Average open rate against visual elements

Figure 3.3 presents the relationship between the average open rate and the emotional valence of the push notification by category. A positive linear correlation emerges between the average open rate and the emotional valence in the clothing and food categories, while it appears that a negative linear correlation exists between the two in the department store and electronics categories. As for the cosmetics category, although the range of values for emotional valence is larger due to a single campaign, a slightly negative linear correlation seems to exist between the average open rate and the emotional valence.

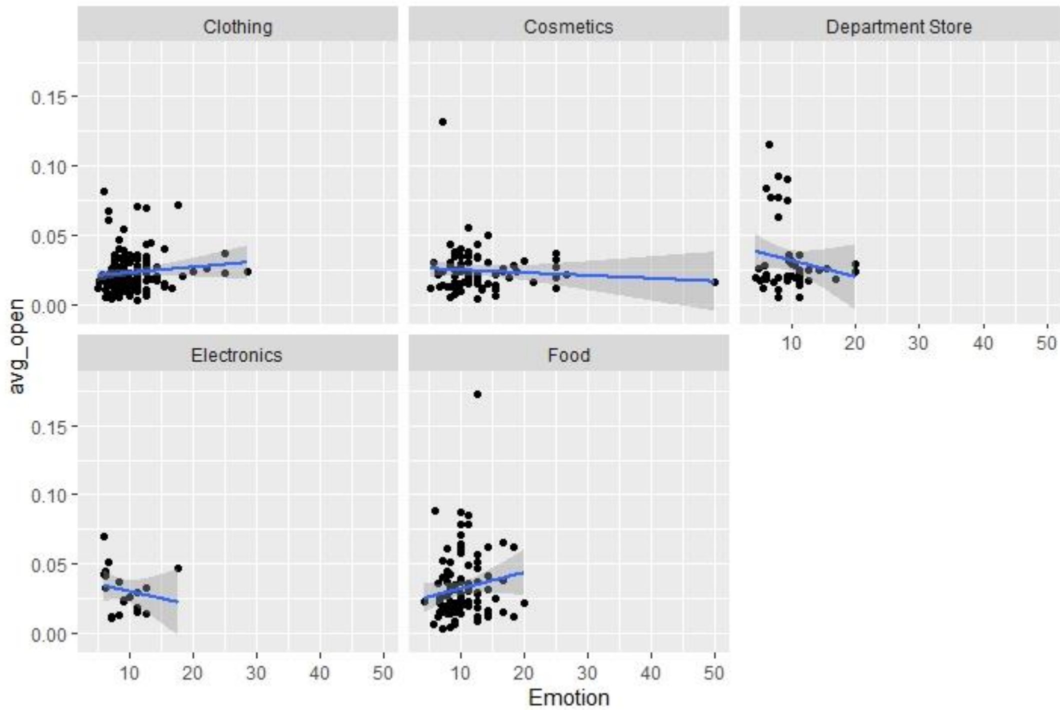

Figure 3.3: Average open rate against emotional valence by category

We then turn to examine the relationships between the average open rate and the assertive design elements (presence of price information, discount amount, discount percentage, restrictions, multiple benefits) of push notifications in Figure 3.4. Starting with price information, it appears that the mean of the average open rate for push notifications is lower when price information is included in the message. However, there seem to be positive linear correlations between the average open rate and both the discount amount as well as the discount percentage variables. When a push notification includes offer restrictions, the mean of the average open rate appears to be lower than when no restriction is involved. Lastly, the mean of the average open rate does not seem to be influenced by the inclusion of multiple benefits.

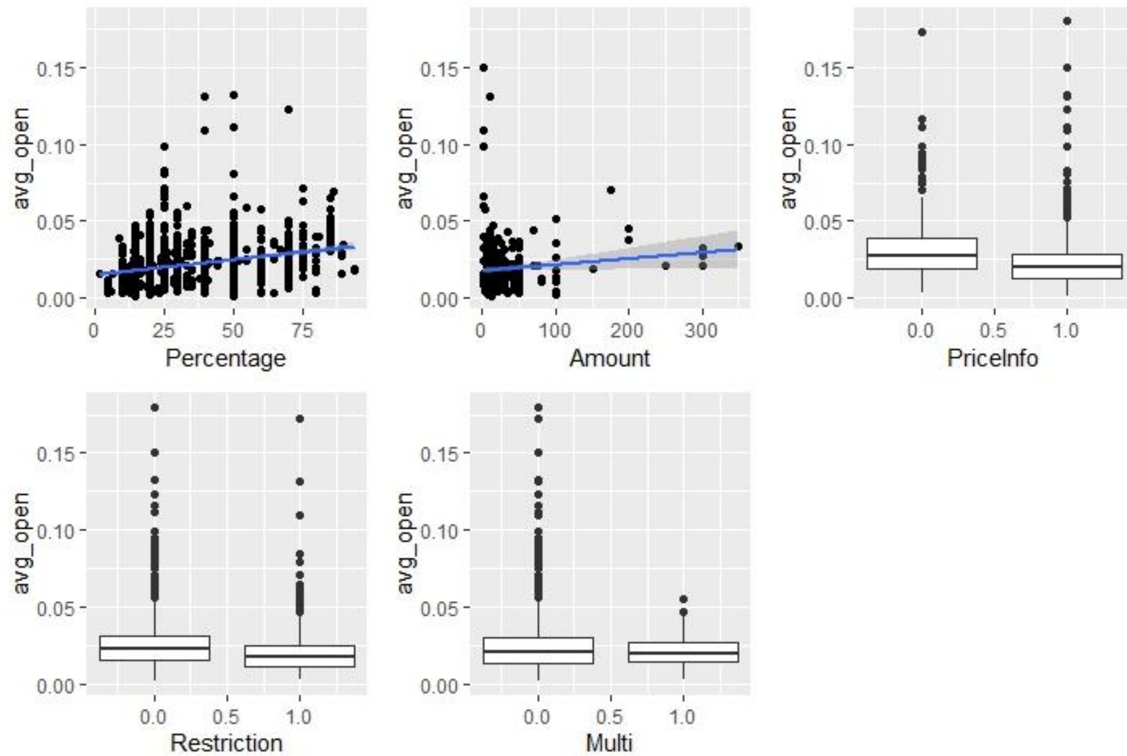

Figure 3.4: Average open rate against assertive elements

Finally, we explore the relationship between the average open rate and the directive design elements of push notifications in Figure 3.5. If the push notification directly calls the user to take actions with the app, the mean of the average open rate is significantly higher than when no such call-to-action is issued. When the urgency or time-sensitivity of the offer is highlighted, the mean of the average open rate also appears to be higher than otherwise.

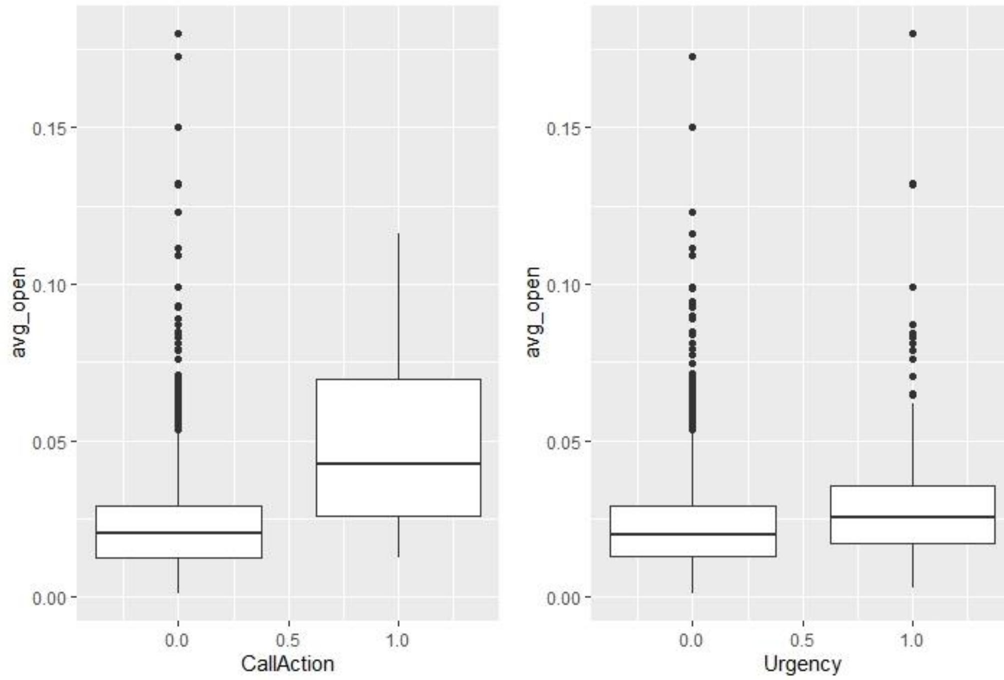

Figure 3.5: Average open rate against directive elements

3.3.4 User Characteristics

Each push campaign is targeted at a group of users who differ in their individual characteristics, and these characteristics can include the user's historical behaviors with the app as well as both time-dependent and time-invariant user attributes. We observe five historical behavioral characteristics for users in our data: the number of push notifications received, the number of push notifications opened, the number of outclicks made, the number of push notifications saved, the time gap since the user's last received push (in days). The user's account age (tenure) on the app (in days) is a time-dependent user attribute, nd we also observe the user's mobile operating platform (android vs ios) as a time-invariant attribute. Marketers have long recognized the importance of observed customer heterogeneity in moderating the effect of marketing interventions (Gupta and Chintagunta 1994; Rossi, McCullough, and Allenby 1996; Ascarza 2018). In our context, we assume that observed user characteristics also moderate the effect of design elements of push notifications on user engagement.

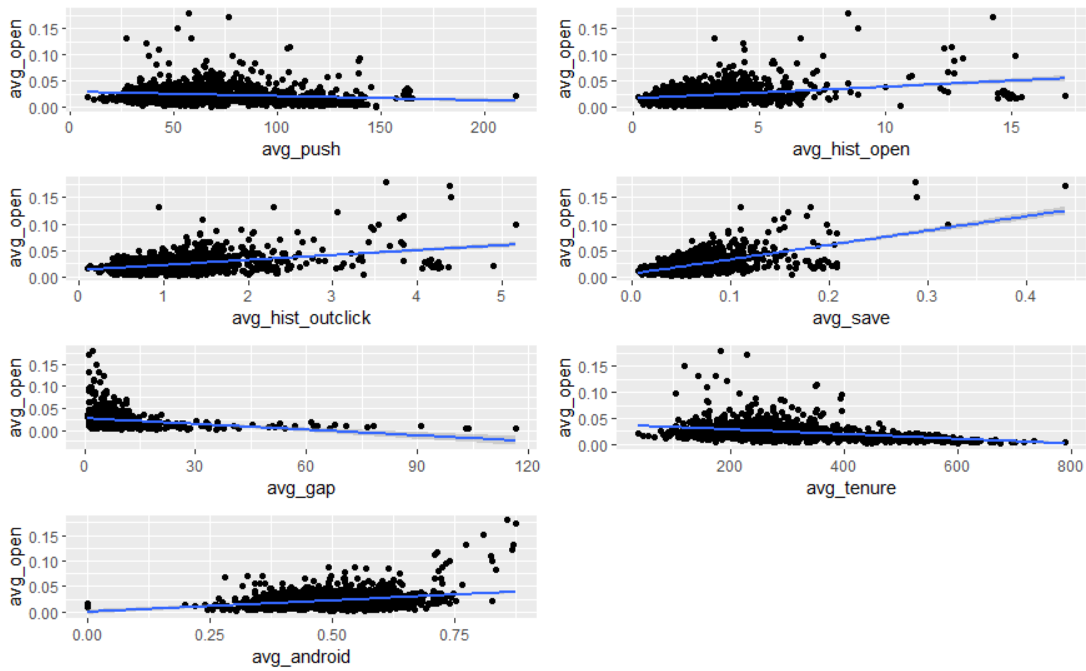

Figure 3.6: Average open rate against observed user heterogeneity

We present visual evidence for the relationships between average user characteristics and average open rate at the campaign level in Figure 3.6. For the average number of pushes received, there does not seem to be a clear linear correlation with the average open rate in either direction. The average number of pushes opened, the average number of outclicks made, and the average number of pushes saved, all measure a user's prior engagement with the app in some way. We can see that the average open rate exhibits positive linear correlations with the average number of pushes opened, the average number of ouclickes made, and the average number of pushes saved, where the positive correlation with the average number of pushes saved is most prominent. This suggests that if a user has engaged with the app from push notifications more often in the past, she is also more likely to engage again when receiving a new push. The average open rate appears to decrease as the average gap since last received push increases, which implies that the longer an average user goes without receiving a push, the less likely she is to engage with a new push by opening it. Furthermore, there does not seem to exist any linear correlation between the average open rate and the average tenure of users receiving the push. Finally, the average open rate seems to positively correlate with the proportion of Android users receiving the push, which could be explained by the fact Android apps have a higher push notification open rate than their iOS counterparts in general.

3.3.5 Aggregate regression analysis

To investigate the relationship between user open rate of push notifications and design elements in the push message, while controlling for observed user characteristics, we perform a regression analysis at the aggregated-campaign level. In particular, we estimate the following specification:

$$AvgOpenRate_j = \alpha + \beta_1 MessageLength_j + \beta_2 NumEmoji_j + \beta_3 FaceEmoji_j + \beta_4 EmotionalValence_j + \beta_5 PriceInfo_j + \beta_6 DiscountAmount_j + \beta_7 DiscountPercentage_j + \beta_8 Restrictions_j + \beta_9 MultipleBenefits_j + \beta_{10} CallToAction_j + \beta_{11} Urgency_j + \beta_{12} AvgPushReceived_j + \beta_{13} AvgPushOpened_j + \beta_{14} AvgOutclicks_j + \beta_{15} AvgPushSaved_j + \beta_{16} AvgGap_j + \beta_{17} AvgTenure_j + \beta_{18} AvgPlatform_j + \delta D_{j,Category} + \epsilon_j$$

where $AvgOpenRate_j$ is the average open rate of push j; $MessageLength_j$, $NumEmoji_j$, $FaceEmoji_j$, $Emotion_j$, $PriceInfo_j$, $DiscountAmount_j$, $DiscountPercentage_j$, $Restrictions_j$, $MultipleBenefits_j$, $CallToAction_j$, $Urgency_j$ are the 11 design elements for the corresponding push message; $AvgPushReceived_j$, $AvgPushOpened_j$, $AvgOutclicks_j$, $AvgPushSaved_j$, $AvgGap_j$, $AvgTenure_j$, $AvgPlatform_j$ are the average observed user characteristics for users targeted by push j; $D_{j,Category}$ is a set of indicator variables for the product category that campaign j is in, and ϵ_j is a normally distributed error term.

Table 3.4 reports the results from the aggregate regression. The (unreported) intercept of 0.02171 is statistically significant, which suggests that the baseline average open rate is 2.171% across all push campaigns. Both the number of emojis and the urgency indicator are significantly positively correlated with the average open rate. The estimated coefficient of 0.00114 for the number of emojis suggests that as the push campaign included one extra emoji (up to a maximum of 3), the average open rate of the push campaign is increased by 0.114%. If the push campaign included language that highlights the urgency or time sensitivity of the offer, the average open rate of the campaign increases by 0.353%. In contrast, the estimated coefficients for the price information, multiple benefits, and call to action indicators are all statistically significantly negative. Specifically, specifying price information in a push message is associated with a 0.425% decrease in the average open rate. Including multiple benefits in the push campaign is associated with a 0.197% decrease in the campaign's average open rate. Furthermore, including language that directly calls for user's taking action with the app is associated with a 1.07% decrease in the average open rate.

Variable Category	Independent Variable	Estimate	Standard Error
Design Elements	Message Length	-1.29×10^{-4}	1.19×10^{-4}
	Number of Emojis	$\mathbf{1.14\times10^{-3}}$***	3.8×10^{-4}
	Face Emoji	-8.35×10^{-4}	1.21×10^{-3}
	Emotional Valence	-5.20×10^{-5}	7.19×10^{-5}
	Price Information	$\mathbf{-4.25\times10^{-3}}$***	1.19×10^{-3}
	Discount Amount	1.14×10^{-5}	1.31×10^{-5}
	Discount Percentage	3.78×10^{-5}	1.60×10^{-5}
	Restrictions	-3.89×10^{-4}	6.21×10^{-4}
	Multiple Benefits	$\mathbf{-1.97\times10^{-3}}$***	9.74×10^{-4}
	Urgency	$\mathbf{3.53\times10^{-3}}$***	9.40×10^{-4}
	Call to Action	$\mathbf{-1.07\times10^{-2}}$***	3.27×10^{-3}
User Characteristics	AvgPushReceived	$\mathbf{-5.45\times10^{-4}}$***	3.71×10^{-5}
	AvgPushOpened	$\mathbf{2.66\times10^{-3}}$***	6.90×10^{-4}
	AvgOutclicks	3.90×10^{-3}	2.27×10^{-3}
	AvgPushSaved	$\mathbf{2.38\times10^{-1}}$***	2.27×10^{-2}
	AvgGap	$\mathbf{-2.34\times10^{-4}}$***	3.63×10^{-5}
	AvgTenure	$\mathbf{2.13\times10^{-5}}$***	6.89×10^{-6}
	AvgPlatform	-3.67×10^{-3}	3.51×10^{-3}
Category Fixed Effects	Cosmetics	1.75×10^{-3}	1.19×10^{-3}
	Department Store	$\mathbf{-2.13\times10^{-3}}$*	9.80×10^{-4}
	Electronics	-6.40×10^{-4}	1.31×10^{-3}
	Food	1.44×10^{-3}	1.00×10^{-3}

*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

Table 3.4: Aggregate regression analysis results

Among the observed user characteristics, the estimated coefficients are statistically significantly negative for AvgPushReceived and AvgGap. In particular, the estimated coefficient of -0.00055 on AvgPushReceived suggests that each additional push the average user has received is associated with a 0.055% decrease in the user's average open rate of the current push. While the estimated coefficient of -0.00023 on AvgGap implies that each additional day the average user has not received a push is associated with a 0.023% decrease in the user's average open rate of the current push. Conversely, the estimated coefficients for AvgPushOpened and AvgPushSaved are statistically significantly positive. Specifically, the estimated coefficient of 0.00266 for AvgPushOpened suggests that, as the average user has opened one extra push in the past, there is a 0.266% increase the user's open rate of the current push.[11] The estimated coefficient of 0.238 on AvgPushSaved suggests that each additional push the average user has saved in the past is associated with a 23.8% increase in the user's average open rate of the current push. Finally, each additional day in the average user’s account age is associated with a 0.002% increase in the average open rate of the current push .

With the clothing category as baseline, we see that only the coefficient for the department store category indicator is statistically significantly negative. The estimate of -0.00213 on the department store indicator suggests that, on average, push campaigns in the department store category are associated with a 0.213% lower average open rate compared to campaigns in the clothing category.

[11] It may seem surprising that the coefficient on AvgOutclicks is statistically insignificant given the visual evidence presented above, however a check of correlation reveals that the linear correlation between AvgPushOpened and AvgOutclicks is 0.97, thus leading to only one of the variables picking up their common correlation with AvgOpenRate in the regression analysis due to strong multi-colinearity (in this case AvgPushOpened).

These results provide statistical evidence that user open rate of a push notification is correlated with both the design elements in the push notification message, as well as observed user characteristics. However, it is important to note that these results reflect only the statistical correlations between user engagement and message design elements of push notifications at the aggregate level. To more fully understand the causal relationship between design elements and user engagement, a different empirical approach is needed. This alternative strategy aims to isolate the causal effects of message design elements on push open rate, while simultaneously taking into account the potential moderating effects of observed user characteristics.

3.4 Research Design and Conceptual Framework

3.4.1 Conceptual framework

Our objective is to understand the causal effects of the different design elements on user opening of push notifications. To achieve this objective, we propose the following conceptual framework in Figure 3.7 to delineate the causal inference problem in our context.

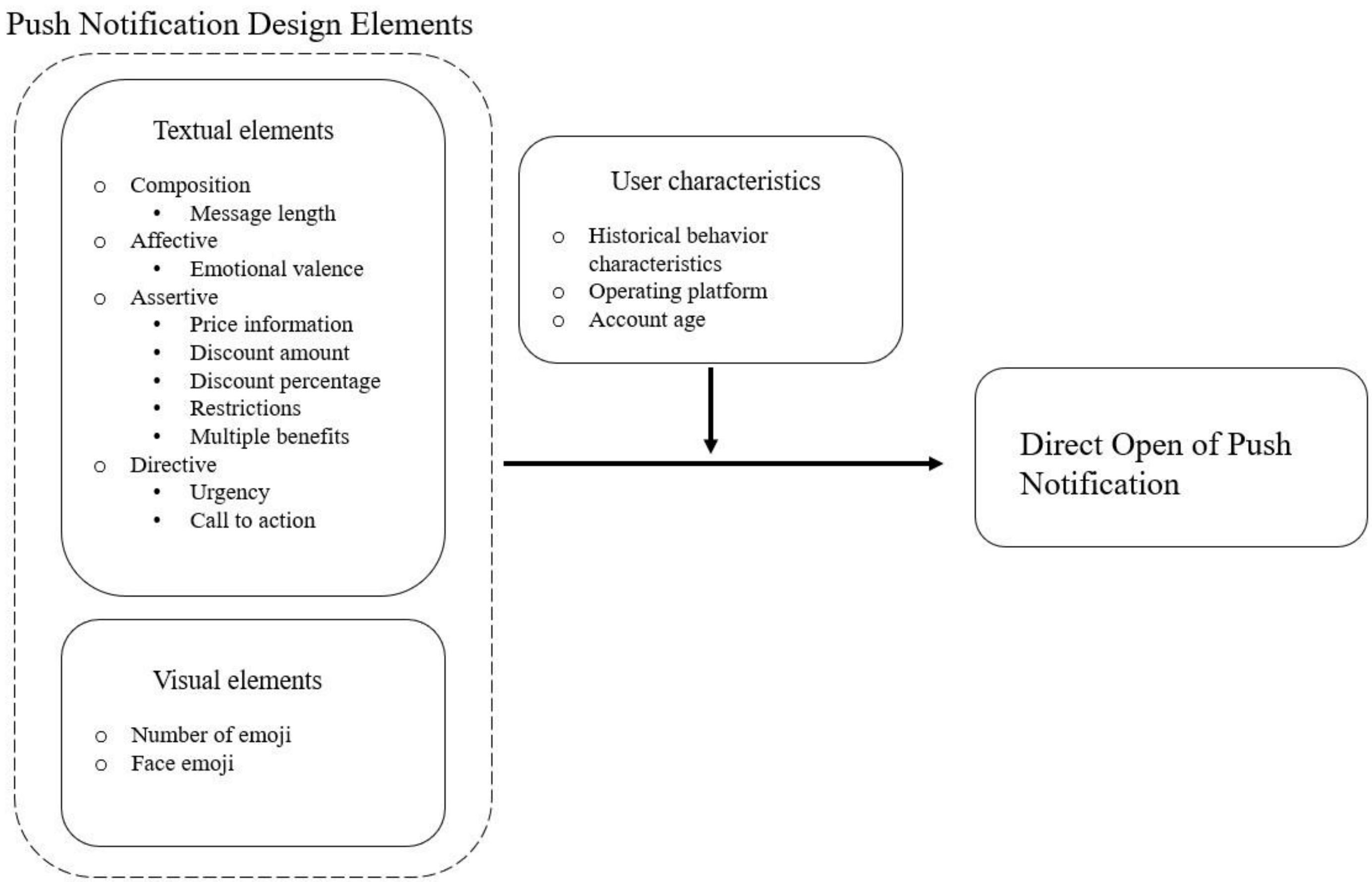

Figure 3.7: Conceptual framework

As discussed earlier, the message design elements encompass both textual and visual components. In accordance with speech act theory, the textual elements are further divided into three subcategories: assertive, affective, and directive. Additionally, the compositional feature of the textual elements, in terms of message length in words, is also considered due to the limited display space for push notifications on mobile devices. Furthermore, the utilization of emojis constitutes a visual component of the message, which is also a source of variation in the message design. We theorize that the variations in textual and visual design elements cause the variation in user engagement in terms of push open rate, and this causal influence is moderated by observed user characteristics.

3.4.2 Causal inference of design elements

As marketers, we are particularly interested in the incremental changes in customer response metrics, such as open rate, that are caused by changes in marketing variables. The reason is that marketing variables are directly under the marketer’s control and can be manipulated to varying degrees. In other words, if the causal impact of a marketing intervention can be understood, it then becomes feasible for the marketer to effectively leverage these marketing variables to achieve more desirable customer response. In the present empirical context, the use of design elements in push campaigns is posited as a type of marketing leverage, and by manipulating this lever marketers can motivate more user engagement with the help of a well-designed push notification. This kind of causal inference is possible in our empirical context, because user engagement outcomes for push notifications are observed at the individual user level.

In identifying the causal effects of design elements on user engagement with push notifications, we encounter three major challenges. First, each push campaign is targeted at a specific group of users by the publisher, and campaigns that are similar in nature (e.g. clothing campaigns offering percentage discounts) also tend to use similar design elements in their messages. As a result, the design elements are not randomly assigned to targeted users across different push campaigns. Second, the push notification for each push campaign includes multiple design elements that appear in different perturbations of their values (for continuous variables) or levels (for discrete variables), thus making it infeasible to disentangle the causal effect of each element using conventional causal inference methods on observational data (e.g. difference-in-differences). Third, as shown by the aggregate regression analysis, observed user characteristics are a part of the interplay that leads to the final user open behavior of a push campaign at the aggregate level. However, it is unclear how a user's individual characteristics interact with the design elements to generate the user's final open decision, therefore imposing *a priori* functional form for the effect of observed user characteristics on user engagement will lead to biases in the estimates for the causal effects of design elements.

We draw from a growing body of empirical research that utilizes machine learning techniques to study the causal effect of policies or interventions using observational or quasi-experimental data, and identify causal forest as the most appropriate tool for helping us overcome these challenges. This method will be adapted to our specific context of studying the causal effects of design elements on user engagement with push notifications.

3.4.3 Causal forest

In this section, we first present the mathematical formulation of the causal inference problem in the context of this study. We then provide a detailed description of the causal forest method, with a specific focus on its relevant desirable features that aid in addressing the aforementioned three challenges that we face in identifying the causal effects of design elements on user engagement with push notifications.

At the heart of the causal inference problem using observational data is the issue of non-observability of counterfactual outcomes: consider a user i in push campaign j whose user characteristics are captured by covariates X_i, and for the sake of argument suppose only one binary treatment W_i is administered (e.g. price information in the push message) to the user, so that the outcome (open) for user i given X_i and W_i is $Y_i(W_i|X_i)$, where we observe either $Y_i(W_i = 1|X_i)$ or $Y_i(W_i = 0|X_i)$ for this user but not both. Our goal is to find the average treatment effect defined by $\tau = E[Y_i(W_i = 1) - Y_i(W_i = 0)]$, with the conditional (on user characteristics) average treatment effect (CATE) given as $\tau(x) = E[Y_i(W_i = 1) - Y_i(W_i = 0)|X_i = x]$. It can be easily seen that the CATE is unobserved at the individual user level.

The standard approach in the literature proceeds by assuming unconfoundedness (Rosenbaum and Rubin, 1983; Athey and Imbens, 2017), that is, the treatment assignment conditional on controls X_i is independent from the potential outcomes:

$$\{Y_i(1), Y_i(0)\} \perp W_i | X_i$$

In other words, users with similar characteristics as captured by the X covariates can be considered as having come from a randomized experiment with respect to the treatment W. This assumption then allows for the recovery of CATE on users that are close in the X-space. Put another way, this means that users similar to each other in terms of their observed characteristics can be regarded as close copies of each other, with the only difference being that some of them have received a push message with treatment W and the others have received a control push instead.

Generalized random forest is a nonparametric tree-based and supervised machine learning method developed to estimate any quantity of interest $\theta(X)$ which can be identified via local moment conditions (Athey, Tibshirani, and Wager, 2019). One of its most popular applications in the recent empirical literature is to study the heterogeneous treatment effects of policy interventions using observational or (quasi-)experimental data (Athey and Wager, 2019; Guo, Sriram, and Manchanda, 2021). For this particular purpose, the generalized random forest is also known as the causal forest (Wager and Athey 2018; Athey and Wager 2019; Athey, Tibshirani, and Wager 2019), and its estimated treatment effects are shown to be pointwise consistent for the true treatment effects and asymptotically normal (Wager and Athey 2018). In this study, we implement the causal forest for causal inference on *multiple* and *concurrent* treatment variables, where we uncover the overall average treatment effects of each treatment variable (marginalized over user characteristics).

The causal forest method accomplishes the estimation for the causal effect of treatment W from a given set of users S in three major building blocks:

1. It first grows single causal trees by making axis-aligned partitions of users into subgroups or leaves, L, based on the covariates X (e.g. $X_1 < c$), where users who received the treatment and those who didn't will both be included in each leaf and have similar propensities for receiving the treatment. The partition is performed based on a revised mean-squared error criterion (Athey and Imbens 2016), until each leaf reaches a minimal size of k observations for either of treated and untreated users. As a result, the leaves of the tree will be larger (containing more user observations) in the X-space where there are lower degrees of observed user heterogeneity, and smaller in the X-space where higher degrees of observed user heterogeneity are present. This tree-based method has the advantage that the researcher does not need to specify *a priori* structures by which the covariates moderate the causal effects of the treatment, hence the algorithm implicitly allows for capturing complex and nonlinear relationships between the covariates and the treatment effect, which in turn leads to a reduction in bias and increase in power over other grouping methods such as propensity score matching which uses a pre-specified functional form.
2. After an individual causal tree is grown and its partitions made, the estimation of treatment effects on W is carried out within each leaf L from calculating the

differences between the average outcomes of the treated and those of the untreated users:

$$\hat{\tau}(x) = \frac{\sum_{\{i:X_i \in L, W_i=1\}} Y_i(W_i = 1)}{|\{i: X_i \in L, W_i = 1\}|} - \frac{\sum_{\{i:X_i \in L, W_i=0\}} Y_i(W_i = 0)}{|\{i: X_i \in L, W_i = 0\}|},$$

which can be viewed as the sample analogue of the CATE $\tau(x) = E[Y_i(W_i = 1) - Y_i(W_i = 0)|X_i = x]$. The causal tree also follows an honesty principle (Athey and Imbens 2016; Wager and Athey 2018), in the sense that the training sample is randomly subdivided into two halves for each tree: one half of the sample is used to grow the tree and make partitions, and the other half is used to carry out the within-leaf treatment effect estimations. In this way, every observation in the training sample can be used for either tree growing or within-leaf estimation but not both. This is because, when the causal tree is grown, extreme values of Y_i are likely to be placed into the same leaf as other extreme values in a partition, hence the within-leaf estimation of the treatment effect containing those extreme values will also lead to more extreme estimates than would have been otherwise, which gives rise to biased estimates of treatment effects. By using a different subsample for estimation, this potential bias is alleviated since the estimation sample will contain independent observations that are on average less extreme than those used in the tree growing sample.

3. Single causal trees are then bootstrap aggregated ("bagged") into an ensemble of B trees, or a causal forest, with indices $b = 1, \ldots, B$. In growing and estimating treatment effects in each single causal tree, a random subset of the training data $S_b \in S$ is used. Furthermore, each new split considered for a tree will use a random subset of the X variables, which is a common practice in random forest to de-correlate the trees and reduce overfitting. A straightforward approach to aggregate estimates from the causal trees is by taking an average over all B trees: $\hat{\tau}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{\tau}_B(x)$. In practice, the causal forest is cast as a type of adaptive locally-weighted nearest neighbor estimator, and it obtains a weighted average estimate of treatment effects across trees using orthogonalized treatment effect and propensity estimates to ensure accurate treatment effect estimates in observational studies (Robinson, 1988; Athey and Wager 2019; Athey, Tibshirani, and Wager 2019). Concretely, the bagged CATE estimate is given by:

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{N} \alpha_i(x) \left(Y_i - \hat{m}^{(-i)}(X_i)\right)\left(W_i - \hat{e}^{(-i)}(X_i)\right)}{\sum_{i=1}^{N}\left(W_i - \hat{e}^{(-i)}(X_i)\right)},$$

where $\alpha_i(x)$ is the data-adaptive weight metric that measures how similar the i-th user is to the focal user with characteristics x. In other words, $\alpha_i(x)$ captures how often the i-th user falls into the same leaf as x over B trees: $\alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \frac{I(X_i \in L_b(x), i \in S_b)}{|\{X_i \in L_b(x), i \in S_b\}|}$, where $L_b(x)$ denotes the leaf containing the user with characteristics x in the causal tree grown in step b. $m(x) = E[Y_i | X_i = x]$ is defined as the marginal outcome at x (marginalized over treatment status), and $e(x) = E[W_i | X_i = x]$ the propensity score at x. $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$ are the "out-of-bag" estimates for $m(x)$ and $e(x)$ respectively, which are formed by only considering the trees b for which $i \notin S_b$.

Finally, the overall average treatment effect (unconditional on x) can be obtained by averaging estimated CATE across all training examples. However, Chernozhukov et al. (2018) show that a more accurate estimate can be obtained by plugging the causal forest predictions into a doubly robust average treatment effect estimator, and the causal forest method implements this approach to recover the average treatment effect.

3.4.4 Causal forest analysis and results

We adapt and implement the causal forest method to study the treatment effects of the eleven proposed design elements on user engagement with push notifications below.

We first choose a push campaign as the control, which is accomplished by identifying a push with minimal numbers of design elements each of which attaining minimal values in the message. This chosen campaign is labeled as the control campaign c_0 and its users as control users.[12] Denoting the set of design elements as d, we can write the design elements for the control campaign as d_0. In particular, we chose the campaign Abercrombie100615 with the corresponding message "Up to 30% off Abercrombie Sale".

[12] In theory, an ideal control campaign will have a blank push message that contains no design elements. Since such a campaign doesn't exist in practice in our data, we instead seek a minimal-design push campaign as the basis for comparison.

This choice of control campaign corresponds to the baseline design elements given by $d_0 = (MessageLength = 5, NumEmoji = 0, FaceEmoji = 0, EmotionalValence = 0, PriceInfo = 1, DiscountAmount = 0, DiscountPercentage = 30, Restrictions = 0, MultipleBenefits = 0, CallToAction = 0, Urgency = 0)$.

Then, for all other push campaigns $c = 1, \ldots, C$ and $c \neq c_0$ that also differ in design elements from the control campaign, we execute the following procedure until all treatment-control campaign pairs have been iterated over:

1. Take campaign c as the treatment campaign, and its users as treatment users. Pool data (Y, X, W) from both the control campaign and the treatment campaign, where Y is the outcome indicator for whether the user opened the push, X is the set of observed user characteristics, and W is the treatment indicator, with $W = 1$ for users in the treatment campaign and $W = 0$ otherwise. Take the difference in design elements between the treatment campaign and the control campaign, $\Delta d_c = d_c - d_{c_0}$, and save this information for all treatment-control campaign pairs. We then train a causal forest on the pooled data (Y, X, W) to obtain the average treatment effect, $\tau_c = CF(Y, X, W)$, which is entirely absorbed by the treatment indicator W. Essentially, the average treatment effect τ_c packs the effects of all design elements into one aggregate estimate for each treatment campaign considered.
2. Once the average treatment effects are estimated for all treatment-control campaign pairs (c, c_0), we then estimate the linear regression: $\tau_c = \alpha + \delta \Delta d_c + \varepsilon$

where the coefficient δ captures the individual treatment effects of each design element, and ε is a normally distributed error term. This step allows for the unpacking of treatment effects by projecting the aggregated average treatment effects onto the space spanned by the differenced design element vectors Δd_c.

To recap, the goal of the causal forest analysis is to uncover the treatment effects of individual design elements on users' opening of push notifications. To accomplish this goal, we fix a push campaign with minimal design elements in its message as the control, then regard all other push campaigns with different design elements as treatment. Once the aggregate treatment effects have been estimated for all treatment-control campaign pairs using causal forest, they are then projected onto the differenced design elements for all treatment-control pairs. Thus, in the subsequent discussions, the interpretation of the unpacked treatment effects for each design element rests on the difference between the level of that design element in the treatment campaign and that of it in the control campaign.

Table 3.5 presents the results from the causal forest analysis from executing the two steps delineated above.

Variable Category	Independent Variable	CF Estimate	Standard Error
	Intercept	**$5.74\times10^{-3***}$**	8.20×10^{-4}
Compositional	Message Length	2.20×10^{-4}	1.30×10^{-4}
Visual	Number of Emojis	**$1.05\times10^{-3**}$**	4.00×10^{-4}
	Face Emoji	-1.92×10^{-3}	1.44×10^{-3}
Affective	Emotional Valence	1.40×10^{-4}	9.00×10^{-5}
Assertive	Price Information	**$-6.78\times10^{-3***}$**	1.36×10^{-3}
	Discount Amount	3.00×10^{-5}	2.00×10^{-5}
	Discount Percentage	**$6.90\times10^{-5***}$**	2.00×10^{-5}
	Restrictions	**$-2.25\times10^{-3**}$**	7.30×10^{-4}
	Multiple Benefits	2.30×10^{-4}	1.15×10^{-3}
Directive	Urgency	2.20×10^{-4}	1.12×10^{-3}
	Call to Action	**$1.09\times10^{-2***}$**	2.13×10^{-3}
Category F.E.	Cosmetics	-1.60×10^{-4}	1.39×10^{-3}
	Department Store	7.70×10^{-4}	1.08×10^{-3}
	Electronics	-2.76×10^{-3}	1.44×10^{-3}
	Food	**$3.72\times10^{-3***}$**	1.04×10^{-3}

*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

Table 3.5: Causal forest results for individual design element

Reviewing the results, first we see the estimated coefficient for message length is insignificant, which suggests that the length of the push message does not have any impact on user engagement with push notifications. While the presence of face emoji does not seem to have any causal influence on user engagement, having more emojis in the message positively drives user open rate. Concretely, the inclusion of each additional emoji (up to a maximum of 3) leads to a 0.105% increase in the user's open rate of push notifications. However, the emotional valence of the push message does not have a significant effect in driving user engagement since its coefficient estimate is insignificant.

Next, we examine the treatment effects of the assertive design elements. We find that including price information in the push message decreases user open rate by 0.678% on average, which amounts to a significant drop in user engagement. Although the absolute dollar discount amount does not appear to have an influence on user open rate, each additional 1% discount specified in the message drives a 0.0069% increase in the average user open rate. Contrarily, specifying restrictions for offer redemption in the message leads to a 0.225% decrease in the average user open rate. Furthermore, the inclusion of multiple benefits in the push message does not lead to significant changes in the user's average open rate.

The last design elements are the directive design elements. We find that using language that emphasizes the urgency of the offer in the push message does not lead to a statistically significant difference in user engagement. However, using language that directly calls the user to take action with the push notification gives rise to a whopping 1.09% increase in user open rate. The effect of call-to-action language is especially remarkable, not only because its magnitude is the largest among all treatment effects for the design elements considered, but also because its significant relative size to the baseline average open rate across all push campaigns (50.44% of the baseline).

The category fixed effects suggest that while push campaigns in the cosmetics, department store and electronics categories generally do not give rise to more or less favorable user engagement, push campaigns in the food category have a 0.372% higher open rate on average than those in the clothing category.

Comparing the results from the causal forest analysis with those from the aggregate regression in Table 3.4, we notice several interesting similarities and contrasts. First, regardless of statistical significance, the same design element generally has a similar effect size in both the causal forest and the aggregate regression analyses. This suggests that the aggregate correlation patterns between design elements and user open at the campaign level, are typically reflective of the magnitude of the causal relationships between them at the individual user level. Second, the two analyses reveal overlapping but different sets of design elements that demonstrate significant statistical relationships with user open rate. In particular, the number of emojis, price information, and call to action are found to be statistically significant in both the aggregate regression and the causal forest analysis. Whereas multiple benefits and urgency are significantly correlated with average open rate at the aggregate campaign level, they do not appear to have significant causal impact on user open rate at the individual level. In contrast, discount percentage and offer restrictions exhibit no statistical correlation with user engagement at the campaign level, but they are shown to have significant causal impact on user open rate at the individual level. Third, design elements that are significantly correlated with average open rate at the campaign level could turn out to have significant causal effects on individual user open behavior in the opposite direction. In particular, using language that directly calls the user to take action with the push was correlated with a 1.07% decrease in the average campaign open rate. However, the effect of call to action is reversed in the causal forest analysis, for using call to action language is found to cause a 1.09% increase in individual users' open rate.

A plausible explanation for the reversal of treatment effects of design elements on user open rate, can be motivated by the statistical phenomenon known as the Simpson's paradox. In short, Simpon's paradox describes the situation in which an effect that persists in subgroups of data disappears or even reverses when the groups of data are pooled. This could result from a failure to account for confounding variables or the causal mechanism through which the effect takes place. Particularly in our context, the treatment effect of call to action may be positive within subgroups of users sharing similar characteristics, which then gives rise to a positive average treatment effect for this design element by pooling the *effects* across user subgroups. However, when *users* are pooled by their treatment status, we ignore the differences in the effect of the treatment on user open rate that are due to differences in user characteristics, hence masking the underlying subgroup treatment effects and opening up the aggregate correlations between the treatment variable and the average open rate to the influence of group sizes. Table 3.6 below depicts a hypothetical scenario that gives rise to the reversal of the treatment effect after aggregating users by their treatment status.

User group	No treatment		Treatment	
	Users	Opened	Users	Opened
A	600	4%	100	5%
B	300	3%	100	4%
C	100	2%	800	2.5%
Total	1000	3.5%	1000	2.9%

Table 3.6: Example of effect reversal for a treatment

Even if we control for average user characteristics in the aggregate correlational analysis, the specific causal pathways by which the treatment effects take place are still unaccounted for, thus making it possible for the actual treatment effects to be reversed in the aggregate. This discussion highlights the importance of allowing for flexible control of confounding variables as well as modeling the mechanism by which causal effects take place when studying causal relationships using observational data. The causal forest method enables us to account for the effects of confounding user characteristics by implicitly capturing the grouping structures through which the treatment effects of the design variables take place, consequently recovering the true average treatment effects.

3.5 CONCLUDING REMARKS

This study examines the causal effects of message design elements on user engagement with push notifications from a mobile coupon publisher app. Building on speech act theory and extant marketing research literature, we categorize and code the message content of 1,589 push notifications into five design categories and eleven design elements. Leveraging the machine learning method causal forest which is designed for causal inference on observational data, we recover the average treatment effect for each design element. We find that 5 of the 11 language design elements play a significant role in driving user engagement with push notifications. Specifically, number of emojis, price information, discount percentage, and call to action are found to have a positive causal impact on user engagement, with call to action giving rise to a staggering 1.09% increase to user open rate, which measures up to 50.44% of the baseline average open rate (2.171%) across all push campaigns. On the other hand, including offer restrictions can cause a 0.225% dip in individual user's open rate. Taken together, these results suggest 3 main takeaways: 1) the simple use of visual design elements such as emojis can boost the effectiveness of push notifications at generating user engagement; 2) assertive elements in the push message can drive differences in user engagement with the push notifications, and the effect of these assertive elements depends on the nature of the information: while including offer restrictions and price information in the push message will lower user engagement, higher specified discount percentages of the offer will give rise to higher user open rate; 3) cheap talk works - using language that directly calls for the user's taking action with the push proves to be a simple but significantly effective strategy to increase user engagement with the push campaign, and this strategy seems to be under-utilized by the publisher since less than 5% of the push campaigns in our data included this element in its message design.

Our research contributes to marketing theory by proposing a theory-based framework for studying what message design elements drive user engagement with mobile marketing communications. While extant literature on mobile marketing has examined the influence of various utilitarian and contextual factors on consumer response to mobile marketing offers, research that focuses on the message design aspect of the marketing communications has been scant. To fill this gap, we build on speech act theory and existing marketing literature to develop a conceptual framework that dissects a given message into its constituent design elements. Subsequently we then study the causal impact of design elements on user engagement. This framework can provide the basis for future research that seeks to examine the effect of message design on user engagement in other forms of marketing communications.

As demonstrated by the comparison of results from the aggregate campaign level analysis, and those from the causal forest analysis, design elements found to have significant causal effects on user engagement can have their effects vanish or be reversed in an aggregate analysis that fails to account for the moderating effect of observed user characteristics. This phenomenon can be explained by the Simpson's paradox, and speaks to the fact that statistical patterns in the aggregate are not necessarily informative about the underlying causal relationships between variables. Thus, a manager making decisions for the message design of push notifications may be misguided by the results of aggregate statistical analyses to under(over)-utilize design elements for future campaigns that are in fact effective (ineffective) drivers of user engagement.

Based on the results from the causal analysis, some practical implications emerge for the manager's design consideration of future push messages. The usage of emojis at the beginning of push messages improves user engagement possibly by invoking a sense of light-heartedness and fun, so it's advisable for the manager to continue to include these visual elements and choose emoji icons that are specifically pertinent to their product offerings and brand image for potentially higher effectiveness. Furthermore, when it comes to informative and assertive design elements, since the inclusion of offer restrictions and detailed price information leads to lower user engagement, it may be worthwhile for the manager to explore what other informative content can help capture the user's attention and give rise to better user engagement. Finally, cheap talk works and is quite effectively so at inducing better user engagement, i.e. using language such as "open this message" and "take a look at this exclusive offer" can boost user open rate by significant proportions only at the change of a few words. Consequently, the manager may be well served to continue exploiting various call-to-action language in push notifications for marketing campaigns with high targeted impact. A caveat about the use of call-to-action language is that its effectiveness could be impacted by how often it is used through different campaigns: it's probable that a tradeoff exists such that the more often users see these call-to-action messages the less effective they become. This kind of dynamics is beyond the scope of this study and can be the focus of future research that explicitly examines the dynamic effect of message design elements across time.

This study also informs several avenues for potentially fruitful future research. First, by focusing specifically on a mobile shopping app, our findings may not generalize to other categories of mobile apps. Notwithstanding, our methodological framework sets up a good template for researchers interested in studying the effects of mobile marketing communications on user engagement in other domains and categories. Second, we do not observe users' interaction and engagement with similar or competing apps in our data, thus we're unable to comment on whether and how the presence of competitive forces will influence the effectiveness of push notifications sent by the focal app. Although taking into account observed user characteristics partially controls for the effect of competitive forces on user engagement, future research may track and quantify such effect to provide a better understanding of the extent of the competitive influence. Third, due to the observational and targeted nature of the push campaigns in our data, we do not consider the dynamics of the effects of design elements on user engagement over time. It is possible that users may get used to push notifications with similar design elements and become less responsive over time, and understanding such wear-out effects of the design elements can help marketers better tailor their mobile push strategies towards each individual user. Finally, we restrict our study to 11 language design elements in 5 design categories, and future research can consider introducing new design elements to study their effects on user engagement in experimental settings. We hope our study can motivate future research endeavors in these directions.

Appendix A: Data integration and variable definitions

We discuss some important units of analysis and clarify the data integration procedures we use in this section.

First of all, we regard different MAC ID's to be associated with mobile devices owned by different potential consumers. While some consumers may carry multiple mobile devices with Wi-Fi capability, given our data, it is empirically infeasible to differentiate this scenario from that in which two or more consumers all with mobile devices are detected by the IoT sensor at about the same time. We convert the Wi-Fi signal data of a MAC ID into observations of pass-by sessions and store visits by the owner. For each MAC ID, we define a pass-by session to be a time interval between when a signal is first picked up from the device and when it is last detected without going missing for a long time in between. We count this session as an “arrival” during the time window in which the beginning time of the session falls: for example, if a MAC ID shows up at 11:57am and is last seen at 12:01pm with continual signal occurrence in between, then this counts as a pass-by session in the 6am-noon time window. Furthermore, if the interval between two signals from the same MAC ID is longer than 400 seconds, then the next signal is treated as the beginning of a new session. A pass-by session is considered to have led to a store visit, if the signal strength exceeds -56 at any point during the session.

Second, to attribute a transaction to a consumer whose WiFi signals suggest that she visited the store around the time of the transaction, we applied the following matching rule: we regard a transaction as a purchase arising from a consumer's store visit, if within the last 60 seconds leading up to the transaction, the consumer's WiFi signals have been detected in-store at some point. We chose 60 seconds, because the time duration of most store visits is less than 60 seconds in each store. n case of multiple devices satisfying the previous condition, we attribute the transaction to the MAC ID whose signals were the last to be seen before the transaction, as a tie-breaking rule. Using this approach, we are able to match over 20% of all sales to a detected mobile device in the Wi-Fi dataset. One caveat associated with the tie-breaking rule is that when multiple devices satisfy the matching condition, it inevitably generates misattribution of transactions to the wrong consumer, when only one of the other consumers present was the actual purchaser. Such concern of misattribution, is alleviated by the fact that the average number of devices present during the 60 seconds prior to a matched transaction is only about 1.2 for each of the 8 stores, and this implies that on average only 1 out of 6 matched transactions is misattributed to the wrong consumer, assuming that all matched transactions were conducted by consumers who used mobile devices; if some of the matched transactions were in fact conducted by consumers who did not carry mobile devices, then the misattribution rate should become larger, however we are unable to determine the magnitude of increase because of a lack of information on the consumers not carrying mobile devices.

Third, for each store visit, we map the corresponding promotion breadth variables to the visit. To do this, we first calculate the category-level promotion breadth for each of the Food, Beverages and Miscellaneous categories, across all four promotion periods; given the time of a store visit, we can then determine the promotion period in which the visit falls, and tag the promotion breadths to the consumer's purchase decision in each category. When constructing the category-level promotion breadth variables, we count only the products in the categories that have been sold at least once during the product period, and this helps avoid exaggerating the actual breadth of price promotions available to the consumer. For example, if the price promotion scheme indicates that 20 products in the beverage category are on promotion, but the store's transaction data shows that 8 of these products have never been sold during the entire data period, then it is likely that the store does not have these products in stock, so that only the other 12 products should be counted towards the proportion of products being on price promotions when constructing the effective promotion breadth for the beverage category in this store.

Finally, since we do not observe demographic information associated with each consumer, so to control for observed heterogeneity across consumers, we use consumers' observable behavioral variables to characterize their differences. These behavioral variables include $Cumulative\ Sessions$, $Cumulative\ Visits$, $Out\ Duration$, $In\ Duration$, $Prior\ Session$, $Prior\ Visit$, $Time\ Since\ Last\ Session$, $Time\ Since\ Last\ Visit$. $Cumulative\ Sessions$ is the running total of a consumer's pass-by sessions made to around the store area up to the current session, and $Cumulative\ Visits$ keeps count of the total number of times when a consumer's pass-by session turned into a store visit up until the current session. $Out\ Duration$ keeps track of the time that the consumer has spent outside the store in a given session, and $In\ Duration$ reflects the time that the consumer has spent inside the store. $Prior\ Session$ and $Prior\ Visit$ indicate whether the consumer has had either a session or a visit prior to the current session respectively. $Time\ Since\ Last\ Session$ and $Time\ Since\ Last\ Visit$ respectively keep track of the time passed since the consumer’s last session and last visit in hours.

Appendix B: Details of the MCMC sampling chain

B.1: Pass-by model

Given each consumer i's number of pass-by sessions n_{iq} in window q across the entire data period, we have for the pass-by model the following joint posterior distribution:

$$f\left(\overrightarrow{\alpha_{\iota 0}}, \alpha, \mu_\alpha, \sigma_\alpha | n_{iq} \forall i, q\right)$$

$$\propto \prod_i \prod_q \left(\lambda_{iq}^{n_{iq}} \cdot e^{-\lambda_{iq}}\right) \cdot \sum_i \left[\sigma_\alpha^{-1} \cdot e^{-\frac{1}{2}\left(\frac{\alpha_{i0}-\mu_\alpha}{\sigma_\alpha^2}\right)^2}\right] \cdot \pi(\alpha) \cdot \pi(\mu_\alpha) \cdot \pi(\sigma_\alpha^2)$$

where $\lambda_{iq} = e^{\alpha_{i0}+Q_q \cdot \alpha_Q + W_q \cdot \alpha_W + Q_q \cdot W_q \cdot \alpha_{QW}}$, $\pi(\alpha)$, $\pi(\mu_\alpha)$ and $\pi(\sigma_\alpha^2)$ are the priors for α, μ_α and σ_α^2 respectively. We construct the MCMC sampling chain in the following steps.

Use a Metropolis-Hastings step and a normal proposal distribution to update the values of $\alpha = \left(\alpha_Q, \alpha_W, \alpha_{QW}\right)'$ with acceptance probability for new values of α^* given by:

$$\rho = min\left\{1, \frac{\prod_i \prod_q \left(\lambda_{iq}^{n_{iq}}(\alpha^*) \cdot e^{-\lambda_{iq}(\alpha^*)}\right)}{\prod_i \prod_q \left(\lambda_{iq}^{n_{iq}} \cdot e^{-\lambda_{iq}}\right)} \cdot \frac{|\Sigma_\alpha|^{-\frac{1}{2}} \cdot e^{\left[-\frac{1}{2} \cdot (\alpha^*-\Theta_\alpha)' \Sigma_\alpha^{-1} (\alpha^*-\Theta_\alpha)\right]}}{|\Sigma_\alpha|^{-\frac{1}{2}} \cdot e^{\left[-\frac{1}{2} \cdot (\alpha-\Theta_\alpha)' \Sigma_\alpha^{-1} (\alpha-\Theta_\alpha)\right]}}\right\}$$

where Θ_α, Σ_α are the prior mean and covariance for α.

For each of the individual random intercept, use a Metropolis-Hastings step with acceptance probability for new values of α_{i0}^* given by:

$$\rho = min\left\{1, \frac{\prod_i \prod_q \left(\lambda_{iq}^{n_{iq}}(\alpha_{i0}^*) \cdot e^{-\lambda_{iq}(\alpha_{i0}^*)}\right)}{\prod_i \prod_q \left(\lambda_{iq}^{n_{iq}} \cdot e^{-\lambda_{iq}}\right)} \cdot \frac{\sigma_\alpha^{-1} \cdot e^{-\frac{1}{2}\left(\frac{\alpha_{i0}^*-\mu_\alpha}{\sigma_\alpha^2}\right)^2}}{\sigma_\alpha^{-1} \cdot e^{-\frac{1}{2}\left(\frac{\alpha_{i0}-\mu_\alpha}{\sigma_\alpha^2}\right)^2}}\right\}$$

Then with newly updated values for α_{i0}^*, we apply a Gibbs step to update μ_α and σ_α^2 respectively by their posteriors:

$$\mu_\alpha \sim N\left(\frac{\frac{m_\mu}{s_\mu^2} + \frac{N_i}{s^2}\overline{\alpha_\iota}}{\frac{1}{s_\mu^2} + \frac{N_i}{s^2}}, \left(\frac{1}{s_\mu^2} + \frac{N_i}{s^2}\right)^{-1}\right)$$

$$\frac{1}{\sigma_\alpha^2} \sim Gamma\left(\nu_1 + \frac{N_i}{2}, \kappa_1 + \frac{1}{2}\sum_i (\alpha_{i0} - \mu_\alpha)^2\right)$$

where N_i is the number of consumers in the data.

B.2: Store visit and purchase models with latent instrumental variable

With the correlated error structure between the probit model for store visit, tobit model for in-store purchases, and the latent instrument equation, we may construct the MCMC sampling chain in the following steps. For purchase decision in one of the three product categories, say category 1, we adjust the variance of the purchase value equation and the error using the conditional distribution $\tau_{it}^1|\varepsilon_{it},\tau_{it}^2,\tau_{it}^3,\xi_{it}$ from the joint normal distribution of the error terms $MVN(0,\Sigma)$. We have $\tau_{it}^1 \sim N\left(\Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot(\varepsilon_{it},\tau_{it}^2,\tau_{it}^3,\xi_{it})',\Sigma_{\tau^1,\tau^1}-\Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot\Sigma_{-\tau^1,\tau^1}\right)$, which then gives rise to $UP_{it}^1 = \left(\delta_{i0}^1 + Z_{it}^1\cdot\delta^1 + \Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot(\varepsilon_{it},\tau_{it}^2,\tau_{it}^3,\xi_{it})',\Sigma_{\tau^1,\tau^1}-\Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot\Sigma_{-\tau^1,\tau^1}\right)$. So, if $y_{it}^1 < 0$ then we sample $UP_{it}^1 < 0$ from this conditional distribution truncated above at 0; if $y_{it}^1 > 0$ then we set $UP_{it}^1 = y_{it}^1$. Now set $\widetilde{UP_{it}^1} = UP_{it}^1 - \delta_{i0}^1 - \Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot(\varepsilon_{it},\tau_{it}^2,\tau_{it}^3,\xi_{it})' \sim N\left(Z_{it}^1\cdot\delta^1,\Sigma_{\tau^1,\tau^1}-\Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot\Sigma_{-\tau^1,\tau^1}\right)$, then we can draw δ^1 with a Gibbs step from the posterior distribution:

$$N\left(\left(\widetilde{\sigma_1}^{-2}Z^{1\prime}Z^1+\Psi_{\delta,1}^{-1}\right)^{-1}\left(\widetilde{\sigma_1}^{-2}Z^{1\prime}\widetilde{UP}^1+\Psi_{\delta,1}^{-1}\mu_{\delta,1}\right),\left(\widetilde{\sigma_1}^{-2}Z^{1\prime}Z^1+\Psi_{\delta,1}^{-1}\right)^{-1}\right)$$

where $\mu_{\delta,1}$ and $\Psi_{\delta,1}^{-1}$ are the prior mean and prior precision matrix for δ_1, and $\widetilde{\sigma_1} = \Sigma_{\tau^1,\tau^1} - \Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{-\tau^1,-\tau^1}^{-1}\cdot\Sigma_{-\tau^1,\tau^1}$.

With δ_1 updated, we then update the individual random intercepts by drawing from $\delta_{i0}^1 \sim N\left(UP_{it}^1 - Z_{it}^1\cdot\delta^1 - \Sigma_{\tau^1,-\tau^1}\cdot\Sigma_{\tau^1,-\tau^1}^{-1}\cdot(\varepsilon_{it},\tau_{it}^2,\tau_{it}^3,\xi_{it})',\left(\frac{1}{\sigma_{\delta i,1}^2}+\frac{n_i}{\widetilde{\sigma_1}^2}\right)^{-1}\right)$, where n_i is the number of observations for consumer i. Then the prior mean and variance for δ_{i0}^1 can both be updated with a Gibbs step:

$$\mu_{\delta i,1} \sim N\left(\frac{N_i\frac{\overline{\delta_{i0}^1}}{\sigma_{\delta i,1}^2}+\frac{m_{\mu i1}}{s_{\mu i1}^2}}{\frac{1}{s_{\mu i1}^2}+\frac{N_i}{\sigma_{\delta i,1}^2}},\left(\frac{1}{s_{\mu i1}^2}+\frac{N_i}{\sigma_{\delta i,1}^2}\right)^{-1}\right)$$

$$\frac{1}{\sigma_{\delta i,1}^{2}} \sim Gamma\left(\nu_{\delta 1}+\frac{N_i}{2}, \kappa_{\delta 1}+\frac{1}{2}\sum_{i}\left(\delta_{i0}^{1}-\mu_{\delta i,1}\right)^{2}\right)$$

where $m_{\mu i1}$ and $s_{\mu i1}^{2}$ are the hyper mean and hyper variance for $\mu_{\delta i,1}$, which are taken to be 0 and 10000 in estimation.

For the store visit decision, we have $\varepsilon_{it} \sim N\left(\Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot (\tau_{it}^{1}, \tau_{it}^{2}, \tau_{it}^{3}, \xi_{it})', \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot \Sigma_{-\varepsilon,\varepsilon}\right)$ for consumers who visited the store, resulting in $UV_{it} = N\left(\beta_{io} + X_{it} \cdot \beta + \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot (\tau_{it}^{1}, \tau_{it}^{2}, \tau_{it}^{3}, \xi_{it})', \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot \Sigma_{-\varepsilon,\varepsilon}\right)$, and $\varepsilon_{it} \sim N\left(\Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \xi_{it}, \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \Sigma_{\xi,\varepsilon}\right)$ for consumers who did not visit the store, resulting in $UV_{it} = N\left(\beta_{io} + X_{it} \cdot \beta + \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \xi_{it}, \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \Sigma_{\xi,\varepsilon}\right)$. So, we sample $UV_{it} < 0$ if $y_{it}^{v} = 0$ and $UV_{it} > 0$ if $y_{it}^{v} = 1$. Now set $\widetilde{UV}_{it} = UV_{it} - \beta_{io} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot (\tau_{it}^{1}, \tau_{it}^{2}, \tau_{it}^{3}, \xi_{it})' \sim N\left(X_{it} \cdot \beta, \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot \Sigma_{-\varepsilon,\varepsilon}\right)$ for consumers who visited the store, and $\widetilde{UV}_{it} = UV_{it} - \beta_{io} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \xi_{it} \sim N\left(X_{it} \cdot \beta, \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \Sigma_{\xi,\varepsilon}\right)$ for consumers who did not visit the store, then we can update β with a Gibbs step from the posterior distribution given by:

$$\beta \sim N\left(\left(X'\Xi^{-1}X + \Psi_{\beta}^{-1}\right)^{-1}\left(X'\Xi^{-1}\widetilde{UV} + \Psi_{\beta}^{-1}\mu_{\beta\beta}\right), \left(X'\Xi^{-1}X + \Psi_{\beta}^{-1}\right)^{-1}\right)$$

where $\mu_{\beta\beta}, \Psi_{\beta}^{-1}$ are the prior mean and prior precision matrix for β, and Ξ is a diagonal matrix whose diagonal elements are $\Xi_{ii} = \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot \Sigma_{-\varepsilon,\varepsilon}$ for consumers who visited the store, and $\Xi_{ii} = \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \Sigma_{\xi,\varepsilon}$ for those who did not visit.

With the updated β, we then update the individual random intercepts by drawing from $\beta_{i0} \sim N\left(UV_{it} - X_{it} \cdot \beta - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot (\tau_{it}^{1}, \tau_{it}^{2}, \tau_{it}^{3}, \xi_{it})', \left(\frac{1}{\sigma_{\beta}^{2}} + \frac{n_i}{\widetilde{\sigma_v}^{2}}\right)^{-1}\right)$ for each consumer who visited the store, and $\beta_{i0} \sim N\left(UV_{it} - X_{it} \cdot \beta - \Sigma_{\varepsilon,\nu} \cdot \Sigma_{\nu,\nu}^{-1} \cdot \xi_{it}, \left(\frac{1}{\sigma_{\beta}^{2}} + \frac{n_i}{\widetilde{\sigma_{nv}}^{2}}\right)^{-1}\right)$ for those who didn't visit, where $\widetilde{\sigma_v}^{2} = \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,-\varepsilon} \cdot \Sigma_{-\varepsilon,-\varepsilon}^{-1} \cdot \Sigma_{-\varepsilon,\varepsilon}$, $\widetilde{\sigma_{nv}}^{2} = \Sigma_{\varepsilon,\varepsilon} - \Sigma_{\varepsilon,\xi} \cdot \Sigma_{\xi,\xi}^{-1} \cdot \Sigma_{\xi,\varepsilon}$. Then the prior mean and variance for β_{i0} can both be updated with a Gibbs step:

$$\mu_\beta \sim N\left(\frac{N_i\dfrac{\overline{\beta_{i0}}}{\sigma_\beta^2}+\dfrac{m_{\beta i}}{s_{\beta i}^2}}{\dfrac{1}{s_{\beta i}^2}+\dfrac{N_i}{\sigma_\beta^2}}, \left(\frac{1}{s_{\beta i}^2}+\frac{N_i}{\sigma_\beta^2}\right)^{-1}\right)$$

$$\frac{1}{\sigma_\beta^2} \sim Gamma\left(\nu_\beta+\frac{N_i}{2}, \kappa_\beta+\frac{1}{2}\sum_i\left(\beta_{i0}-\mu_\beta\right)^2\right)$$

where $m_{\beta i}$ and $s_{\beta i}^2$ are the hyper mean and hyper variance for μ_β, which are taken to be 0 and 10000 in estimation.

Similar to the treatment of β, we adjust the error and variance of the latent instrument equation with $\xi_{it} \sim N\left(\Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot (\varepsilon_{it}, \tau_{it}^1, \tau_{it}^2, \tau_{it}^3)', \Sigma_{\xi,\xi} - \Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot \Sigma_{-\xi,\xi}\right)$ for consumers who visited the store, and $\xi_{it} \sim N\left(\Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \varepsilon_{it}, \Sigma_{\xi,\xi} - \Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \Sigma_{\varepsilon,\xi}\right)$ for those who did not. This gives rise to $\widetilde{x_{it}} = x_{it} - \Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot (\varepsilon_{it}, \tau_{it}^1, \tau_{it}^2, \tau_{it}^3)' \sim N\left(\omega \cdot \gamma_i, \Sigma_{\xi,\xi} - \Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot \Sigma_{-\xi,\xi}\right)$ for consumers who visited the store, and $\widetilde{x_{it}} = x_{it} - \Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \varepsilon_{it} \sim N\left(\omega \cdot \gamma_i, \Sigma_{\xi,\xi} - \Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \Sigma_{\varepsilon,\xi}\right)$ for those who didn't. Then we can update ω with a Gibbs step from the posterior distribution:

$$\omega \sim N\left(\left(\gamma'\Lambda^{-1}\gamma+\Psi_\gamma^{-1}\right)^{-1}\left(\gamma'\Lambda^{-1}\tilde{x}+\Psi_\gamma^{-1}\mu_\gamma\right), \left(\gamma'\Lambda^{-1}\gamma+\Psi_\gamma^{-1}\right)^{-1}\right)$$

where Λ is a diagonal covariance matrix with diagonal elements $\Lambda_{ii} = \Sigma_{\xi,\xi} - \Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot \Sigma_{-\xi,\xi}$ for consumers who visited the store, and $\Lambda_{ii} = \Sigma_{\xi,\xi} - \Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \Sigma_{\varepsilon,\xi}$ for those who did not.

For the class indicator γ, the posterior probability for class k given the prior class probabilities π_k is:

$$Pr\left(\gamma_{it}^k = 1 \middle| \widetilde{x_{it}}\right) = \frac{f\left(\widetilde{x_{it}} \middle| \omega^k, \sigma_\xi^2\right) \cdot \pi_k}{\sum_j f\left(\widetilde{x_{it}} \middle| \omega^j, \sigma_\xi^2\right) \cdot \pi_j}$$

where $f\left(\widetilde{x_{it}} \middle| \omega^k, \sigma_\xi^2\right)$ is the normal PDF at support $\widetilde{x_{it}}$ with mean and variance being ω^k and σ_ξ^2, with $\sigma_\xi^2 = \Sigma_{\xi,\xi} - \Sigma_{\xi,-\xi} \cdot \Sigma_{-\xi,-\xi}^{-1} \cdot \Sigma_{-\xi,\xi}$ for consumers who visited the store, and $\sigma_\xi^2 = \Sigma_{\xi,\xi} - \Sigma_{\xi,\varepsilon} \cdot \Sigma_{\varepsilon,\varepsilon}^{-1} \cdot \Sigma_{\varepsilon,\xi}$ for consumers who didn't. Thus, we update the class indicator to be the one having the highest posterior probability from the above calculation.

The prior class probabilities are then updated from the posterior distribution:

$$\pi \sim Dirichlet\left(r_1 + \sum_{i=1}^{N} I(\gamma_{it}^{1} = 1), \dots, r_K + \sum_{i=1}^{N} I(\gamma_{it}^{K} = 1)\right)$$

where N is the total number of observations.

Finally, with all other parameters updated from the previous steps, we can calculate the errors for all 5 equations for consumers who visited the store, and augment $(\tau_{it}^1, \tau_{it}^2, \tau_{it}^3)$ for consumers who did not visit by

$$(\tau_{it}^1, \tau_{it}^2, \tau_{it}^3)' \sim MVN\left(\Sigma_{-\varepsilon\xi,\varepsilon\xi} \cdot \Sigma_{\varepsilon\xi,\varepsilon\xi}^{-1} \cdot (\varepsilon_{it}, \xi_{it})', \Sigma_{-\varepsilon\xi,-\varepsilon\xi} - \Sigma_{-\varepsilon\xi,\varepsilon\xi} \cdot \Sigma_{\varepsilon\xi,\varepsilon\xi}^{-1} \cdot \Sigma_{\varepsilon\xi,-\varepsilon\xi}\right)$$

Then at last, the covariance matrix for the error terms can be sampled from the following inverse Wishart posterior distribution:

$$\Sigma \sim IW\left(\psi + 5N, \psi I + \sum_{i=1}^{N} (\varepsilon_{it}, \tau_{it}^1, \tau_{it}^2, \tau_{it}^3, \xi_{it})' \cdot (\varepsilon_{it}, \tau_{it}^1, \tau_{it}^2, \tau_{it}^3, \xi_{it})\right)$$

References

Ailawadi, Kusum L., Jie Zhang, Aradhna Krshna and Michael W. Kruger (2010), "When Wal-mart Enters: How Incumbent Retailers React and How This Affects Their Sales Outcomes," *Journal of Marketing Research,* 47 (4), 577-593.

Alemdar, Hande and Cem Ersoy (2010), "Wireless Sensor Networks for Healthcare: A Survey," *Computer Networks,* 54 (15), 2688-2710.

Andrews, Mchelle, Xueming Luo, Zheng Fang, and Anindya Ghose (2015), "Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness," *Marketing Science,* 35 (3), 218-33.

Aral, Sinan and Christos Nicolaides (2017), "Exercise Contagion in a Global Social Network," *Nature Communications*, 8.

Ascarza, Eva (2018), "Retention Futility: Targeting High-Risk Customers Might be Ineffective," *Journal of Marketing Research,* 55 (1), 80-98.

Athey, Susan and Guido Imbens (2016), "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Science,* 113 (27), 7353-60.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019), "Generalized Random Forests," *Annals of Statistics*, 47 (2), 1148-78.

Athey, Susan and Stefan Wager (2019), "Estimating Treatment Effects with Causal Forests: An Application," *Observational Studies,* 5, 36-51.

Bell, David R. and James M. Lattin (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17 (1), 66-88.

Bell, David R., Jeongwen Chiang and V. Padmanabhan (1999), "The Decomposition of Promotional Response: An Empirical Generalization," *Marketing Science* 18 (4), 504-526.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192-205.

Bezawada, Ram and Koen Pauwels (2013), "What is Special about Marketing Organic Products? How Organic Assortment, Price, and Promotions Driver Retailer Performance," *Journal of Marketing*, 77 (1), 31-51.

Bleier, Alexander, Colleen M. Harmeling, and Robert W. Palmatier (2019), "Creating Effective Online Customer Experiences," *Journal of Marketing*, 83 (2), 98-119.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, 29 (2), 201-215.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," *Econometrics Journal,* 21 (1), C1-C68.

Chintagunta, Pradeep K. (1993), "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households," *Marketing Science,* 12 (2), 184-208.

Danaher, Peter J., Michael S. Smith, Kulan Ranasinghe, and Tracey S. Danaher (2015), "Where, When, and How Long: Factors That Influence the Redemption of Mobile Phone Coupons," *Journal of Marketing Research*, 52 (October), 710-25.

Dube, Jean-Pierre, Zheng Fang, Nathan Fong, and Xueming Luo (2017), "Competitive Price Targeting with Smartphone Coupons," *Marketing Science,* 36 (6), 944-75.

Erdem, Tulini and Michael P. Keane (1996), "Decision-making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science,* 15 (1), 1-20.

Fang, Zheng, Bin Gu, Xueming Luo and Yunjie Xu (2015), "Contemporaneous and Delayed Sales Impact of Location-based Mobile Promotions," *Information Systems Research*, 26 (3), 552-564.

Fong, Nathan, Zheng Fang, and Xueming Luo (2015), "Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions." *Journal of Marketing Research*, 52 (5), 726-35.

Fong, Nathan, Yuchi Zhang, Xueming Luo, and Xiaoyi Wang (2018), "Targeted Promotions on an E-Book Platform: Crowding Out, Heterogeneity, and Opportunity Costs," *Journal of Marketing Research*, 56 (2), 310-23.

Ghose, Anindya, Avi Goldfarb and Sang Pil Han (2013), "How is the Mobile Internet Different? Search Costs and Local Activities," *Information Systems Research,* 24 (3), 613-631.

Ghose, Anindya, Beibei Li and Siyuan Liu (2019), "Mobile Targeting Using Customer Trajectory Patterns," *Management Science,* 65 (11), 5027-49.

Ghose, Anindya, Hyeokkoo Eric Kwon, Dongwon Lee and Wonseoko Oh (2019), "Seizing the Commuting Moment: Contextual Targeting Based on Mobile Transportation Apps," *Information Systems Research,* 30 (1), 154-174.

Guo, Tong, S. Sriram, and Puneet Manchanda (2019). "The Effect of Information Disclosure on Industry Payments to Physicians," *Journal of Marketing Research*, 58 (1), 115-40.

Guadagni, Peter M. and John D. C. Little (1984), "A Logit Model of Brand Choice Calibrated on Scanner DATA," *Marketing Science,* 27 (1), 29-48.

Gupta, Sachin, Pradeep K. Chintagunta, Dick R. Wittink and Anil Kaul (1998), "Do Household Scanner Panels Provide Representative Inferences from Brnd Choices: A Comparison with Store Data," *Journal of Marketing Research*, 33 (4), 383-398.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342-355.

Hoffman, Donna L. and Thomas P. Novak (2015), Emergent Experience and the Connected Consumer in the Smart Home Assemblage and the Internet of Things. Unpublished manuscript. URL: *http://dx.doi.org/10.13140/RG.2.1.3966.1282.*

Hui, Sam K., J. Jeffrey Inman, Yanliu Huang and Jacob Suher (2013), "The effect of In-store Travel Distance on Unplanned Spending: Applications to Mobile Promotion Strategies," *Journal of Marketing,* 77 (2), 1-16.

Hui, Sam K., Yanliu Huang, Jacob Suher and J. Jeffrey Inman (2013), "Deconstructing the 'First Moment of Truth': Understanding Unplanned Consideration and Purchase

Conversion Using In-store Video Tracking", *Journal of Marketing Research*, 50 (4), 445-462.
Kahn, Barbara E. and David S. Schmittlein (1989), "Shopping Trip Behavior: An Empirical Investigation," *Marketing Letters*, 1 (1), 55-69.
Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research* 26 (4), 379-390.
Lee, Dokyun, Kartik Hosanagar, and Harikesh S. Nair (2018), "Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook," *Management Science* 64 (11), 5105-31.
Li, Chenxi, Xueming Luo, Cheng Zhang and Xiaoyi Wang (2017), "Sunny, Rainy, and Cloudy with a Chance of Mobile Promotion Effectiveness," *Marketing Science*, 36 (5), 762-779.
Li, Hongshuang and P. K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51 (1), 40-56
Li, Xueni, Kimmy Wa Chan, and Sara Kim (2019), "Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters," *Journal of Consumer Research*, 45 (5), 973-87.
Luo, Xueming, Michelle Andrews, Zheng Fang, and Cee Wei Phang (2014). "Mobile Targeting," *Management Science*, 60 (7): 1738-56.
McShane, Lindsay, Ethan Pancer, Maxwell Poole, and Qi Deng (2021), "Emoji, Playfulness, and Brand Engagement on Twitter," *Journal of Interactive Marketing*, 53 (1): 96-110.
Manchanda, Puneet, Asim Ansari and Sunil Gupta (1999), "The 'Shopping Basket': A Model for Multicategory Purchase Incidence Decisions', *Marketing Science*, 18 (2), 95-114.
Moe, Wendy W. (2003), "Buying, Searching, or Browsing: Differentiating between Online Shoppers Using IN-store Navigational Clickstream," *Journal of Consumer Psychology*, 13 (1-2), 29-39.
Moe, Wendy W. and Peter S. Fader (2004), "Dynamic Conversion Behavior at E-commerce Sites," *Management Science*, 50 (3), 326-335.
Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579-595.
Ng, Irene C. L. and Susan Y. L. Wakenshaw (2017), "The Internet-of-Things: Review and Research Directions." *International Journal of Research in Marketing*, 34 (1), 3-21.
Niraj, Rakesh, V. Padmanabhan and P. B. Seetharaman (2008), "A Cross-category Model of Households' Incidence and Qunatity Decisions," *Marketing Science*, 27 (2), 225-235.
Ordenes, Francisco Villarroel, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2019), "Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research*, 45 (5), 988-1012.
Robinson, Peter M. (1988), "Root-n-Consistent Semiparametric Regression," *Econometrica: Journal of the Econometric Society*, 56 (4), 931-54.

Rosenbaum, Paul R. and Donald B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, 70 (1), 41-55, 1983.
Rubin, Donald B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 66(5): 688-701.
Ruiz-Grcia, Luis, Loredana Lunadei, Pillar Barrerio and Jose Ignacio Robla (2009), “A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends,” *Sensors*, 9, 4728-50.
Seiler, Stephan and Fabio Pinna (2017), “Estimating Search Benefits from Path-tracking Data: Measurement and Determinants,” *Marketing Science*, 36 (40), 565-589.
van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel and Philip Hans Franses (2010), “Retrieving Unobserved Consideration Sets from Household Panel Data,” *Journal of Marketing Research*, 47 (1), 63-74.
Wager, Stefan and Susan Athey (2018), “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” *Journal of the American Statistical Association*, 113 (523), 1228-42.
Xu, Heng, Xin Luo, John M. Carroll and Mary Beth Rosson (2011), “The Personalization Privacy Paradox: An Exploratory Study of Decision-making Process for Location-aware Marketing,” *Decision Support Systems*, 51 (1), 42-52.
Xu, Lizhen, Jason A. Duan and Andrew Whinston (2014), “Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion,” *Management Science*, 60 (6), 1392-1412.
Yoo, Youngjin, Richard J. Boland Jr., Kalle Lyytinen and Ann Majchrzak (2012), “Organizing for Innovation in the Digitized World,” *Organization Science*, 23 (5), 1398-1408.