ABSTRACT

Title of Dissertation **MARKETING IN MOBILE, OMNI-CHANNEL AND MULTI-FORMAT CONTEXTS**

Xian Gu
Doctor of Philosophy, 2019

Chair: **Professor P. K. Kannan**
Marketing Department

Co-Chair: **Associate Professor Liye Ma**
Marketing Department

This dissertation studies the marketing issues in mobile, omni-channel and multi-format contexts. In the first essay, while conventional wisdom indicates that apps have a positive impact on customer spending, I critically examine this premise by estimating the impact of app adoption on customers' omni-channel spending in the context of a hotel chain and identifying the factors contributing to such impact. Exploiting the variation in customers' timing of app adoption and difference-in-differences approach, I find that app adoption has a significant negative impact on total customer spending. This negative effect is robust to controlling for customer self-selection bias, measuring effects across alternative time frames, customer spending patterns and app usage behaviors, using different measures of purchase and alternative model specifications, and using different random samples. A survey of app adopters reveals that customers who adopt the focal app are also most likely to adopt competitors' apps, and therefore are likely to search more and shop around, leading to decreased share of

wallet with the firm. My analysis also reveals that the negative impact on spending is lower for those customers who use the apps for mobile check-in than those who use the apps for just searching. So by encouraging customers to engage with the full functionality of the app, firms can possibly mitigate the negative impact.

In the second essay, I further develop an integrated dynamic structural model to investigate consumers' decision to adopt the mobile app, and its subsequent impacts on their decisions to purchase using alternative shopping channels. The policy simulations show that consumers are less likely to make reservations with the focal firm if the firm had not introduced their apps.

Finally, in the third essay, I investigate the strategy of extending the product line by providing an additional premium version as a means to spur demand for the existing premium version. I highlight how extending the results of the standard product line model is insufficient in such cases due to the conceptual nuances that the presence of the free version introduces in a freemium context. I conduct a randomized field experiment with an online content provider, the National Academies Press, which offers book titles in a PDF version for free and sells the paperback version for a premium. Overall, I show that paperback titles accompanied by an additional premium version, either an ebook or a hardcover format, have higher sales than those in the control condition. The positive impact on paperback sales is stronger for titles that are more popular or lower in price, and the effect of introducing the ebook version is higher when the ebook price is closer to the paperback price. By analyzing customer choices at the individual level, I identify the existence of the compromise effect and the attraction effect in the extended product line setting, a significant contribution not only in the freemium context but also to the product line literature.

MARKETING IN MOBILE, OMNI-CHANNEL AND MULTI-FORMAT CONTEXTS

By

Xian Gu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor P. K. Kannan, Chair
Associate Professor Liye Ma, Co-Chair
Professor Michel Wedel
Professor David Godes
Professor Guido Kuersteiner

Table of Contents

List of Tables

List of Figures

Chapter 1: The Dark Side of Mobile App Adoption: Examining the Impact on Customers' Multichannel Purchase

1.1 Introduction

Mobile applications (apps) provide brands an excellent conduit to interact directly with customers, enhance shopping and consumption experiences, build customer loyalty, and thereby beat the competition. According to the *2017 U.S. Mobile App Report* by Lipsman and Lella (2017), approximately 75 percent of smartphone users have installed retail apps in their smartphones. Moreover, their mobile app usage accounts for 47 percent of their total time spent using digital devices in the retailing category, dominating desktop (40 percent) and mobile website (13 percent) in usage time. Firms now expect higher spending from customers who adopt mobile apps and use omni-channel options. Over the past several years, mobile apps have become a major source of retail e-commerce revenues. At the same time, many brands have increasingly turned to revolutionizing services offered via their apps, which deliver additional convenience for customers not possible in traditional shopping channels. For example, several major hotel chains have offered keyless entry service which enables customers to unlock their hotel room using the mobile app. Our research specifically applies to mobile apps which primarily serve as a shopping channel and are commonly provided by many retailers, travel firms, etc.

Introducing mobile app raises several critical questions for firms, most important of which is how mobile apps impact customers' total spending with the firm. A related issue is understanding the factors contributing to the impact so that retailers can take actions to enhance the positive impact or reduce the negative impact as the case may be. We investigate these questions in the context of customer spending in an omni-channel environment. We leverage a unique dataset from a leading international hotel chain that introduced its mobile app in August 2011. The dataset covers a four-year period between 2011 and 2014 and contains rich information about hotel reservations, customer app adoption, app usage, and demographics. Focusing on a random sample of 60,000 customers who ultimately adopted the app at six different points of time (making up six cohorts), we track the total and channel-specific spending of each customer and examine the impact of customer app adoption on their total spending with the firm as well as their spending in each channel.

A major methodological concern in analyzing such time-series data, however, is that there could be unobserved factors driving customer app adoption which could be correlated with their purchase behavior. We address this concern in several ways. First, we construct multiple treatment groups and control groups by taking advantage of the fact that customers adopted the mobile app at different points of time. For example, a treatment group adopting the mobile app earlier in time is compared with a control group which adopts the mobile app later in time, thereby controlling for the time-invariant unobserved factors that impact app adoption, while providing sufficient time-window for a valid comparison of the two groups. We then use difference-in-differences approach on the data from the two groups to account for the bias impacting customer spending due to time invariant unobservables (e.g., business traveler) and time variant unobservables that are common to all individuals (e.g., seasonal effects). Furthermore, we augment the difference-in-differences approach with the Heckman correction to fully control for customer self-selection bias, that is, the selection on observables (e.g., past purchase pattern of customers) and unobservables (e.g., idiosyncratic demand shock). Finally, we use propensity score matching to increase the comparability of the treatment group and the control group based on customer pre-adoption behavior and demographics. We replicate the analysis using multiple treatment-control pairs spread over time to account for the impact updated versions of the app.

Based on past findings in academic and practitioner literature, we expect mobile app adopters to increase their spending with the firm more than non-adopters do. However, we find that app adoption has a quantitatively non-negligible and statistically significant negative impact on customer spending (at least 10 percent) across all treatment and control pairs. In particular, these patterns are more evident for customers without elite loyalty membership and customers who tend to use the firm's non-digital channels more often for booking. At the channel level, we also find that the average dollar spent per customer is negatively affected by app adoption in firm-owned channels as well as in third-party channels. However, the share of spending in firm-owned channels is positively affected by app adoption. Our results indicate that marketers may not always gain an increase in customer spending after app adoption. To understand the underlying mechanisms for our findings, we conducted a survey with the firm's and competitors' app adopters. The survey results suggest that customers who adopt the focal app are also most likely to adopt competitors' apps and to shop around using the apps, thereby becoming less loyal to the focal firm and possibly decreasing their share of wallet. An additional difference-in-differences analysis of customers who adopt the app and book rooms with the firm reveals that those who used the apps not only for searching and but also for mobile check-in had a lower negative impact as compared to those who used the apps just for searching. This may suggest the firm can attenuate the negative impact by encouraging customers to use the full functionality of the mobile apps rather than use it just for searching.

Our research relates to a broad body of literature on mobile marketing. A prominent number of prior studies have focused on mobile advertising (e.g., Luo et al. 2013; Bart et al. 2014; Danaher et al. 2015; Fang et al. 2015; Fong et al. 2015; Andrews et al. 2015) and various customer behavior in the mobile context such as user content generation (e.g., Ghose and Han 2011) and word-of-mouth (e.g., Burtch and Hong 2014). Ghose (2017) recently provides a good summary of researches related to the mobile topics. Our paper primarily contributes to an emerging stream of research on mobile applications and differentiates from prior research in several ways. First, our findings suggest a strikingly negative effect of app adoption on customer spending, which contrasts with prior research and thus is worthy of further investigation. Most importantly, we point out the highly competitive nature of the mobile app market for firms could be a reason for such impact. Also, we test the impact of mobile adoption over time by customer cohorts who adopted the app at different points of time. Therefore, we are able to test whether this effect is homogeneous across cohorts as the mobile app quality and functionality evolves over time. From a methodological viewpoint, we exploit the sequential nature of adoption of the app by different cohorts to generate treatment-control pairs and performance several robustness checks to validate the results.

Finally, the firm studied in our research generates a portion of its revenue from a number of third-party channels such as global distribution systems and online travel agencies to whom they have to pay commissions for the bookings. Mobile apps provide a direct channel for customers to be in contact with the firm, and thus afford customers the option to choose the direct channel over third-party channels. We show that mobile app adoption has a differential impact on spending in firm-owned channels and spending through third-party channels. In fact, the introduction of a mobile app impacts spending in third-party channels more negatively than in firm-owned channels. This suggests that mobile channel has positive spillovers on other firm-owned channels because of the synergistic aspects of mobile, and thus firms can still benefit from the mobile app by strengthening the relative share of its own channels against the third-party channels.

1.2 Research Background

Many firms have been allocating an increasing share of marketing budget towards the development and maintenance of their mobile apps. Customers can use mobile apps to search for information, read reviews, make purchases, review orders, and manage personal shopping accounts. In addition, many firms have been providing post-sale service and other customer service through mobile apps. Brands expect that these app investments will translate into enhanced shopping and consumption experiences, and thereby increased customer purchases and retention. There is indeed strong evidence for this expectation. Recent research into the impact of mobile apps show that a mobile app is an important component of an omni-channel experience, providing synergistic value to customers at all stages of the customer journey (Boyd et al. 2017).

Prior research has reported extensively on the positive impact of mobile app on individual spending. Dinner et al. (2018) uses data from a clothing and apparel retailer that operates primarily online and has a single brick-and-mortar store. They show that increased access to mobile apps increases both online and offline purchases, with online sales benefiting more from the app. Similarly, Kim et al. (2015) find, using data from a coalition loyalty program, that mobile app adoption as well as continued app usage increase customer spending. However, the mobile apps do not serve as a full-fledged shopping channel in either of these two studies. Instead, the apparel retailer's app only allows customers to purchase one specific product that is promoted in each week, and the loyalty program app does not incorporate a purchase functionality.

In other studies more relevant to our research, Wang et al. (2015) recognize the positive impact of the introduction of mobile app on purchase incidence for an Internet-based grocery retailer. Similarly, Huang et al. (2016) focus on an e-retailer that introduced a mobile shopping app in addition to its incumbent web channel. They find that app adoption increases customer total purchase across all channels, while slightly cannibalizing sales in the Internet channel. In the B2B context, Gill et al. (2017) find that apps increase customer engagement and overall sales. Focusing on a video game retailer, Narang and Shankar (2019) show that adoption of mobile apps led to a 24 percent increase in customer purchases for the retailer over a year.

Furthermore, extant literature on multichannel marketing also supports a firm's expectations about growing revenue through mobile app launches. Research shows that customers who shop in multiple channels contribute higher profits to the firm than those who do not (Venkatesan et al. 2007). In a study by Montaguti et al. (2015), the authors conduct a randomized field experiment with a multichannel book retailer and show using a propensity score matching method that customers who use both online and offline channels are more profitable than they would be if they used a single channel. Likewise, Li et al. (2015) find that online channel adoption increases purchases of light buyer segments. Viewing the multichannel impact from a different perspective, Konuş et al. (2014) demonstrate a negative impact on sales revenues when the firm eliminated their catalog, which served as a search channel, indicating the synergistic effect of multiple channels. Results from these studies support the premise that customers will increase their total spending if they were to adopt an additional channel such as a mobile app.

However, introducing a mobile app can be a double-edged sword for firms. We argue that launching a mobile app is not a sufficient condition to increase customer purchases and spending. First, if a firm's mobile app does not possess a superior design and quality, then customers may be dissatisfied with their usage experiences. In particular, customers adopting the mobile app expect to use the app frequently and derive benefits. If the app does not meet their quality expectations this could lead to dissatisfaction. For instance, Jumio (2015) find the top reasons why smartphone users fail to complete a transaction or register an account are: 1) slow loading speed (36 percent); 2) difficulty in navigation (31 percent); 3) difficulty in transacting on a small screen (28 percent); and 4) complicated payment process (22 percent). More seriously, 34 percent of customers do not attempt to repeat the transaction after failing to complete it on the first try. This negative effect may not only impact the spending in the mobile channel negatively, but could also spillover to other channels and decrease total spending.

Another factor contributing to the firms not realizing the positive impact of mobile app adoption is the level of competition that may exist in the mobile environment. Mobile app adopters seldom adopt just one app in a given category, especially if they spend significant amount of money in that category. In a highly competitive market of mobile apps provided by competing brands, customers may adopt several competitors' apps. Therefore, customers are more likely to shop around, which in turn may decrease their spending at a particular firm reducing their share of wallet. Court et al. (2017) provides empirical evidence on this issue. The authors highlight that mobile shopping apps and new technologies have changed the customer decision journey at the product consideration and purchase stages. Among the thirty product categories they studied, they find that only three were loyalty driven, with customers repurchasing the same brand rather than shop around. Therefore, from the firm's perspective, the competitive nature of the app environment in the product category could be a limiting factor on the increase on total spending they could expect with app launch.

In summary, there could be several reasons as to why the positive impact of mobile app adoption may not be realized. First, the quality of app may not be good. Second, even if the app quality is good, it may not be as good as competitors' apps. And finally, even if the app is quite competitive in quality, the fact that customers may shop around using multiple apps may lead to reduced spending at the focal firm.

In the following sections, we examine these issues in detail using our data.

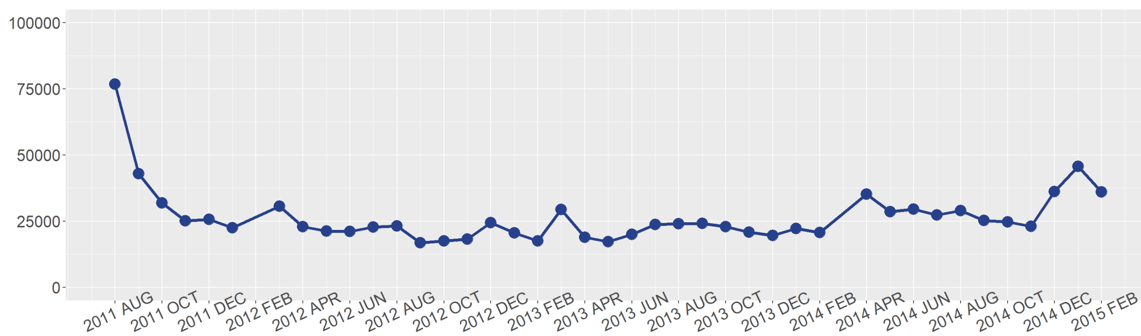
1.3 Data Description

We use a data set provided by a leading international hotel chain based in the U.S. This data set has detailed information on customers' hotel reservations, mobile app usage, and demographics. Our sample period covers 47 months of data from January 2011 to November 2014. Our focal firm launched its mobile app in August 2011, allowing customers to search for hotel rooms, make reservations, check upcoming reservations, and manage rewards membership accounts. Over our sample period, the firm did not provide any monetary incentive such as promotions or discounts for customers to adopt the mobile app. We observe the date when a customer adopted the mobile app, and all customers in our dataset adopted the hotel's mobile app before February 2015.

The app was improved over time with a series of version releases and a variety of functions added to the app. In particular, the firm announced a major update to its app in June 2013, by introducing the mobile check-in feature. This service feature allows customers to check-in and check-out of the hotel using the firm's mobile app. When the room is ready, the customer is automatically notified by an in-app alert, and thus can avoid waiting in line at the front desk by picking up her key at a mobile check-in kiosk. At the end of her stay, the customer can check-out via the app and receive a copy of the bill by email, again, without waiting in line at the front desk. The firm aims to use the mobile check-in service as an important feature for enhancing their customers' stay experience. In our data set, we observe whether each customer used the mobile check-in function in each month and the number of times the customer used the app for searching in each month.

In order to examine the effect of mobile app adoption on customer spending, we specifically focus on existing customers of the focal firm rather than including customers acquired during our sample period. This restriction ensures that our analysis is based on comparisons between existing customers instead of comparisons between existing customers and new customers for whom we do not have historical pre-adoption data. Figure 1.1 shows the number of existing customers that adopted the app in each month. The total number of existing customers who adopted the mobile app before February 2015 is 1,085,609. A large group of customers adopted the app in the first few months of app launch.

Figure 1.1: Number of App Adopters per Month



Notes: The values for February 2012 and March 2014 are missing.

We define six different customer cohorts on the basis of the timing of their app adoption: August 2011, April 2012, January 2013, October 2013, June 2014, and February 2015. We construct our sample by randomly drawing 10,000 customers from all adopters in each of the six cohorts, adding up to a total number of 60,000 customers. Table 1.1 gives the mean value of the demographic variables of customers in each cohort. In general, customers from different cohorts are similar in terms of age, gender and length of member tenure. The reason that we define customer cohorts based on the timing of app adoption is that we aim to identify the effect of app adoption by using different customer cohorts as the treatment group and the control group. We select the six cohorts in such a way that there is sufficient time gap between them and sufficient heterogeneity across cohorts based on the month of adoption that we are considering, ruling out the possibility that our estimated effect could be biased due to seasonality. In addition, the six selected months are spread over the entire observed time period of customer app adoption between August 2011 and February 2015 covering the releases of different versions of the app. The details of constructing the treatment group and the control group will be explained in the next section.

Table 1.1: Demographics of Customers in Each Cohort

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Cohort 6**
Adoption Timing	Aug 2011	Apr 2012	Jan 2013	Oct 2013	Jun 2014	Feb 2015
Average Age (years)	46	46	47	47	48	49
Percentage of Female Customers	20%	25%	25%	29%	32%	30%
Average Member Tenure (months)	75	78	80	80	79	81

The 60,000 customers made a total of 1,937,124 reservations between January 2011 and November 2014. The hotel chain generates revenue from both firm-owned channels and third-party channels. We refer firm-owned channels as inner channels, which are operated by the hotel chain. The inner channels include online (i.e., online website), offline (i.e., property and call center), and mobile (i.e., mobile app and mobile website) channels. The hotel chain sets the same price and assortment across its inner channels. The third-party channels (i.e., outer channels) include global distribution system (GDS), which are primarily used by travel agents, online travel agencies (OTA) such as Expedia and Travelocity, and other channels operated by third-party corporations. On average, 63 percent of the reservations were booked directly through the firm-owned channels as opposed to 37 percent through third-party channels. Table 1.2 shows the share of reservations in each channel.

1.3.1 The Trend in Customer Spending

When we examine the trend in the average total spending across all channels and the average spending in inner channels and outer channels, we find that the average

Table 1.2: **Share of Reservation Amount ($) across Channels**

Reservation Channel	Share
Inner Channels	63.18%
- Online Website (Online)	37.53%
- Property and Call Center (Offline)	23.20%
- Mobile App and Website (Mobile)	2.45%
Outer Channels	36.82%
- Global Distribution System (GDS)	22.96%
- Online Travel Agencies (OTA)	0.99%
- Other	12.87%

total spending follows an upward trend over time (see Figure 1.2). This increasing trend in the average total spending may lead the firm to believe that their customer spending is increasing across all cohorts, thus masking the real problem that we find based on the empirical analysis. In addition, the spending in the inner channels is more than that in the outer channels, consistent with the channel-specific share of reservations in Table 1.2. Finally, the figure also shows a clear seasonality effect such that the spending tends to have drop significantly in November and December, which needs to be addressed when we estimate the treatment effect on spending.

We also compare average customer spending before and after their app adoption across each channel in Table 1.3. The average total spending per month before app adoption is $260.10, and increases by $29.28 during post-adoption period. Moreover, the average spending increases in every channel except for the offline channel and OTA channel. However, such increases in spending could be a consequence of a multitude of factors such as promotion campaigns, economic and industry shocks, as well as the adoption of the mobile app. The model free evidence presented in these tables could mask the real impact of mobile app adoption due to the confounding effects. Therefore, we employ a number of formal empirical methodologies to parse out the confounding factors and identify the real impact of app adoption.

Figure 1.2: Customer Average Monthly Spending ($)

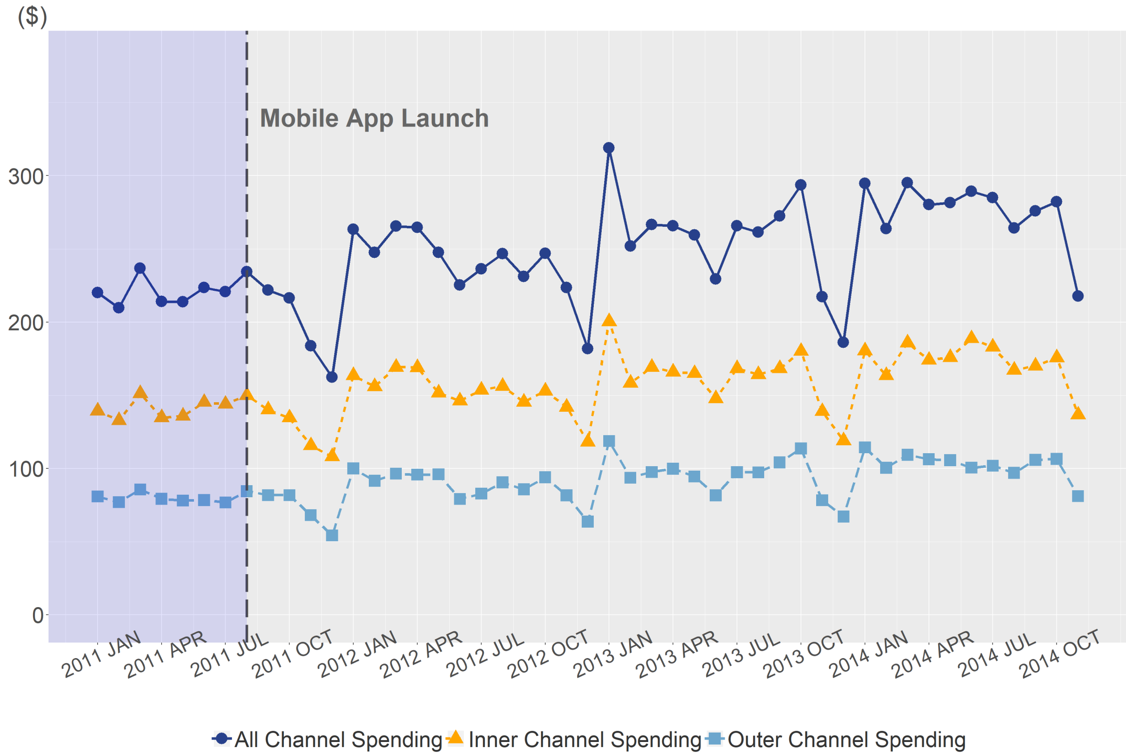


Table 1.3: Customer Average Monthly Spending before and after App Adoption ($)

	Before	After	Difference
All Channels	260.10	289.38	29.28
- Inner Channels	164.94	183.81	18.87
- Online Website (Online)	99.10	113.12	14.02
- Property and Call Center (Offline)	64.74	58.40	−6.34
- Mobile App and Website (Mobile)	1.10	12.29	11.19
- Outer Channels	95.15	105.57	10.42
- Global Distribution System (GDS)	61.08	66.30	5.22
- Online Travel Agencies (OTA)	2.68	2.56	−0.12
- Other	31.40	36.71	5.31

1.3.2 The Trend in Customer Spending by Cohort

When we break our sample down into six cohorts, we observe different patterns in the trend for average total spending for each cohort as shown in Figure 1.3. First, the average customer spending for a cohort reaches its peak when the customers adopt the mobile app. Although the peak of average spending and app adoption seems to occur in the same month, it is not the case that each individual customer within the cohort spent at their highest level when they adopted the app and dropped their spending after adoption. In fact, only 7.3 percent of customers had their highest level of spending in the month of app adoption while 46.6 percent of customers had their highest level of spending after app adoption.

Therefore, the peak of average spending at adoption is mainly due to aggregation bias. To further clarify this, we provide an example consisting of the spending of four real customers from Cohort 1 and their average spending in Figure A1 in the appendix. Although the highest level of the monthly total spending of each customer occurs after app adoption, the highest level of monthly total spending could still occur at the adoption. In fact, the trend in average spending by cohort reveals that customers tend to have high level of spending around the month of app adoption, whereas they may still have a higher level of spending after adoption with these high spending of each customer occurring in different months. Overall, the peak of average total spending implies that, on average, customers have a higher motivation to adopt the app when they make a purchase because they believe they will have a chance to use the app during the stay or later.[1]

Despite the seemingly decreasing trend of spending for each individual cohort, in the aggregate the combined trend could still be increasing as shown in Figure 1.2. This is because the average spending of each cohort that has not adopted the app yet

[1] The adoption takes place when each customer's spending level is high. To rule out the regression-to-the-mean effect, we extends the length of treatment window from one month to five months in our robustness checks and still find a consistently negative impact of app adoption.

Figure 1.3: Customer Average Monthly Spending by Cohort ($)

Notes: The vertical dashed line denotes the month when the cohort of customers adopted the mobile app.

increases successively over time, and thus results in another aggregation bias. Finally, we find a positive relationship between customers' timing of app adoption and their initial average spending level. That is, customers who adopted the app earlier are those who have higher spending levels than the later adopters.

We also summarize the spending of each of the six cohorts before and after app adoption in Table 1.4. Customers who adopted the app earlier spend more (prior to adoption) than those who adopt the app later. In particular, in our data, the earliest adopters spent approximately two times more than those who adopted the apps much later. We observe this positive relationship between timing of app adoption and pre-adoption spending levels for both inner channels and outer channels. In addition, the average spending of customers in Cohort 1 decreases after app adoption, whereas the average spending increases for customers in Cohort 2, 3, 4, and 5 after adoption. Although customers may still have an increasing trend of spending after app adoption, it could be the case that they might have spent even more if they had not adopted the mobile app. Therefore, the impact of mobile app adoption on spending still remains unclear.

1.4 Empirical Analysis

The objective of our research is to estimate the causal effect of customer app adoption on their purchases (total spending). The dependent variable is the individual average monthly spending constructed from the records of hotel reservations. Ideally, we are interested in comparing the spending of app adopters after the adoption to the spending of app adopters if they had not adopted the app. However, the latter is a counterfactual that we cannot observe in reality.

In order to identify the average treatment effect of app adoption, we construct the treatment groups and the control groups using customer cohorts. Next, we estimate the impact of app adoption on customer spending by applying difference-in-differences

Table 1.4: Customer Average Monthly Spending before and after App Adoption by Cohort ($)

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Cohort 6**
Adoption Timing	Aug 2011	Apr 2012	Jan 2013	Oct 2013	Jun 2014	Feb 2015
Pre-Adoption Spending ($)						
All Channels	408.85	263.77	234.25	223.90	169.71	142.84
- Inner Channels	274.76	169.34	147.56	131.17	101.89	84.47
- Share of Inner Channels (%)	70.08	68.36	66.46	62.50	63.94	63.83
- Online Website (Online)	177.65	104.14	82.09	75.91	55.74	42.98
- Property and Call Center (Offline)	94.32	64.19	64.83	54.68	45.67	41.24
- Mobile Website (Mobile)	2.79	1.01	0.64	0.58	0.48	0.25
- Outer Channels	134.09	94.43	86.69	92.73	67.82	58.37
- Global Distribution System (GDS)	87.82	60.67	55.25	59.37	42.28	34.22
- Online Travel Agencies (OTA)	3.51	2.87	2.47	2.71	1.82	2.17
- Other	42.77	30.89	28.97	30.65	23.71	21.97
Post-Adoption Spending ($)						
All Channels	344.98	279.52	284.57	296.73	241.10	–
- Inner Channels	232.45	183.59	180.45	178.43	144.14	–
- Share of Inner Channels (%)	71.15	70.35	67.73	64.32	65.41	–
- Online Website (Online)	148.62	112.72	107.68	110.22	86.37	–
- Property and Call Center (Offline)	68.51	56.73	60.66	57.27	48.83	–
- Mobile Website and App (Mobile)	15.32	14.14	12.11	10.94	8.93	–
- Outer Channels	112.52	95.93	104.12	118.30	96.96	–
- Global Distribution System (GDS)	69.12	58.12	64.57	76.15	63.55	–
- Online Travel Agencies (OTA)	2.26	2.49	2.18	3.30	2.55	–
- Other	41.14	35.32	37.37	38.86	30.85	–

approach, which is an empirical methodology widely applied in the field of marketing (e.g., Chevalier and Mayzlin 2006; Goldfarb and Tucker 2011; Gill et al. 2017). In order to reduce self-selection bias, which is a major concern of many observational research, we use a formal Heckman correction to address selection on unobservables as well as selection on observables. Alternatively, we use propensity score methods to match customers based on their observed attributes to increase the comparability between customers in the treatment group and the control group. We obtain consistent results across different analyses. The roadmap of our empirical analyses and robustness checks are summarized in Table 1.5.

Table 1.5: Roadmap of Empirical Analysis

Section	Approach	Objective
4.1	Selection of cohorts across time	• To construct multiple pairs of the treatment group and the control group. • To control for unobserved time-invariant factors impacting app adoption decision. • To take early and late adopters as well as app upgrades into account.
4.2	Difference-in-Differences (DID)	• To control for observed and unobserved time-varying factors impacting purchase that are common to all customers (e.g., seasonal effects). • To control for observed and unobserved time-invariant factors impacting purchase that vary across customers (e.g., business traveler). This reduces part of potential bias due to self-selection problem.
4.3	DID with Heckman correction	• To fully control for self-selection problem.
4.4	DID with propensity score matching	• To increase the comparability between customers who adopt the app and who have not adopted the app yet.

1.4.1 Treatment and Control Groups

We have six customer cohorts that have adopted the app one time or another. For each cohort and their corresponding month of app adoption, we compare them with customers in the other cohorts that have not adopted the app yet in that particular month. For example, customers in Cohort 1 adopted the app in August 2011 while the other cohorts had not adopted yet at that time. Therefore, we can match customers in Cohort 1 with those in Cohorts 3, 4, 5, and 6, with Cohort 1 being the treatment group and the other cohorts being the control groups. Similarly, when we use Cohort 2 as the treatment group, Cohorts 4, 5, and 6 are used as the control groups. By comparing a treatment group to a control group that ultimately adopted the app later in time, we account for in our model the time invariant unobservables that impact adoption of the app (see Table 1.5).

We do not pair a treatment group (e.g., Cohort 1) with its closest cohort (e.g., Cohort 2) because the time gap between the two corresponding months of app adoption (e.g., August 2011 and April 2012) is quite short that the post-treatment period is too close to the adoption month of the control cohort. This may result in biased estimation of the treatment effect as the spending of customers in the control group could be affected by their own app adoption decision, and thus the control group becomes disqualified as a good counterfactual for the treatment group. Overall, we construct 10 pairs of the treatment group and the control group: $(1,3)$, $(1,4)$, $(1,5)$, $(1,6)$, $(2,4)$, $(2,5)$, $(2,6)$, $(3,5)$, $(3,6)$ and $(4,6)$.

1.4.2 Difference in Differences

We employ the difference-in-differences (DID) approach, which essentially compares the change in spending of customers in the treatment group before and after app adoption to the change in spending of customers in the control group (Angrist and Pischke 2008). The change in spending of the control group is an estimate for the counterfactual change in spending of the treatment group in the absence of mobile app adoption. The comparison between these two changes enables us to account for observed and unobserved time-varying factors affecting spending that are common to all customers. A few examples of the unobserved time-varying factors include marketing campaigns such as TV advertisement, economic shocks, and seasonal trends.

Moreover, the DID approach addresses a major methodological concern, that is, customers who chose to adopt the app could be different from customers who chose not to adopt. There could be unobserved factors driving customer app adoption which could also be correlated with their purchase behavior. This so-called self-selection problem arises in our context because customers decided whether they would adopt the mobile app by themselves. For example, business travelers generally travel more frequently than leisure travelers, and thus will spend more on hotel stays. The business travelers may also have been the customers that chose to the adopt mobile app because they have higher requirements for efficiency and control over their trip while "on the go". In this case, the traveler type (business vs. leisure) might drive app adoption as well as correlate with spending, but we do not observe it in the data. By taking the difference between spending before and after app adoption of the same customer, the DID approach controls for time-invariant factors affecting spending that can be either observed or unobserved.

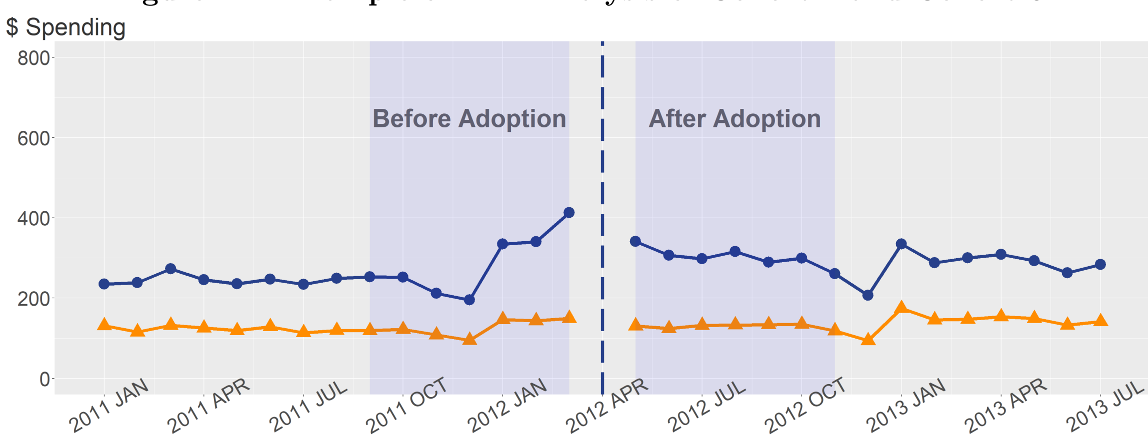
A critical identifying assumption of DID is that the change in spending of customers in the control group should represent the counterfactual of customers in the treatment group. Although it is not possible to test this assumption directly, we can check whether the pre-adoption spending of customers in the treatment group shares a similar time trend with that of customers in the control group. If the spending of customers in the two groups follows a similar trend before app adoption, we would expect the trends of spending to continue developing in a similar way in the post-adoption period if customers in the treatment group had not adopted the app. We evaluate the pre-adoption time trends for the control and treatment customers through a graphical inspection: Figure A2 suggests that the time trends of spending of customers in the treatment and control groups were similar before app adoption, and thus supports the validity of identification for treatment effect using the DID approach.

The difference-in-differences model is formally given by the following specification. For customer i in period t,

$$y_{it} = \beta_0 + \beta_1 \text{adopt}_i + \beta_2 \text{post}_t + \beta_3 \left(\text{adopt} \times \text{post}\right)_{it} + \mathbf{x}_{it}\gamma + \mathbf{w}_i\theta + \nu_i + \epsilon_{it} \tag{1.1}$$

where y_{it} is the log value of spending of customer i in period t. There are two periods: $t = 0$ denotes the period before app adoption of the treatment group, while $t = 1$ denotes the period after app adoption of the treatment group. For each pair of the treatment group and the control group, we define the pre-adoption period as the six months prior to adoption, and the post-adoption period as the six months after adoption. Given that Figure 1.3 exhibits extraordinary spikes of spending in the adoption months, we do not include the month in which the adoption occurred so as to reduce potential bias. Figure 1.4 shows the example of analyzing Cohort 2 and Cohort 6 using the DID approach, in which the shaded areas are the pre-adoption period and the post-adoption period, with a one-month gap in between.

Figure 1.4: Example of DID Analysis of Cohort 2 and Cohort 6



The indicator variable adopt_i takes on the value one if customer i adopted the app in the post-adoption period and zero other wise. The coefficient β_1 captures the average differences in spending between the treatment group and the control group, such as traveler type. post_t is an indicator variable that takes on the value one for the post-adoption period and zero otherwise. The coefficient β_2 captures the average differences in spending between post- and pre-adoption periods, such as the seasonal trends in the travel industry.

The coefficient of interest is β_3, is the difference-in-differences estimate of the average effect of app adoption on spending, given as

$$\begin{aligned}\beta_3 = &\left[\mathbf{E}\left(y_{it}|\text{adopt}_i = 1, \text{post}_t = 1, \mathbf{x}_{it}, \mathbf{w}_i\right) - \mathbf{E}\left(y_{it}|\text{adopt}_i = 1, \text{post}_t = 0, \mathbf{x}_{it}, \mathbf{w}_i\right)\right] \\ &- \left[\mathbf{E}\left(y_{it}|\text{adopt}_i = 0, \text{post}_t = 1, \mathbf{x}_{it}, \mathbf{w}_i\right) - \mathbf{E}\left(y_{it}|\text{adopt}_i = 0, \text{post}_t = 0, \mathbf{x}_{it}, \mathbf{w}_i\right)\right]\end{aligned} \tag{1.2}$$

where β_3 measures the difference in the change in spending over time across the treatment and control groups. ν_i is the individual fixed effect, and ϵ_{it} is the error term which is assumed to be correlated across the two periods for each customer and independent across customers.

In addition, there may be observed attributes that vary across time and are correlated with both customer spending and mobile app adoption; for example, it could be that customers make reservations for luxury hotels, and thus have higher spending. At the same time, they may also decide to adopt the mobile app given that they expect luxury hotels to provide more benefits and services via mobile app than economy hotels would provide. Therefore, we directly control for a number of observed time-varying characteristics to augment the DID models (Gill et al. 2017). $\mathbf{x}_{it}$ represents a vector of reservation characteristics that vary across both customer and time. $\mathbf{w}_i$ is customer i's demographics that do not vary over time. Table 1.6 provides a full list of dependent variables and control variables in the DID model with variable definitions.

Table 1.6: Definitions of Variables in Difference-in-Differences Model

Variable	Definition
Dependent Variables: $\mathbf{y_{it}}$	
Spending	Log value of spending amount ($)
Share of spending	Ratio of the spending of a specific channel to the total spending
Control Variables: $\mathbf{x_{it}}$ & $\mathbf{w_i}$	
<u>Reservation Channel</u>	
Online	Fraction of months in which the customer booked through firm's online website
Offline	Fraction of months in which the customer booked through firm's property and call center
GDS	Fraction of months in which the customer booked through the global distribution system
OTA	Fraction of months in which the customer booked through the online travel agencies
<u>Characteristics of Reservations</u>	
Luxury hotel	Fraction of months in which the customer booked hotels with a luxury brand
Premium hotel	Fraction of months in which the customer booked hotels with a premium brand
Upscale hotel	Fraction of months in which the customer booked hotels with a upscale brand
Economy hotel	Fraction of months in which the customer booked hotels with a economy brand
Franchised hotel	Fraction of months in which the customer booked franchised hotels
Domestic hotel	Fraction of months in which the customer booked hotels located in the U.S.
Weekend	Fraction of months in which the customer booked stays containing Friday and/or Saturday nights
Holiday	Fraction of months in which the customer booked stays containing holidays
Airport	Fraction of months in which the customer booked hotels located in an airport area
Downtown	Fraction of months in which the customer booked hotels located in a downtown area
Resort	Fraction of months in which the customer booked hotels located in a resort area
Suburban	Fraction of months in which the customer booked hotels located in a suburban area
Expressway	Fraction of months in which the customer booked hotels located in an expressway area
Metro	Fraction of months in which the customer booked hotels located in a metro area
<u>Customer Demographics</u>	
Platinum member	Fraction of months in which the customer had a platinum membership
Gold member	Fraction of months in which the customer had a gold membership
Silver member	Fraction of months in which the customer had a silver membership
Basic member	Fraction of months in which the customer had a basic membership
Age	Age of the customer (years)
Female	Dummy variable = 1 if the customer is female, and 0 otherwise
Member tenure	The length of time that the customer has become a rewards member of the firm (months)

1.4.3 Modeling Self-Selection

The major challenge for an unbiased estimation of treatment effect arises from the fact that customers self-selected themselves to adopt the mobile app. The DID approach, based on the way we have chosen the cohorts to compare, has been able to eliminate bias from time-invariant factors and time-varying factors that are common to all customers impacting spending. Additionally, the way we have selected the control cohorts as those who adopt the app at a later time eliminates the time invariant factors that impact app adoption. However, the DID approach may still suffer from an additional source of bias — unobserved factors that vary across both customers and time periods. Therefore, we have to model self-selection formally in order to fully control for any potentially remaining endogeneity bias.

Specifically, we use the Heckman correction (Heckman 1979; Gill et al. 2017; Narang and Shankar 2019), which is described by the outcome equation and the selection equation as follows:

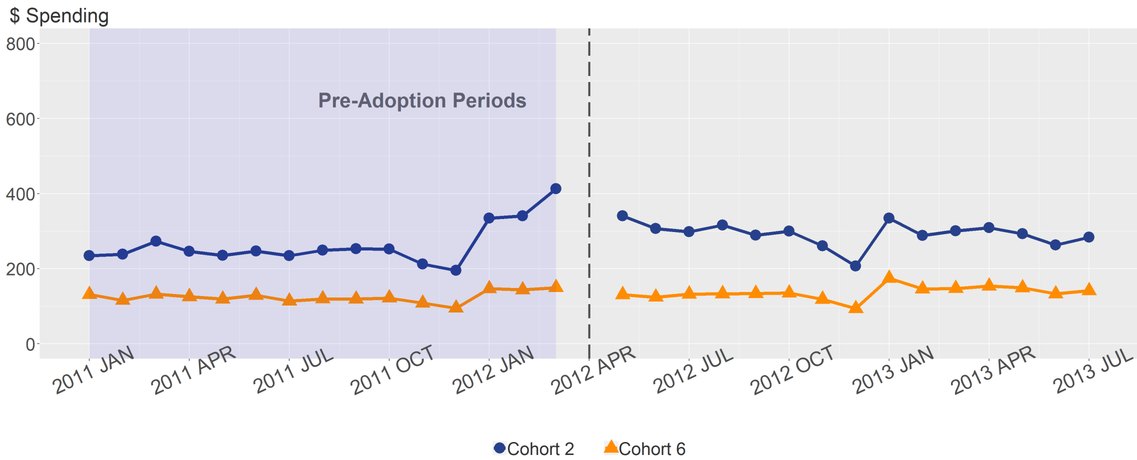
$$y_{it} = \beta_0 + \beta_1 \text{adopt}_i + \beta_2 \text{post}_t + \beta_3 \left(\text{adopt} \times \text{post}\right)_{it} + \mathbf{x}_{it}\gamma + \mathbf{w}_i\theta + \epsilon_{it} \tag{1.3}$$

$$\text{Prob}\left(\text{adopt}_i = 1\right) = \mathbf{z}_i\alpha + e_i \tag{1.4}$$

We use the difference-in-differences specification as the outcome equation, in which we control for the bias due to unobservables which only vary across customers or only vary across time periods. The selection equation is a binomial probit model, which models the customers' decision to adopt the mobile app as a function of observed covariates prior to the adoption, denoted by $\mathbf{z_i}$. Figure 1.5 provides an example of Cohort 2 and Cohort 6 where the shaded area represents the pre-adoption periods of the treatment group. It is important to note that the pre-adoption period of the selection equation is longer than that of the outcome equation.

The control variables in $\mathbf{z_i}$ are constructed as the mean values per month. First,

Figure 1.5: Example of Pre-Adoption Periods of Cohort 2 and Cohort 6



we include a dummy variable indicating whether a customer used the firm's mobile website before. One would expect that customers who used the mobile website are more likely to use the mobile devices during their shopping process and thus are more likely to adopt the app. Second, we include the length of booking window as a predictor for app adoption. For customers who prefer to make travel plans in advance, they may be more likely to adopt the app since it could be a good tool for travel plans. Third, we include a set of variables measuring the share of spending through each shopping channel. One would expect that customers who mostly use the digital channels are more likely to adopt the app. Fourth, Figure 1.3 depicts customers' tendency of increasing their spending prior to adoption. To capture the predictive power of this increased spending prior to adoption, we define a variable: $\Delta\text{Spending} = \text{Spending}_{m-1} - \text{Spending}_{m-4}$, where m is the month when the adoption occurred. Fifth, we include a set of variables describing the reservation characteristics such as the number of room nights, the brand, and type and location of the reserved hotels, which may also improve the prediction of customers' probability of adopting mobile app. Finally, we include several demographic variables including customers' rewards membership, age, and gender.

The error term in the outcome equation, ϵ_{it}, and the error term in the selection equation, e_i, are assumed to be correlated and follow a bivariate normal distribution. We estimate the bivariate normal model by Heckman's two-step estimation approach using the ETREGRESS module in STATA. It is important to note that the system of the two equations is identified using exclusion restrictions. First, the control variables in the outcome equation are constructed using the six-month data prior to the adoption, whereas those in the selection equation are constructed using data during all pre-adoption months, which contributes to the identification. Moreover, the dummy variable of mobile website usage and the length of booking window are two exogenous variables that only appear in the selection equation. Customers who used the firm's mobile website before are more likely to adopt the mobile app as they have such habit of using mobile devices during their shopping process. However, given that below two percent of spending are made through mobile websites, this dummy variable is unlikely to predict customer total purchases. Although the length of booking window may relate to the hotel price a customer will receive and thus relate to customer's spending amount, it should not predict the change in the number of bookings and the number of room nights, which are also used as alternative dependent variables measuring customer purchases.

1.4.4 Treatment Effect with Observable Heterogeneity

Another concern that could complicate our DID identification strategy is that the treatment effect may vary across customers as a function of observed covariates. For example, the impact of app adoption may be more consequential for customers who stay frequently at the focal hotel than for those who stay occasionally. Suppose there are some particularly frequent customers among those who adopted the app (treatments). Problems may emerge for our DID estimates if there are no comparable frequent customers among those who did not adopt the app (controls).

Our task is then to increase the comparability between the treatment group and the control group, and thereby eliminate bias in the estimated treatment effects due to observable heterogeneity. In principle, we would like to construct a control group by finding non-adopters that have similar observed attributes to those of the app adopters. However, the problem of "curse of dimensionality" will arise if the vector of observed attributes $\mathbf{z_i}$ is high-dimensional. Instead of matching on a high-dimensional vector $\mathbf{z_i}$, we use propensity score matching (PSM) methods to match on a scalar — the "propensity score", $P(\mathbf{z})$ (Rosenbaum and Rubin 1983).

The propensity score $P(\mathbf{z})$ is the probability of customer adopting the mobile app in our context. We estimate the propensity score for each observation by estimating the predicted values of the adoption probability from a binomial logit model as a function of customer's historical shopping behavior and individual characteristics.

$$P(\mathbf{z}_i) = \text{Prob}(\text{adopt}_i = 1 | \mathbf{z}_i) = \mathbf{z}_i \alpha + \xi_i \tag{1.5}$$

where $\mathbf{z}_i$ is a vector of pre-adoption covariates, which is identical to what we use in the selection equation of Heckman correction; it captures consumer i's spending level in each channel, reservation characteristics and demographic information. The error term ξ_i follows a standard logistic distribution.

Using the estimated propensity scores $\hat{P}(\mathbf{z})$, we identify the control and treatment customers on the basis of how sufficiently close the scores are across the two groups. As a common practice, we exclude all control and treatment customers whose propensity score lies outside the common support of the propensity scores of the two groups. Then we use a 1:1 matching by finding the nearest neighbor in terms of $\hat{P}(\mathbf{z})$ in the set of non-adopters to each of the adopter. Matches were only accepted if a non-adopter is within 0.01 unit of standard deviation of $\hat{P}(\mathbf{z})$ for an adopter. As a result, we dropped approximately 30 percent of the customers from the original sample. The matched sub-samples are identified using the MatchIt package in R.

The propensity score matching method ensures that all customers are similar in terms of the predicted probability of app adoption based on the observed attributes — customer demographics and purchase behavior prior to the adoption month of the treatment group. We check the quality of propensity score matching by comparing the distribution of $\hat{P}(\mathbf{z})$ of the treatment group and that of the control group before and after matching. By visual inspection, there was negligible difference in the distribution of propensity scores of the two groups after matching (see Figure A3 in Appendix A). We also perform a Kolmogorov-Smirmov test to verify that the distribution of estimated propensity scores of customers from the treatment and the control groups are roughly the same after matching (see Table A2).

1.5 Results

1.5.1 Treatment Effect on Customer Spending

We present the estimated treatment effect of the standard DID model, the DID model combined with Heckman selection equation, and the DID model on matched sample in Table 1.7. The results are for the total spending across all channels as well as spending through the inner channels and spending through the outer channels. Each of the ten columns reports the DID estimates for one pair of the treatment group and the control group.

Surprisingly, we find a very strong negative impact of app adoption on total spending from all model specifications for all pairs of the treatment group and the control group. In particular, the adoption of mobile app is associated with a more than 10 percent negative effect on total spending. Especially for customers in Cohort 4, they would have spent approximately 38 percent more had they not adopted the app. The differential magnitude of the impact of app adoption across cohort pairs could relate to the fact that customers of each cohort downloaded different versions of the focal app when they first adopted it, as the firm frequently updated its app over time. Yet the impact remains consistently negative.

Regarding the effects across shopping channels, app adoption has a significant negative impact on the spending through both inner channels and outer channels. In general, the adoption of mobile app is associated with a 5 percent∼20 percent negative impact on spending of inner channels and a 10 percent∼18 percent negative impact on spending of outer channels. Moreover, app adoption has positive impacts on the share of inner channel spending for customers in all cohorts.

The estimates of the standard DID model and the estimates of the DID model with the Heckman selection equation are particularly close, indicating that we do not achieve much improvement in identification of the treatment effect by using Heckman correction. There are two possible reasons: first, most of the factors related with the self-selection bias are observed and are explicitly controlled for in the DID model; second, a majority of the self-selection bias is due to time invariant unobservables (e.g., business traveler) and time variant unobservables that are common to all customers (e.g., seasonal effects), which has already been accounted for by the DID model. The DID estimates on the matched sub-sample are generally different in magnitude from the estimates obtained from the other two methods. This is mainly because the standard DID model and the DID model with Heckman correction are applied to the full sample, whereas we exclude approximately 30 percent of observations when we construct the matched sub-sample.

We also investigate how the spending in each specific channel is affected by the app adoption. The estimation results are reported in Table 1.8. The effect is different across channels as well as across cohorts. For customers in Cohort 1, app adoption has a negative impact on spending through the firm's online website, properties, call center and the global distribution system. However, spending at the online travel agencies is not significantly affected by app adoption. Similarly, app adoption by customers in Cohort 2, 3 and 4 has a negative impact on spending in every channel,

Table 1.7: Impact of Customer App Adoption on Spending

	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
All Channels										
DID	−0.137***	−0.126***	−0.0613**	−0.101***	−0.310***	−0.222***	−0.188***	−0.230***	−0.195***	−0.377***
	(0.0251)	(0.0254)	(0.0258)	(0.0258)	(0.0258)	(0.0262)	(0.0263)	(0.0261)	(0.0261)	(0.0268)
DID-Heckman	−0.138***	−0.125***	−0.0572**	−0.0979***	−0.306***	−0.217***	−0.183***	−0.227***	−0.189***	−0.379***
	(0.0268)	(0.0268)	(0.0270)	(0.0271)	(0.0284)	(0.0286)	(0.0287)	(0.0290)	(0.0290)	(0.0295)
DID-Matching	−0.146***	−0.113***	−0.0910***	−0.139***	−0.250***	−0.178***	−0.180***	−0.195***	−0.161***	−0.334***
	(0.0293)	(0.0316)	(0.0340)	(0.0350)	(0.0292)	(0.0312)	(0.0324)	(0.0301)	(0.0308)	(0.0309)
Inner Channels										
DID	−0.0626**	−0.0559**	−0.00428	−0.0156	−0.184***	−0.105***	−0.0900***	−0.0927***	−0.0505**	−0.195***
	(0.0252)	(0.0250)	(0.0252)	(0.0251)	(0.0253)	(0.0254)	(0.0254)	(0.0254)	(0.0252)	(0.0258)
DID-Heckman	−0.0649**	−0.0521*	0.00379	−0.0107	−0.177***	−0.0969***	−0.0836***	−0.0866***	−0.0440	−0.196***
	(0.0269)	(0.0267)	(0.0267)	(0.0267)	(0.0277)	(0.0274)	(0.0274)	(0.0277)	(0.0274)	(0.0281)
DID-Matching	−0.0415	0.0148	0.0196	−0.00307	−0.106***	−0.0179	−0.0325	−0.0338	−0.0195	−0.161***
	(0.0296)	(0.0311)	(0.0332)	(0.0336)	(0.0288)	(0.0302)	(0.0311)	(0.0293)	(0.0295)	(0.0296)
Share of Inner Channels										
DID	0.0162***	0.0224***	0.0229***	0.0322***	0.00909*	0.0126**	0.00807	0.0148***	0.0246***	0.00764
	(0.00484)	(0.00515)	(0.00556)	(0.00562)	(0.00528)	(0.00563)	(0.00576)	(0.00555)	(0.00572)	(0.00564)
DID-Heckman	0.0155***	0.0220***	0.0231***	0.0320***	0.0101*	0.0134**	0.00881	0.0154**	0.0249***	0.00661
	(0.00534)	(0.00560)	(0.00587)	(0.00597)	(0.00571)	(0.00603)	(0.00616)	(0.00601)	(0.00619)	(0.00618)
DID-Matching	0.0279***	0.0452***	0.0412***	0.0466***	0.0261***	0.0339***	0.0259***	0.0281***	0.0334***	0.0183***
	(0.00543)	(0.00606)	(0.00680)	(0.00711)	(0.00580)	(0.00645)	(0.00686)	(0.00619)	(0.00659)	(0.00647)
Outer Channels										
DID	−0.140***	−0.139***	−0.126***	−0.154***	−0.159***	−0.132***	−0.106***	−0.171***	−0.169***	−0.180***
	(0.0286)	(0.0285)	(0.0279)	(0.0275)	(0.0288)	(0.0280)	(0.0277)	(0.0277)	(0.0273)	(0.0290)
DID-Heckman	−0.139***	−0.141***	−0.128***	−0.155***	−0.163***	−0.133***	−0.106***	−0.173***	−0.168***	−0.181***
	(0.0330)	(0.0324)	(0.0315)	(0.0314)	(0.0324)	(0.0313)	(0.0312)	(0.0313)	(0.0310)	(0.0326)
DID-Matching	−0.190***	−0.234***	−0.204***	−0.220***	−0.206***	−0.207***	−0.147***	−0.191***	−0.200***	−0.219***
	(0.0329)	(0.0345)	(0.0356)	(0.0359)	(0.0323)	(0.0329)	(0.0335)	(0.0312)	(0.0315)	(0.0329)

Clustered standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.8: Impact of Customer App Adoption on Spending (Multichannel)

	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
Online Website										
DID	−0.0721***	−0.0653***	−0.0454**	−0.0454**	−0.182***	−0.130***	−0.0921***	−0.0804***	−0.0527**	−0.162***
DID-Heckman	−0.0749***	−0.0615***	−0.0356	−0.0381*	−0.176***	−0.121***	−0.0820***	−0.0748***	−0.0437*	−0.165***
DID-Matching	−0.032	−0.0108	−0.00332	0.0222	−0.127***	−0.0516**	−0.0186	−0.0503**	−0.0187	−0.133***
Property & Call Center										
DID	−0.0390*	−0.0444**	−0.0177	−0.0455**	−0.122***	−0.0858***	−0.0751***	−0.0843***	−0.0594***	−0.0957***
DID-Heckman	−0.0382	−0.0447*	−0.0194	−0.0480**	−0.122***	−0.0882***	−0.0794***	−0.0860***	−0.0634***	−0.0944***
DID-Matching	−0.0624**	−0.0703***	−0.0372	−0.0936***	−0.107***	−0.0878***	−0.0851***	−0.0707***	−0.0681***	−0.102***
Global Distribution System										
DID	−0.0378**	−0.0641***	−0.0359**	−0.0428***	−0.0815***	−0.0824***	−0.0617***	−0.0405**	−0.0450***	−0.0675***
DID-Heckman	−0.0374*	−0.0656***	−0.0387**	−0.0437**	−0.0840***	−0.0845***	−0.0621***	−0.0421**	−0.0433**	−0.0699***
DID-Matching	−0.0614***	−0.115***	−0.106***	−0.0819***	−0.0789***	−0.0805***	−0.0557***	−0.0439**	−0.0438**	−0.0580***
Online Travel Agencies										
DID	−0.00651	−0.00182	−0.00653	−0.00139	−0.0131**	−0.00289	−0.00327	−0.0017	−0.00593	−0.0111*
DID-Heckman	−0.00650	−0.00190	−0.00650	−0.00155	−0.0129**	−0.00291	−0.00329	−0.00171	−0.00622	−0.0109*
DID-Matching	−0.00751	−0.000280	−0.00508	−0.00287	−0.0167***	−0.00347	−0.00768	−0.00300	−0.0142**	−0.0118*

Clustered standard errors are used

*** p<0.01, ** p<0.05, * p<0.1

with spending online having the largest negative impact.

1.5.2 Customer Spending by Segment

In order to have a comprehensive understanding about the impact of customer app adoption on their spending, we estimate the treatment effects by customer segments. We construct three partitions of segments: 1) elite customers who have platinum or gold membership and non-elite customer who have basic or silver membership; 2) heavy spenders whose total spending is among the top 25 percent of all customers and light spenders whose total spending is among the bottom 25 percent of all customers; and 3) customers who have used the firm's digital channels (i.e., online website, mobile website and mobile app) for making reservations and customers who have never used digital channels for making reservations (both prior to and after app adoption).

Overall, we find strong differential impacts of app adoption on total spending across customer segments in each of the cohorts (see Table A1 in the appendix). First, when we segment customers based on rewards membership, we find that non-elite customers in general experience a larger negative impact on their spending than elite customers. Moreover, such differences between elite and non-elite customers are most evident among the earliest adopters. Second, we investigate how the impact of adoption differs across heavy spenders and light spenders. We find that the app adoption has a larger negative effect on light spenders than on heavy spenders, and the effect on the light spenders is consistently significant across all cohorts. Third, we also divide customers into segments based on their bookings using the firm's digital versus the firm's non-digital channels. Irrespective of the cohort considered, customers who use digital channels and those who do not use digital channels have negative impacts on their spending after app adoption, while the non-digital channel users have a much larger negative impact than the digital channel users.

1.5.3 Mobile App Adoption

Table A3 and Table A4 in the appendix report the estimates of the probit regression and the logit regression. The binary outcome takes a value one if the customer adopted the mobile app and zero otherwise. The regressors comprised a wide range of reservation and customer characteristics. The reservation variables include mobile website usage, the length of booking window, number of room nights booked, the amount of spending increase prior to adoption, share of spending through each shopping channel, hotel brand and location, and whether the hotel stays are during weekend and holiday periods. The customer characteristics include rewards membership and demographics.

There are a number of significant explanatory variables of mobile app adoption. The results are consistent with our expectations: 1) customers are more likely to adopt the app if they have used the mobile website before; 2) the likelihood of app adoption is higher when customers increase their spending before adoption; and 3) the likelihood of app adoption is also higher when customers have larger share of spending through the firm's online website, indicating that mobile app adoption is most likely correlated with consumers' past purchase behavior on the online website. The predictive power of reservation characteristics are heterogeneous across different pairs of cohorts. In general, customers booking U.S. hotels and/or booking for weekend stays are more likely to adopt the mobile app. The probability of older and female customers adopting the app is lower compared to younger or male customers.

1.6 Underlying Mechanisms and Managerial Implications

Contrary to conventional wisdom, our findings show that mobile app adoption by customers has a negative impact on their total spending post-adoption. Consequently, it is important to understand the underlying causes or mechanisms driving these results. In this section, we put forth the possible explanations for the negative effect and systematically examine them with additional data and analysis. The summary of all the possible explanations and the corresponding testing methods are summarized in Table 1.9.

1.6.1 Regression-to-the-Mean Effect.

One primary concern about our findings is that the negative impact on spending after adoption is due to the regression-to-the-mean effect. That is, customers tend to return to their average spending level or even a lower level after they spend a large amount of money prior to app adoption. To rule out this possibility, we conduct two additional analyses: (1) we extend the time frame of our sample from six months before app adoption and six months after app adoption to one year before app adoption and one year after app adoption (see Figure 1.6(a)); and (2) we extend the length of the treatment window from one month to five months. That is, we drop out not only the month in which customers adopted the app but also the two months preceding and succeeding that month (see Figure 1.6(b)).[2]

We confirm that adoption of the mobile app has a negative impact on total spending and a positive impact on the share of inner channel spending. The estimated treatment effects are even stronger when we use the extended time frame of two years. The DID estimates are reported in Table A5.

1.6.2 Lower Price.

One might argue that the negative impact on spending is purely an effect caused by decline in price because it is easier for customers to find better deals, lower prices or lower-end hotels by using the app. Although this is not likely to be a major concern

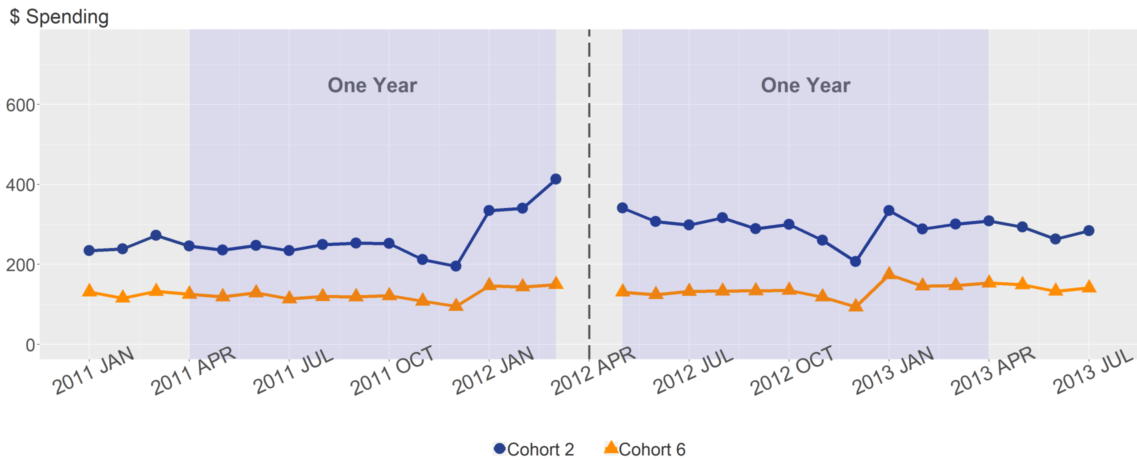
[2] The alternative time frames do not work for some pairs of treatment group and control group in our main analysis as the post-adoption period of the treatment group may cover or be too close to the adoption month of the control group.

Table 1.9: Underlying Mechanisms and Testing Methods

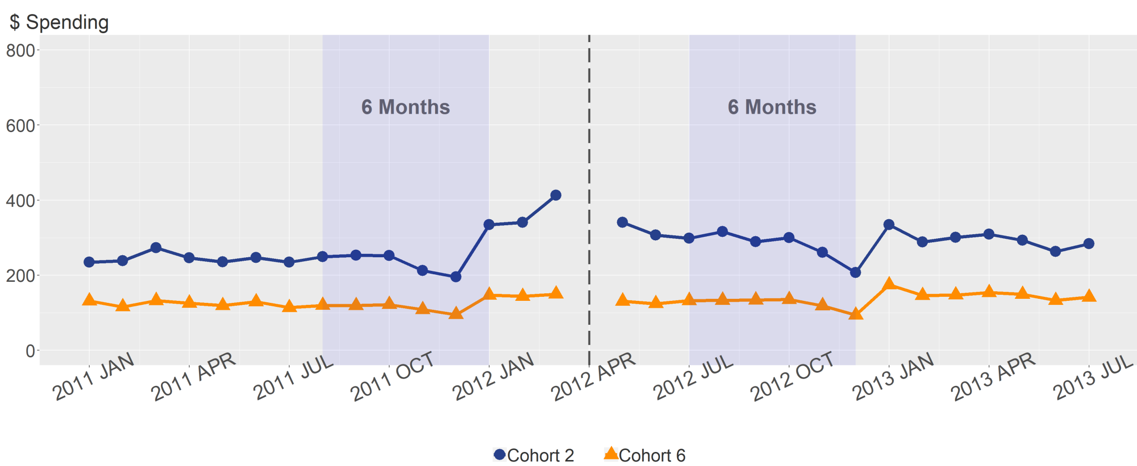
Section	Possible Underlying Mechanisms	Conclusion	Testing Methods
6.1	Regression-to-the-mean effect	No	• Use the five-month period as the app adoption gap window • Use one year before app adoption as the pre-treatment period and one year after app adoption as the post-treatment period
6.2	It is easier for customers to find better deals, lower prices or lower-level hotels using the app	No	• Use alternative dependent variables: the number of room nights and the number of bookings
6.3	Customers may adopt the app and just use it once	No	• Apply analysis to customers who actively use the app after adoption
6.4	The negative effect could be merely driven by customers who increased their spending just before app adoption.	No	• Apply analysis to customers who did increase spending before app adoption
6.5	Model specification problem	No	• Use multi-period panel DID model
6.6	Sample selection bias	No	• Use alternative random sample of customers from the six cohorts • Use alternative customer cohorts
6.7	The mobile app does not provide a good design or high quality	No	• Use multiple treatment groups • Conduct survey for hotel app users • Examine the usage data of the focal app
6.7	Customers who adopted the focal app find mobile apps in general to be good research tools and, therefore, start to use more apps and become more likely to shop around.	Yes	• Conduct survey for hotel app users • Examine the usage data of the focal app

in our context as prices do not vary across inner channels, we repeat the DID analysis with two different dependent variables: the number of room nights and the number of bookings in log values, which are alternative measurements of customer purchase. As shown in Table A6, the impact of app adoption on the total number of room nights and number of bookings is significantly negative for all customers. If the negative impact on spending is merely due to lower price, then we would observe customers making more reservations with the focal hotel at that lower price. However, our results show that customers actually make fewer bookings with the focal hotel than they would have if they had not adopted the app. It further suggests that customers are turning to other hotels because it is less likely that there was demand shrinkage

Figure 1.6: Examples of Extensions of Time Frame and Treatment Window



(a) Extending Time Frame to Two Years



(b) Extending Treatment Window to Five Months

among all customers after app adoption.

1.6.3 App Adopter May not Use the App Actively.

In the main analysis, we do not distinguish whether adopters actively use the app after installing it. This gives rise to the concern that the negative impact on total spending is not attributable to app adoption if customers do not actively use the app after adopting it. We address this concern by restricting our sample to customers who are among the top 50 percent active app users, which is measured by the average number of searches using app in a month. We are able to show that customers in all cohorts have a significant negative impact on their total spending (see Table A6).

1.6.4 App Adopters may Increase Spending Before App Adoption.

Another concern is that a spending spike in the treatment group prior to app adoption could cause the estimated negative effect. We address this concern by comparing non-adopters only to adopters who had no such spending spike. We define adopters without a spending spike as adopters whose ΔSpending is less than the mean value of ΔSpending of non-adopters plus one unit of the standard deviation of it. Again, the results in Table A6 show a consistent negative impact of app adoption on total spending.

1.6.5 Model Specification Problem.

Although we adopted multiple empirical methodologies in our main analysis, it is also of interest to test for the impact using multi-period panel analysis, where month_τ is a set of dummy variables for each month with $\tau = 1, 2, \ldots, T$ and μ_i captures unobserved individual fixed effects.

$$y_{it} = \beta_0 + \beta_1 \text{adopt}_i + \sum_{\tau=1}^{T} \beta_{2,\tau} \text{month}_\tau + \beta_3 \left(\text{adopt} \times \text{post}\right)_{it} \qquad (1.6)$$
$$+ \mathbf{x}_{it}\gamma + \mathbf{w}_i\theta + \mu_i + \epsilon_{it}$$

The estimates for total spending are significantly negative for customers in Cohorts 1 and 4 (see Table A6). It is generally insignificant for other cohorts.

1.6.6 Sample Selection Bias.

In addition, we apply DID analysis to an alternative random sample consisting 4,000 random customers from each of the six cohorts. The alternative sample does not overlap with the sample we use in the main analysis. Furthermore, we apply DID analysis to alternative four customer cohorts as additional treatment groups: 1) alternative cohort 1 adopted the app in December 2011; 2) alternative cohort 1 adopted the app in August 2012; 3) alternative cohort 1 adopted the app in June 2013; and 4) alternative cohort 1 adopted the app in February 2014. We find a negative impact on customer total spending in all of these different samples (see Table A6). This analysis ensures that our findings are not due to bias in sample selection.

1.6.7 App Quality and Shopping-Around Behavior

One of the remaining possible reasons is that adopters find the quality of the app to be poor, or if the absolute quality of the app is good, find it to be not as good as competitors' apps. Given dissatisfaction with app, the customers may not use the app very often, which could have a negative spillover across other channels and, as a result, lower spending with the firm. The other remaining possible reason is that customers who adopted the focal app may find mobile apps in general to be good research tools for comparing hotel accommodations and, therefore, may also have adopted competitors' apps. Customers that have adopted multiple apps are more likely to shop around and shift their spending to competitors. This comparison-shopping could then result in an overall lower share of wallet for the focal firm.

To examine the above questions, we conducted an in-depth survey to measure customer adoption of hotel apps (e.g., Hilton, Marriott) and travel intermediary apps (e.g., Expedia, Booking.com). Specifically, we examined customers of the focal firm who adopted the firm's app at points in time corresponding to the cohorts, and supplemented this sample with adopters of competitor hotel brand apps (some of whom may not have adopted the focal firm's app). We also collected information on customers' app usage behavior, evaluation of the quality of apps, frequency of travel, and demographics. We sent the survey by email in the first week of May 2017 with the help of a market research firm. Overall, we obtained 448 complete responses, with each respondent having at least one hotel app in their smartphone. Around 53 percent of the respondents use the focal hotel app.

Since the respondents adopted the apps over time, we used included retrospective questions to identify which hotel app they adopted the earliest, the hotel app they adopted most recently, and similar questions for travel intermediary apps. The questions also focused on their usage of hotel and intermediary apps for different functionalities, perceptions of app quality, and relative rank of app quality among all hotel and travel intermediary apps (for focal brand customers these questions focused on the focal hotel app, while for non-focal-hotel customers the questions focused on the first hotel app they downloaded).

With regard to the quality perceptions of our focal hotel app as compared to other hotel apps, results in Table 1.10 show that there is no statistically significant difference between the quality rating of our focal hotel app and the quality rating of other hotel apps.[3] Moreover, our focal hotel app has comparable quality ratings with

[3] In the survey, we measure customers' perceptions of app quality by asking them to rate the statement – "Overall, I am very satisfied with XXX app quality" – based on a five-point scale between strongly

other apps when the respondents were asked to compare the focal app to all the other hotel and travel website apps they use. Among the respondents who have the focal hotel app, 86 percent indicated that the focal app was among the best or very good compared with the other hotel app and the travel website apps the respondents have. The corresponding figures for non-focal-hotel app users were lower at 73 percent.[4] In addition, when we plot the usage of the apps among customers in each cohort, we see steady and slightly increasing usage of the apps over time (see Figure 1.7). These results indicate it is unlikely that customers have a poor perception of the quality of the focal-hotel app.

Table 1.10: T-Test: Adopters vs. Non-Adopters of the Focal Hotel App

Variables	**Adopters**	**Non-Adopters**	**Difference**	**t-value**
Rating of App Quality	4.41	4.31	0.10	1.56
Rating of App Quality (comparison)	4.31	4.12	0.19	2.84***
# Apps of Hotel Brands (e.g., Hilton)	2.69	1.55	1.14	9.42***
# Apps of Travel Websites (e.g., Expedia)	3.23	2.27	0.96	4.74***

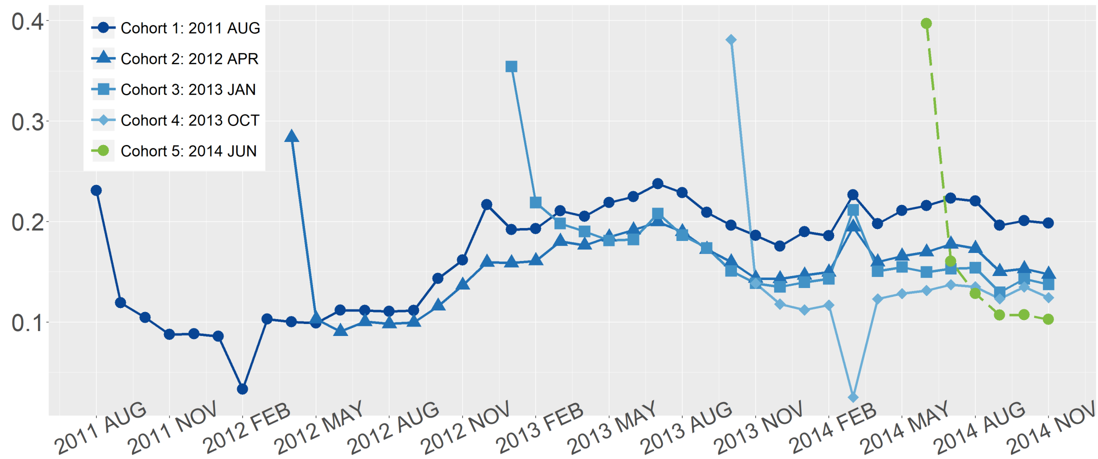
*** p<0.01, ** p<0.05, * p<0.1

Another interesting observation from the survey is that the customers who use the focal hotel app have significantly more hotel apps as well as more travel intermediary apps in their smartphones (on average 1.14 apps more) than those who do not use the focal hotel app. This implies that the focal-hotel app users have more alternatives in their consideration set of apps to use for booking rooms as compared to non-focal-hotel brands. About 24 percent of focal hotel app users have just one hotel app in their smartphone (the focal hotel app). Among the other 76 percent none indicated that the focal hotel app was the first app they downloaded. This could indicate a

disagree and strongly agree. For respondents who adopted the focal app, XXX app is the focal app. For respondents who did not adopt the focal app, XXX app is the first hotel app they adopted.

[4] The survey question is stated as "How is the quality of XXX app compared with other hotel and travel website apps you have?". The respondents were required to choose one option out of five: among the best, very good, average, bad and among the worst. For respondents who adopted the focal app, XXX app is the focal app. For respondents who did not adopt the focal app, XXX app is the first hotel app they adopted.

Figure 1.7: Percentage of Customers Who Use the App per Month



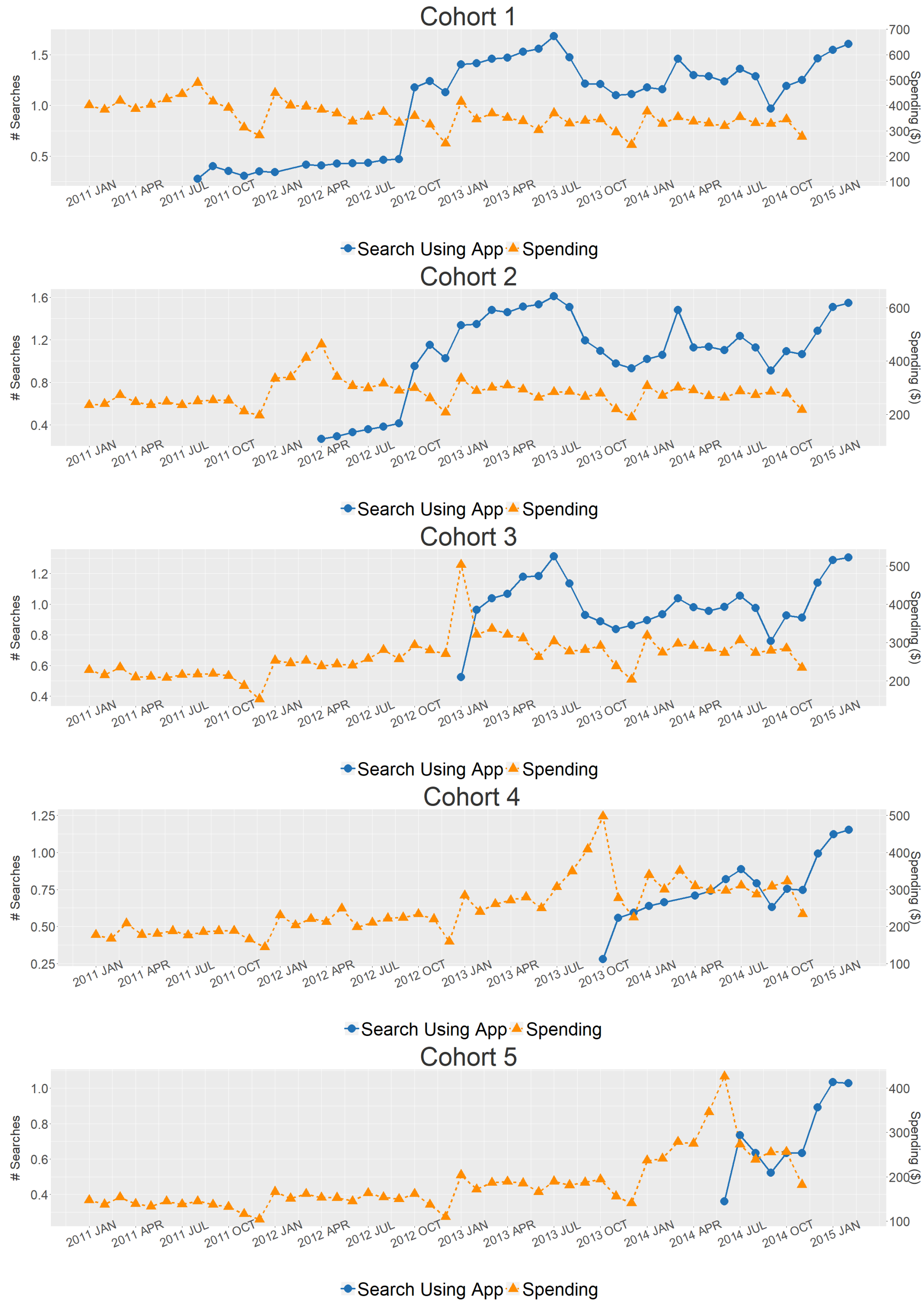
reflection of their preference for the focal hotel app as compared to the others.

1.6.8 Discussion

The results of the survey dispel poor quality of the focal hotel app as the reason for the negative impact of app adoption, in absolute terms as well as in comparison with other apps. While the survey results are an instantaneous snap shot of customer perceptions at a specific point in time, we also look for evidence in the app usage data of customers. A time series graph of focal hotel app usage over time by the different cohorts of customers shows that the usage is steady (see Figure 1.7). Additionally, we find that, in contrast to the trend of average spending, customers keep using the app as a research tool to search for hotel rooms over time. This is reflected in Figures 1.8. Notably, customers who adopted the app earlier make more searches than those who adopted later. This is clearly inconsistent with the explanation of poor app quality as customers would be unlikely to keep using the app if they thought its quality was poor.

Finally, since the apps get updated with new versions over time, it is unlikely that a perception of poor app quality could have remained constant throughout the extended time period of our analysis. Yet the negative impact of add adoption is

Figure 1.8: Average Monthly Number of Searches and Spending by Cohort



consistent across all cohorts. This leads us to examine another possible factor that could impact spending – customers that are comfortable using one app are more likely to use multiple different apps to comparison shop.

Figure 1.8 provides clear descriptive evidence supporting "shopping around" explanation. We observe that the cohorts' search volume increased over time while spending levels at the focal hotel decreased. These results seem to point to the fact that customers are using the app for upper purchase funnel activities such as search; and given that the customers of the focal firm's apps are also likely to possess apps of other hotels and travel intermediaries, it is likely that the focal hotel's customers are also searching and booking hotel rooms on other apps. It is also important to note that the focal firm's app is not usually the first travel apps the customers have downloaded. Users of the focal hotel's app often have other apps in their consideration set. Given that the mobile technology enables ease of shopping and high spenders tend to do more shopping, the impact of this comparison shopping leads to less concentrated booking at any one firm and a corresponding a drop in the share of wallet for the focal firm. Court et al. (2017) highlight that mobile shopping apps and new technologies have changed the customer decision journey at the product consideration and purchase stages. Among the thirty product categories they studied, they find that only three were loyalty driven, with customers repurchasing the same brand rather than shop around. Therefore, from the firm's perspective, the competitive nature of the app environment in the product category leads to less loyalty, more promiscuous shopping across service providers, and more diffuse spending across the different firms.

So what can the firm do? We are not advocating that the firm discontinue its mobile apps. As an increasing number of firms adopt technologies that make it easier for customers to interact and engage with the firm, customers expect to see such technologies provided by all firms. Technologies such as mobile apps have thus become less of a differentiator and more of a threshold cost of doing business. We hypothesize that if the firm had not adopted its mobile apps, the decreasing trend in spending would be much steeper. Rather than eschew mobile apps, we argue that the firm should mitigate the dark side of mobile app adoption by increasing customer engagement with the firm by encouraging customers to fully use all the functionalities of the app.

1.6.9 Impact of App Engagement

Formally, we investigate how mobile app usage behavior and the extent of app engagement impacts customer spending after they have adopted the mobile app. Among those customers who made hotel bookings through any channel after their app adoption, we compare the total spending of those customers who used their mobile app for search only (lower level of app engagement) to the total spending of those customers who used their mobile for search as well as for mobile check-in (higher level of engagement). There is, however, the self-selection issue involved in customers choosing to use the mobile check-in feature. When comparing those customers who used the mobile check-in (the treatment group) to those customers who did not use the mobile check-in (the control group), the estimated effect on customer spending could be due to customer's increased app engagement plus the self-selection bias.

However, not all hotel properties introduced the mobile check-in service, and there is randomness in hotels' choice of introduction (see the appendix). This implies that there are two types of customers among those who did not use the mobile check-in service. On the one hand, there is a group of customers who booked at the hotels with mobile check-in service, but did not choose to use the mobile check-in feature in their app. On the other hand, there is another group of customers who only booked at the hotels without mobile check-in service, and thus did not have the option to use the mobile check-in feature. Some of these customers could have self-selected themselves to use the mobile check-in feature if they had had the option to use it. We use the group of customers who had the option to use the mobile check-in feature but did not use it as one control group (control group 1) and the group of customers who did not have the option to use the mobile check-in feature as another control group (control group 2) to compare the treatment group against.

As explained in the appendix, we leverage the variation in mobile check-in service introduction by the hotel properties to our advantage in isolating the self-selection bias and obtaining an unbiased estimate of the treatment effect of using the mobile check-in feature (i.e., increased level of app engagement) on spending. Given that there are multiple timings of the treatment, we use the difference-in-differences approach in a panel setting (Dranove et al. 2003) for the two comparisons: (1) the treatment group vs. the control group 1; and (2) the treatment group vs. the control group 2.

$$\begin{aligned} y_{it} = & \beta_0^c + \beta_1^c \text{check-in}_i + \sum_{\tau=1}^{T} \beta_{2,\tau}^c \text{month}_\tau + \beta_3^c \left(\text{check-in} \times \text{post}^c\right)_{it} \\ & + \mathbf{x}_{it}\gamma^c + \mathbf{w}_i\theta^c + \sum_{\kappa=1}^{6} \lambda_\kappa^c \text{cohort}_\kappa + \mu_i + \epsilon_{it}^c \end{aligned} \tag{1.7}$$

where y_{it} is the log value of customer i's total spending in period t. The indicator variable check-in$_i$ takes on the value one if customer i ever used mobile check-in service and zero other wise. The indicator variable month$_\tau$ takes on the value one for the period $t = \tau$. The indicator variable post$_t^c$ takes on the value one for the periods after customer i's first usage of mobile check-in and zero otherwise. $\mathbf{x}_{it}$ is a vector of reservation characteristics, and $\mathbf{w}_i$ is customer i's demographics and cohort fixed effects. The indicator variable cohort$_\kappa$ takes on the value one if customer i is from the cohort κ. μ_i is the individual specific error term and ϵ_{it}^c is the serially correlated error term. We also replicate the above analysis by focusing on customers who made hotel bookings through the mobile channel (as against any channel as we did before).

The results (see details in the appendix) show that for the group of customers who made hotel bookings through any channel, the impact of using mobile check-in on spending is significantly positive ($\hat{\beta}_{mc} = 0.0450$, p-value < 0.001). For the group of customers who made hotel bookings through the mobile channel, the corresponding estimate is also significantly positive ($\hat{\beta}_{mc} = 0.0479$, p-value < 0.001). The estimated effects suggest that using mobile check-in service helps increase customer spending. Specifically, encouraging customers to use all the features of the app across all stages of purchase funnel – especially post-purchase functions such as mobile check-in, keyless entry, and mobile service requests while in the hotel – could help in increasing customer engagement with the app and the hotel, thereby reducing the negative impact of app adoption on spending. Such positive impact of increased engagement through the app is consistent with the findings of Gill et al. (2017).

1.7 Conclusion

Many brands use mobile apps as a synergistic new channel for selling products, offering services, and increasing engagement. Adoption of such apps clearly can have an impact on customers' purchase behavior. From the firm's perspective, understanding the nature of this impact is critical, especially since accountability for investments in new technologies has become increasingly important in the digital era. In this paper, we examine this impact in the context of a major international hotel chain, which launched its mobile app in August 2011.

While the conventional wisdom based on recent academic work and practitioner articles is that mobile apps have a positive impact on customer spending and possibly retention, in practice, the revenue benefits may not always be realized. In this research, we highlight a striking case in which the adoption of mobile apps may actually have a negative impact on existing customers' spending. While the result may be a function of service category, the level of competition, and the role of intermediaries in the market, it is important to realize that app adoption can lead to unintended consequences in some cases. However, our results also point out that increasing customer engagement by encouraging customers to fully utilize the app's functionality can mitigate the negative impact of app adoption. From this viewpoint, our paper contributes to the emerging literature on mobile apps by highlighting interesting mechanisms that could lead to negative impact. From the context of a firm, we also show that app adoption impacts the share of inner channels positively, thereby indicating that if designed properly to encourage customer engagement, apps can lead to significant advantages for the firm.

A second and a methodological contribution of our paper is that it provides an illustration of how observational data on customer adoption over time can be purposed in a way that we can tease out the causal impact of adoption. For example, first, we take advantage of the fact that customers adopted the mobile app at different points of time to construct the treatment and control groups that control for the time invariant effects impacting adoption. Second, we control for potential bias from various sources using difference-in-differences approach. Third, we control for any remaining self-selection bias by augmenting the difference-in-differences model with Heckman correction and propensity score matching. Finally, we confirm that the negative impact is robust to alternative time frames, customer spending pattern and app usage behavior, different measures of purchase, panel regression, and alternative random samples.

One of the limitations of our study is that we focus only on the spending of existing customers. One could argue that introduction of new technology may lead to increases in customer acquisition which could potentially have positive impact on revenues. However, establishing whether a potential customer became a customer of the firm just due to the availability of the app might be difficult. We leave this important question for future research.

Chapter 2: A Structural Model of Mobile App Adoption and Multichannel Choice of Purchase

2.1 Introduction

The mobile applications are growing rapidly as a new shopping channel over the past few years. Many firms from a variety of industries have introduced their own branded mobile app to facilitate their customers' shopping journey and improve consumption experiences. As a result, mobile apps have become a major source of e-retail revenues. According to Statistia, there are 28 brands that have over ten million monthly mobile app users in the United States as of July 2018, with the top three apps being Amazon, Walmart and eBay. At the mean time, introducing mobile app raises several critical questions for the firms, most importantly, how mobile apps impact consumers' purchase decisions.

The mobile app is different from the traditional channels including the online shopping websites. One reason is that the mobile app is more flexible than all the other shopping channels, especially when consumers are searching for information. The mobile app can be accessed anytime and anywhere, so it is ideal for early information searches and is more frequently used on the go, in a public location such as in a store, or for impulse purchases. Another reason is that the mobile app can provide new services which cannot be realized through traditional shopping channels. For example, Tiffany introduced an innovative app service called "Ring Finder", which uses the augmented reality technology to enable Tiffany's consumers to see how the engagement rings look like on their fingers. This app service has greatly improved consumer shopping experiences outside Tiffany stores.

Given the benefits of mobile apps, the firms usually expect more purchases from the consumers who adopt their shopping apps because they believe that consumers will be able to engage with the firm more often by using the app, and thus become more loyal. As a result, instead of choosing the competitors' products, consumers will consolidate their purchases with one firm. Prior literature on mobile apps also reports extensively on the positive effect of mobile app adoption on consumer purchases. The authors have shown the positive results in different product categories, such as clothing retailing, grocery shopping, and B2B contexts (e.g., Dinner et al. 2018; Kim et al. 2015; Wang et al. 2015; Huang et al. 2016; Gill et al. 2017; and Narang and Shankar 2019). The literature also provides several theories to explain the positive effect of mobile apps (e.g., Gill et al. 2017). First, the heavy spenders have more purchase occasions, and hence they are naturally more likely to adopt the app and make more purchases with the firm (i.e., self-selection). Second, consumers who adopt the mobile app may receive more marketing messages or different marketing messages through the app, which will also induce them to make more purchases with the firm. Finally, the mobile app can provide more service benefits and shopping convenience for consumers. Therefore, the consumers become happier with the firm, and thus increase their purchases.

In this paper, we investigate the impact of mobile app adoption in a multichannel environment, and therefore we specifically interested in how the introduction of mobile app impacts consumers' choices of shopping channel. There are two types of shopping channels in our research context of a hotel group: the firm-owned channels and the third-party channels. This question is critical here because the travel firms, such as hotel chains and airlines, worry about losing shares from the firm-owned channels to the third-party channels. One reason is that the travel firms need to pay a relatively high percentage of commission fees to the third-party channels for any purchases through them. Another reason is that the travel firms are also concerned that the consumers who use the third-party channels are more likely to be attracted by the competitors' products on the third-party channels. Therefore, the travel firms also care about whether they can attract

There is a vast literature on dynamic programming models in both marketing and economics (e.g., Rust 1987; Erdem and Keane 1996; Song and Chintagunta 2003; Hendel and Nevo 2006; Gowrisankaran and Rysman 2012). In particular, our model builds on Seiler (2013) and Pires (2016), where the authors develop a two-stage model to characterize consumers' search behavior as well as purchase decisions. Following the two-stage structure, we present a dynamic model in which consumers decide whether they adopt a firm's mobile app and which shopping channels they use to make a purchase with the firm in each period. In the first stage, we model consumers' decisions of mobile app adoption based on their app adoption cost and anticipated future app usage. In the second stage, we model consumers' choices of shopping channel, depending on their app adoption decisions. Moreover, our model allows for consumer heterogeneity in their marginal utility related to the competitive apps by using individual-specific coefficients. For the model estimation, we use the Bayesian estimation algorithm proposed by Imai et al. (2009) and Ching et al. (2012) to estimate our model.

We find that there are more consumers making reservations with the focal firm after they adopt the app. This positive effect is larger on the firm-owned channels than the third-party channels. Furthermore, when there are more competitive apps on the market, the adoption cost of the focal firm's app increases, which leads to the competition effect that consumers being less likely to adopt the focal app if there are more options on the market. We also conduct the counterfactual simulations and find that consumers are less likely to make reservations with the focal firm if the firm had not introduced their apps.

2.2 Data

We use a panel of individual-level data of hotel reservations and app adoption. The data are obtained from a leading international hotel group based in the US. We focus on the sample period from January 2011 through November 2014, which constructs for a total number of 47 months. During this time period, our focal firm launched its branded mobile app, offering a variety of basic functions such as accessing the rewards membership accounts, searching for hotels and making reservations. Later in 2013, the firm introduced a brand new feature to the app, i.e., Mobile Check-In. The value proposition of this service is for consumers to check-in and check-out using the branded mobile app. When the room is ready, the consumer will be automatically notified by an in-app alert, and then she can pick up her key at a mobile check-in kiosk without waiting in the long line at front desk. At the end of her stay, she can check out via the app and get a copy of bill by email.

2.2.1 Data Summary

We use a random sample of 4,859 consumers. We observe the date when each consumer adopts the mobile app, and whether she uses the mobile check-in service in each month. Figure 2.1 shows the number of consumers who adopted the app (i.e., who install the app for the first time) in each month, with the average number being 110 consumers per month. In our sample, there are 4,200 consumers who adopted the app between August 2011 and November 2014 and 659 customers who adopted the app after November 2014. It is important to note that the firm did not provide any monetary incentives to encourage consumers to adopt the app. With the number of the app adopters varying over time, 14.5% of consumers adopted the app in the first year of app introduction, and 49.7% of consumers adopted the app before Mobile Check-In service was provided. In addition to app adoption, Figure 2.2 presents the number of consumers who used the mobile check-in service in each month.

Figure 2.1: The Monthly Number of New App Adopters

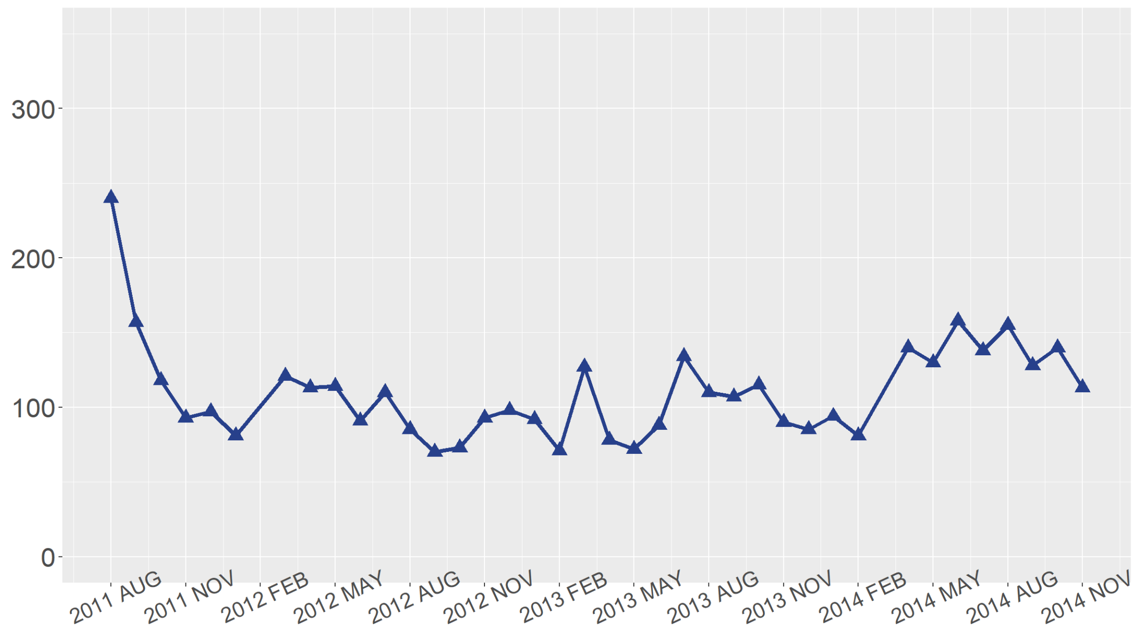
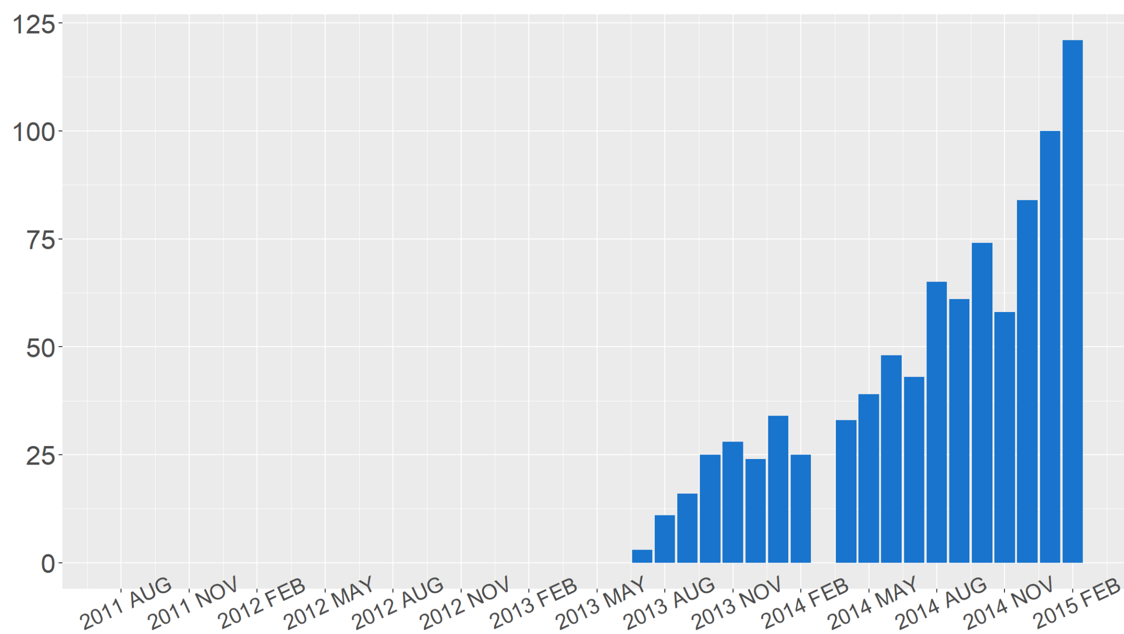


Figure 2.2: The Number of Consumers Who Use Mobile Check-In Service



Besides the mobile app data, we also observe each consumer's reservation information with this hotel group between January 2011 and November 2014. Overall, the consumers made 54,583 reservations during the 47 months. For each reservation, we observe the following information: reservation date, booking channels, hotel features and hotel stay periods. The mobile app, website and offline shopping channels are operated by the hotel group, so we refer those as the firm-owned channels as opposed to the third-party channels including the Global Distribution System (GDS) and the Online Travel Agencies (OTA). In addition, the hotel group sets the same price and assortment across its own shopping channels.

Figure 2.3 presents the evolution of the number of purchasing consumers over time. Despite of the seasonality effects, the figure shows an increasing trend in the number of purchasing consumers in all channels as well as in the firm-owned channels. However, the time trend for the third-party channels is rather flat.

Figure 2.3: The Number of Purchasing Consumers

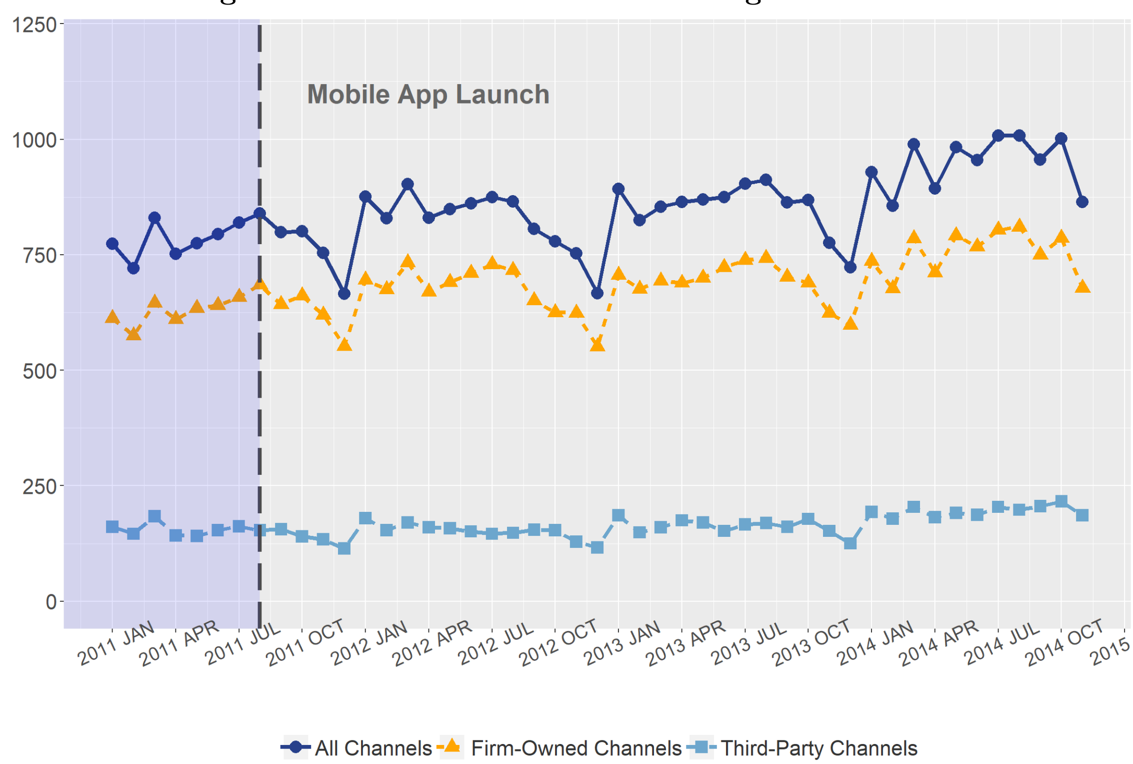


Table 2.1 reports the summary statistics of the variables to be used in our model

estimation. On average, the consumers made 84% of the reservations through the firm-owned channels, and 16% through the third-party channels. A major proportion of booked hotels are upscale brands, which accounts for 46% of all the reservations. In addition, 3.1% of the reservations are luxury hotels, 33.4% are premium hotels, and 17.5% are economy hotels.

Table 2.1: The Characteristics of Hotel Reservations

	Obs.	**Mean**	**S.D.**	**Min**	**25%**	**75%**	**Max**
Firm-Owned Channel	54,583	0.840	0.367	0	1	1	1
Third-Party Channel	54,583	0.160	0.367	0	0	0	1
Luxury Hotel	54,583	0.031	0.174	0	0	0	1
Premium Hotel	54,583	0.334	0.472	0	0	1	1
Upscale Hotel	54,583	0.460	0.498	0	0	1	1
Economy Hotel	54,583	0.175	0.380	0	0	0	1

2.2.2 Descriptive Evidence

In this section we examine the data patterns and provide descriptive evidence which inspires the major design choices of our structural model. We consider the impact of mobile app adoption on the number of reservations through the firm-owned channels and the third-party channels. The regression equation is given by the following specification.

$$Y_{it} = \alpha + X_{it}'\beta + Z_t'\delta + \sum_{j=1}^{J} \theta_j T_{jt} + \mu_i + \epsilon_{it} \tag{2.1}$$

where Y_{it} is the number of reservations booked by consumer i in month t. X_{it} captures customer i's tenure since their app adoption and the reservation characteristics in month t. Z_t includes the number of competitive hotels apps and the number of travel website apps available in the market in month t. T_{jt} is the dummy variable for the month fixed effect. Finally, μ_i is the individual fixed effect, and ϵ_{it} is the error term.

Table 2.2: The Number of Reservations

	(1)	(2)	(3)
Variables	**All Channels**	**Firm-Owned Channels**	**Third-Party Channels**
Tenure of Focal App	−0.00109***	−0.000362*	−0.000732***
	(0.000176)	(0.000188)	(0.000163)
Tenure of Focal App^2	6.25e-06	−8.53e-06*	1.48e-05***
	(3.85e-06)	(4.93e-06)	(4.74e-06)
Competitive Hotel Apps	0.00950***	0.00346	0.00604**
	(0.00207)	(0.00298)	(0.00273)
Travel Website Apps	−0.00882***	−0.000906	−0.00791*
	(0.00333)	(0.00505)	(0.00466)
Month Fixed Effects	Yes	Yes	Yes
Reservation Characteristics	Yes	Yes	Yes
R^2	0.8829	0.7698	0.0985
Observations	226,540	226,540	226,540

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.2 presents the estimation results of the above fixed-effect model. The results suggest that the relationship between the number of reservations and the consumers' tenure of focal app is negative, which is different from the previous research. Another interesting finding is that the relationship between the number of reservations and the number of competitive hotel apps is positive while the relationship between the number of the reservations and the number of travel website apps is negative. The results of the descriptive model suggests that both the focal app and the availability of other apps could have an impact on consumer purchases.

2.3 Model

We model consumer choices about the mobile app adoption and shopping channel as a two-stage process (e.g., Seiler 2013; Pires 2016). In each period t, consumer i decides whether they adopt the mobile app in the first stage, then she makes the choice of shopping channel for hotel reservations with the focal hotel group in the second stage.

The app adoption choice is denoted by d^a_{it}, where $d^a_{it} = 1$ if consumer i chooses to adopt the focal app at time t and zero otherwise. Once consumer i chooses to adopt the app at time t, we assume she does not make the app adoption decision in the future periods. The choice of shopping channel j is denoted by $d^p_{ijt} = 1$ with $j = 0, 1, 2$, where $j = 1$ stands for the firm-owned channels (e.g., firm's website, mobile app and call center), $j = 2$ stands for the third-party channels (e.g., Expedia), and $j = 0$ stands for no purchase with the focal firm. We assume each consumer only chooses one channel in a period, i.e., $\sum_{j=0}^{2} d^p_{ijt} = 1$.

2.3.1 The First Stage: App Adoption

In the app adoption stage, consumer i receives

$$u_{it}^{a} = \begin{cases} -c_{it} + \epsilon_{it,1}^{a}, & \text{if } d_{it}^{a} = 1 \\ \epsilon_{it,0}^{a}, & \text{if } d_{it}^{a} = 0 \end{cases} \tag{2.2}$$

where c_{it} is a one-time cost incurred by app adoption (e.g., limited memory size in smartphone and learning cost). In particular, $c_{it} = c_0 + c_i q_t$ with c_0 being the intercept for the adoption cost, c_i being the individual marginal cost, and q_t capturing the number of competitive hotel apps (e.g., Starwood) and the number of available travel website apps (e.g., Expedia) in the market. We assume c_i's prior distribution is $c_i \sim N\left(c, \sigma_c^2\right)$. For the hyperparameters c and σ_c, we use a normal (diffuse) prior distribution on c and an inverse-gamma (diffuse) prior on σ_c^2.

However, u_{it}^{a} only captures the flow utility in the app adoption stage. When deciding to adopt the app, the consumer anticipates she will have to make the channel choice decision in the same time period. The full expected period utility (i.e., utility from both stages) can be expressed as

$$E\left[u_{it} | d_{it}^{a} = 1\right] = E\left[\max_{d_{ijt}^{p}} \left\{u_{ijt}^{p}\right\}\right] - c_{it} + \epsilon_{it,1}^{a} \tag{2.3}$$

$$E\left[u_{it} | d_{it}^{a} = 0\right] = E\left[\max_{d_{ijt}^{p}} \left\{u_{ijt}^{p}\right\}\right] + \epsilon_{it,0}^{a} \tag{2.4}$$

where u_{ijt}^{p} is the utility in channel choice stage.

2.3.2 The Second Stage: Choice of Shopping Channel

If consumer i decides to choose channel j to make purchases, she receives

$$u_{ijt}^{p} = \delta_{ijt} + u_{it}^{c} + \epsilon_{ijt}^{p} \tag{2.5}$$

which consists of the purchase utility δ_{ijt}, the consumption utility u_{it}^{c}, and the random shock ϵ_{ijt}^{p}. The purchase utility is given by

$$\delta_{ijt} = \begin{cases} \alpha_j, & \text{for } j = 1,2 \\ b_i q_t, & \text{for } j = 0 \end{cases} \tag{2.6}$$

where α_j captures the average preference towards channel j (e.g., usage convenience); and q_t is the the number of competitive hotel apps and the number of travel website apps in the market. Similar as c_i, we assume b_i's prior distribution is $b_i \sim N\left(b, \sigma_b^2\right)$ with b having a normal (diffuse) prior distribution and σ_b having an inverse-gamma (diffuse) prior distribution.

The consumption utility u_{it}^{c} is derived from staying at the booked hotels in period t. No matter whether consumer i makes purchases in period t, she will still receive the consumption utility as long as she has made hotel reservations in previous periods for the current period's hotel stays. $u_{it}^{c} = 0$ if consumer i does not have hotel stays in period t.

2.3.3 The Benefits from App Adoption

After consumer i adopts the app, she receives or loses additional utility by using the app during the purchase process as well as the consumption process.

For the purchase utility, we assume

$$\alpha_j = \begin{cases} \alpha_j, & \text{before app adoption} \\ \alpha_j + \lambda_j, & \text{after app adoption} \end{cases} \tag{2.7}$$

For the consumption utility, we assume

$$u_{it}^c = \begin{cases} u_{it}^{c,h}, & \text{before app adoption} \\ u_{it}^{c,h} + u_{it}^{c,a}, & \text{after app adoption} \end{cases} \tag{2.8}$$

However, consumer i will receive $u_{it}^{c,a}$ only if her reservations are booked through the firm-owned channels because she cannot use the mobile app during her stay if the reservations are booked through the third-party channels. That is,

$$u_{it}^{c,h} = \nu^h z_{it}^h \widetilde{d}_{it} \tag{2.9}$$

$$u_{it}^{c,a} = \nu^a z_{it}^a \widetilde{d}_{i1t} \tag{2.10}$$

where z_{it}^h captures the hotel brand level (e.g, economy, upscale, premium or luxury levels); z_{it}^a includes an intercept term and the availability of mobile check-in in the reserved hotels; $\widetilde{d}_{it}$ is an indicator variable which equals to one if consumer i has hotel stays in period t; $\widetilde{d}_{i1t}$ is an indicator variable which equals to one if consumer i has hotel stays in period t and the hotel stays are booked through the firm-owned channels.

2.3.4 Consumer Dynamic Optimization Problem

Consumer i chooses: (1) a finite sequence of decision rule for mobile app adoption $\{d_{it}^a\}_{t=t_0}^{t_i}$, where t_0 is the time period when the focal firm launched its mobile app, and t_i is the time period when consumer i decides to adopt the app, which is endogenously determined; and (2) an infinite sequence of decision rule for hotel reservation channel $\left\{d_{ijt}^{p}, \forall j\right\}_{t=0}^{\infty}$.

The objective of consumer i is to maximize her expected present value of utility over the entire future horizon conditional on the initial state variables:

$$\max_{\{d_{it}\}_{t=0}^{\infty}} E\left[\sum_{t=0}^{\infty} \beta^{t} U_{it}\left(d_{it}\right) \middle| x_{i0}, \epsilon_{i0}\right] \tag{2.11}$$

β is a discount factor, d_{it} is the set of choice variables $\left\{d_{it}^{a}, \left\{d_{ijt}^{p}\right\}_{j=0}^{2}\right\}$, x_{it} is the set of observed state variables $\left\{q_t, z_{it}^{h}, z_{it}^{a}\right\}$, and ϵ_{it} is the set of unobserved state variables $\left\{\epsilon_{ijt}^{p}, \epsilon_{it}^{a}\right\}$. We assume ϵ_{it} are independently and identically distributed extreme value type I.

$$U_{it}\left(d_{it}\right) = \begin{cases} \sum_{j=0}^{2} d_{ijt}^{p}\left(\delta_{ijt} + \epsilon_{ijt}^{p}\right) + u_{it}^{c,h}, & \text{if } t < \bar{t} \\ \sum_{j=0}^{2} d_{ijt}^{p}\left(\delta_{ijt} + \epsilon_{ijt}^{p}\right) + u_{it}^{c,h} + \epsilon_{it,0}^{a}, & \text{if } d_{it}^{a} = 0 \\ \sum_{j=0}^{2} d_{ijt}^{p}\left(\delta_{ijt} + \lambda_j + \epsilon_{ijt}^{p}\right) + u_{it}^{c,h} + u_{it}^{c,a} - c_{it} + \epsilon_{it,1}^{a}, & \text{if } d_{it}^{a} = 1 \\ \sum_{j=0}^{2} d_{ijt}^{p}\left(\delta_{ijt} + \lambda_j + \epsilon_{ijt}^{p}\right) + u_{it}^{c,h} + u_{it}^{c,a}, & \text{if } t > t_i \end{cases} \tag{2.12}$$

Let $\theta = (\theta_h, \theta_c, \theta_i)$ be the vector of all parameters, $\theta_h = (c, \sigma_c, b, \sigma_b)$ be the vector of hyper-parameters, $\theta_c = \left(c_0, \alpha, \lambda, \nu^{h}, \nu^{a}, \beta\right)$ be the vector of common parameters, and $V(x_{it}, \epsilon_{it}; \theta)$ be the value function:

$$V(x_{it}, \epsilon_{it}; \theta) = \max_{\{d_{i\tau}\}_{\tau=t}^{\infty}} E\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} U_{i\tau} \middle| x_{it}, \epsilon_{it}\right] \tag{2.13}$$

In the infinite horizon dynamic programming problem, the value function and the policy function do not depend on time or individual. So we can eliminate the time and individual subscript. This infinite horizon dynamic programming problem can be solved through the Bellman Equation:

$$V(x,\epsilon;\theta) = \max_{d} \left\{ U + \beta E_{x',\epsilon'} \left[V(x',\epsilon';\theta) | x, \epsilon \right] \right\} \tag{2.14}$$

2.3.5 Choice Specific Value Function

The dynamics of the problem is clear when we look at the choice specific value functions. Before the focal firm launched its mobile app, the choice specific value function in the channel choice stage is the current period's utility plus the discounted expected value of the next period.

$$v_j^{\mathrm{bf}}(x,\epsilon;\theta) = \delta_j + \epsilon_j^p + u^{c,h} + \beta E \left[\max_{d^{p'}} \left\{ \sum_{j=0}^{2} d_j^{p'} \cdot v_j^{\mathrm{bf}}(x',\epsilon';\theta | x, \epsilon, d) \right\} \right] \tag{2.15}$$

After the app launch, consumers will decide whether to adopt the app in the app adoption stage. We model consumers as forward-looking individuals, who trade off the cost of app adoption with the benefits from future app usage. The choice specific value function in the app adoption stage if the consumer chooses not to adopt the app is the expected value from the optimal option from the next stage plus the error term.

$$v_{na}(x,\epsilon;\theta) = E \left[\max_{d^p} \left\{ \sum_{j=0}^{2} d_j^p \cdot v_j(x,\epsilon;\theta | d^a = 0) \right\} \right] + \epsilon_0^a \tag{2.16}$$

When consumer i chooses to adopt the app, the choice specific value function is the expected value from the optimal option from the next stage plus the error term minus the adoption cost.

$$v_a(x,\epsilon;\theta) = E \left[\max_{d^p} \left\{ \sum_{j=0}^{2} d_j^p \cdot v_j(x,\epsilon;\theta | d^a = 1) \right\} \right] - c + \epsilon_1^a \tag{2.17}$$

The value function in the channel choice stage depends on the app adoption decision in the previous stage. Before app adoption, the choice specific value function in the channel choice stage is specific to a choice of channel j.

$$v_j(x,\epsilon;\theta) = \begin{cases} \delta_j + \lambda_j + \epsilon_j^p + u^{c,h} + u^{c,a} + \beta E\left[\max\limits_{d^{p'}}\left\{\sum\limits_{j=0}^{2} d_j^{p'} \cdot v_j^{\text{af}}(x',\epsilon';\theta|x,\epsilon,d)\right\}\right] & \text{if } d^a = 1 \\ \delta_j + \epsilon_j^p + u^{c,h} + \beta E\left[\max\limits_{d^{a'}}\left\{v_{na}(x',\epsilon';\theta|x,\epsilon,d),\ v_a(x',\epsilon';\theta|x,\epsilon,d)\right\}\right], & \text{if } d^a = 0 \end{cases} \tag{2.18}$$

After app adoption, the choice specific value function in the channel choice stage is the current period utility plus the discounted expected value of the next period.

$$v_j^{\text{af}}(x,\epsilon;\theta) = \delta_j + \lambda_j + \epsilon_j^p + u^{c,h} + u^{c,a} + \beta E\left[\max_{d^{p'}}\left\{\sum_{j=0}^{2} d_j^{p'} \cdot v_j^{\text{af}}(x',\epsilon';\theta|x,\epsilon,d)\right\}\right] \tag{2.19}$$

Let $w(x;\theta) = v(x,\epsilon;\theta) - \epsilon$. The value function can be expressed in terms of their choice-specific counterparts

$$V_p^{\text{bf}}(x,\epsilon;\theta) = \max_{d^p}\left\{\sum_{j=0}^{2} d_j^p \cdot \left[w_j^{\text{bf}}(x;\theta) + \epsilon_j^p\right]\right\} \tag{2.20}$$

$$V_a(x,\epsilon;\theta) = \max_{d^a}\left\{w_{na}(x;\theta) + \epsilon_0^a,\ w_a(x;\theta) + \epsilon_1^a\right\} \tag{2.21}$$

$$V_p(x,\epsilon;\theta) = \max_{d^p}\left\{\sum_{j=0}^{2} d_j^p \cdot \left[w_j(x;\theta) + \epsilon_j^p\right]\right\} \tag{2.22}$$

$$V_p^{\text{af}}(x,\epsilon;\theta) = \max_{d^p}\left\{\sum_{j=0}^{2} d_j^p \cdot \left[w_j^{\text{af}}(x;\theta) + \epsilon_j^p\right]\right\} \tag{2.23}$$

Finally, the expected value function is

$$E\left[V(x',\epsilon';\theta)|x,\epsilon,d\right] = E_{x',\epsilon'}\left[V(x',\epsilon';\theta)|x,\epsilon,d\right] = E_{x'}\left[E_{\epsilon'}\left[V(x',\epsilon';\theta)\right]|x,\epsilon,d\right] \tag{2.24}$$

We refer to $E_\epsilon\left[V(x,\epsilon;\theta)\right]$ as the $E_\epsilon max$ function. To simplify notation, we define

$W(x;\theta) = E_\epsilon [V(x,\epsilon;\theta)]$. The assumption that ϵ's are i.i.d. extreme value type I distributed implies

$$W_p^{\text{bf}}(x,\epsilon;\theta) = \ln \left[\sum_{j=0}^{2} \exp \left\{ w_j^{\text{bf}}(x;\theta) \right\} \right] \tag{2.25}$$

$$W_a(x,\epsilon;\theta) = \ln \left[\exp \{ w_{na}(x;\theta) \} + \exp \{ w_a(x;\theta) \} \right] \tag{2.26}$$

$$W_p(x,\epsilon;\theta) = \ln \left[\sum_{j=0}^{2} \exp \{ w_j(x;\theta) \} \right] \tag{2.27}$$

$$W_p^{\text{af}}(x,\epsilon;\theta) = \ln \left[\sum_{j=0}^{2} \exp \left\{ w_j^{\text{af}}(x;\theta) \right\} \right] \tag{2.28}$$

2.4 Estimation

2.4.1 Identification

The identification of the parameters is standard. First, c_0 is the average adoption cost of the app, which is unobserved to the researchers. It is identified by the data pattern that consumers having a general tendency of delaying app adoption over time. Intuitively, the app adoption cost may capture the learning cost of installing and using the app, the fact that consumers having limited space in their smartphones, and etc.

Second, α_j is the average preference towards channel j relative to the outside option. It is identified by the variation in consumers' choices of different channels. Third, λ_j is the average change in preference towards channel j after app adoption. It is identified by the variation that consumers changing their channel choice patterns after they adopt the app. Fourth, ν is the marginal consumption utility from hotel characteristics and app features. It is identified by the variation in consumers' reservations of hotels with different characteristics.

Moreover, c_i is the individual marginal cost related to the availability of the competitive hotel apps and the travel website apps. It is identified if consumers have similar reservation records (i.e. similar benefits), but choose to adopt at different time periods as the availability of other apps change over time. b_i is the individual marginal utility of the outside option related to the availability of the competitive hotel apps and the travel website apps. It is identified if consumers have similar purchase utility from the firm-owned channels and the third-party channels, but some consumers choose to make a purchase while others do not. Finally, β is the discount factor, which is identified by the sequence of consumers' purchase choices.

2.4.2 Likelihood

The likelihood function is $L\left(d|x;\theta\right) = \prod_{i=1}^{I} L_i\left(d_i|x_i;\theta\right)$, where the individual likelihood function for consumer i is given by

$$L_i\left(d_i|x_i;\theta\right) = \left[\prod_{t=0}^{t_0-1} \Pr\left(d^p_{ijt}|x_{it};\theta\right)\right] \cdot \left[\prod_{t=t_0}^{t_i-1} \Pr\left(d^a_{it}|x_{it};\theta\right) \cdot \Pr\left(d^p_{ijt}|d^a_{it};x_{it};\theta\right)\right] \cdot \left[\prod_{t=t_i}^{T} \Pr\left(d^p_{ijt}|x_{it};\theta\right)\right] \tag{2.29}$$

where d_i is the set of choice variables $\left\{d^a_i, \left\{d^p_{ij}\right\}_{j=0}^{2}\right\}$; x_i is the set of observed state variables; t_0 is the time period when the focal firm launched its mobile app; and t_i is the time period when consumer i decides to adopt the app.

Given our assumption that the error terms following the i.i.d. extreme value distribution, the choice probabilities can be expressed as

$$\Pr\left(d_{ijt}^{p}=1|x_{it};\theta;t<t_0\right)=\frac{\exp\left\{w_j^{\text{bf}}\left(x_{it};\theta\right)\right\}}{\sum_k \exp\left\{w_k^{\text{bf}}\left(x_{it};\theta\right)\right\}} \tag{2.30}$$

$$\Pr\left(d_{it}^{a}=1|x_{it};\theta\right)=\frac{\exp\left\{w_a\left(x_{it};\theta\right)\right\}}{\exp\left\{w_a\left(x_{it};\theta\right)\right\}+\exp\left\{w_{na}\left(x_{it};\theta\right)\right\}} \tag{2.31}$$

$$\Pr\left(d_{it}^{a}=0|x_{it};\theta\right)=1-\Pr\left(d_{it}^{a}=1|x_{it};\theta\right) \tag{2.32}$$

$$\Pr\left(d_{ijt}^{p}=1|d_{it}^{a};x_{it};\theta\right)=\frac{\exp\left\{w_j\left(x_{it};\theta;d_{it}^{a}\right)\right\}}{\sum_k \exp\left\{w_k\left(x_{it};\theta;d_{it}^{a}\right)\right\}} \tag{2.33}$$

$$\Pr\left(d_{ijt}^{p}=1|x_{it};\theta;t\geqslant t_i\right)=\frac{\exp\left\{w_j^{\text{af}}\left(x_{it};\theta\right)\right\}}{\sum_k \exp\left\{w_k^{\text{af}}\left(x_{it};\theta\right)\right\}} \tag{2.34}$$

2.4.3 Estimation Procedure

We use the Bayesian estimation algorithm proposed by Imai et al. (2009) and Ching et al. (2012) to estimate our two-stage dynamic model. The parameter space is divided into three parts $\theta=(\theta_h,\theta_c,\theta_i)$, where $\theta_h=(c,\sigma_c,b,\sigma_b)$ being the vector of hyper-parameters, $\theta_c=\left(c_0,\alpha,\lambda,\nu^h,\nu^a,\beta\right)$ being the vector of common parameters, and $\theta_i=(c_i,b_i)$ being the vector of heterogeneous parameters.

The IJC algorithm has three major blocks for each Markov-chain Monte Carlo (MCMC) iteration r:

1. Draw the hyper-parameters θ_h^r and $\sigma_{\theta_h}^r$ from the posterior distributions based on the parameters from iteration $r-1$.

   $$\theta_h^r \sim N\left(\frac{1}{I}\sum_{i=1}^{I}\theta_i^{r-1},\left(\frac{\sigma_{\theta_h}^{r-1}}{\sqrt{I}}\right)^2\right)$$

   $$\sigma_{\theta_h}^r \sim \left[IG\left(\frac{1+I}{2},\frac{1+\sum_{i=1}^{I}\left(\theta_i^{r-1}-\theta_h^r\right)^2}{2}\right)\right]^{0.5}$$

2. For each consumer i, draw the heterogeneous parameters θ_i^r by the Metropolis-

Hastings (M-H) algorithm.

3. Draw the common parameters θ_c^r by the Metropolis-Hastings (M-H) algorithm.

The details of the estimation procedure are explained in Appendix B.

2.5 Results

2.5.1 Parameter Estimates

We present our parameter estimates in Table 2.3 and 2.4. Table 2.3 reports the estimates of the common parameters $\theta_c = \left(c_0, \alpha, \lambda, \nu^h, \nu^a, \beta\right)$, and Table 2.4 reports the estimates of the hyper-parameters in the distribution of the heterogeneous parameters $\theta_h = (c, \sigma_c, b, \sigma_b)$. The total number of MCMC iterations is 500,000, and we discard the initial 200,000 burn-in iterations. Most of the parameter estimates enter the utility function with the expected sign and significance, and the trace plots of the parameters are presented in Appendix B.

First, a consumer would incur a positive cost for app adoption, with a posterior mean of 4.685 and a posterior standard deviation of 0.3301. It suggests that consumers tend to delay their app adoption given that it would incur an adoption cost. Second, a consumer would obtain a negative flow utility from purchasing through the firm-owned channels and the third-party channels relative to the outside option, reflecting that the frequency of hotel reservations on average is low. While both coefficients are statistically significant, the firm-owned channels have a higher posterior mean than the third-party channels. Moreover, after app adoption, consumers would increase their preference towards both the firm-owned channels and the third-party channels. In comparing the magnitudes of these two coefficients, app adoption has a larger positive effect on the firm-owned channels than on the third-party channels.

Additionally, a consumer would obtain a positive consumption utility of 2.686 from staying at the luxury or premium hotels compared to staying at the upscale

Table 2.3: Common Parameters

Variables	Interpretation	Posterior Mean	Posterior S.D.
c_0	Intercept of app adoption cost	**4.685**	0.3310
α_1	Average preference towards the firm-owned channels	$-$**0.8155**	0.005530
α_2	Average preference towards the third-party channels	$-$**2.089**	0.006916
λ_1	Change in average preference towards the firm-owned channels	**0.6835**	0.007665
λ_2	Change in average preference towards the third-party channels	**0.2840**	0.01071
ν_1	Consumption utility from luxury or premium hotels	**2.686**	0.1378
ν_2	Consumption utility from economy hotels	$-$**6.514**	0.1740
ν_3	Consumption utility from app adoption	$-$**0.5570**	0.1014
ν_4	Consumption utility from mobile check-in service	**2.217**	0.1619
β	Discount factor	**0.7731**	0.03317

Boldface type indicates that the 95% posterior credible interval for the parameter estimate does not contain zero.

Table 2.4: Hyper-Parameters in the Distribution of the Heterogeneous Parameters

Variables	**Interpretation**	**Mean**	**S.D.**
$c_{i,1}$	Cost of app adoption associated with competitive hotel apps	**2.377**	**0.014347**
		(0.0002061)	(0.0001451)
$c_{i,2}$	Cost of app adoption associated with travel website apps	**2.374**	**0.014346**
		(0.0002049)	(0.0001462)
$b_{i,1}$	Utility of outside option associated with competitive hotel apps	$-$**0.2727**	**0.014346**
		(0.0002044)	(0.0001457)
$b_{i,2}$	Utility of outside option associated with competitive hotel apps	$-$**0.2772**	**0.014347**
		(0.0002062)	(0.0001458)

Boldface type indicates that the 95% posterior credible interval for the parameter estimate does not contain zero.

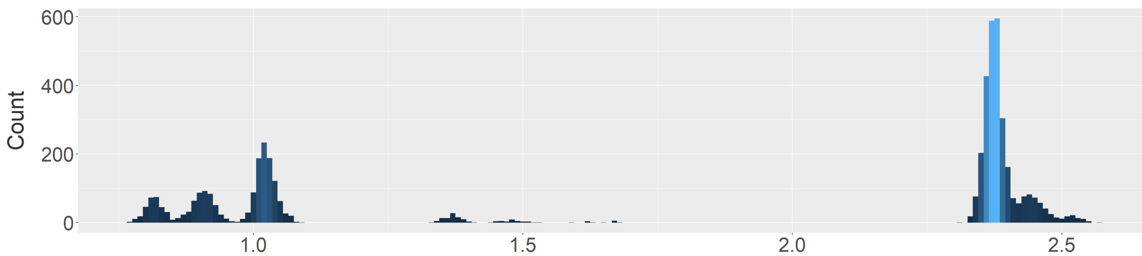
hotels. On the contrary, a consumer would obtain a negative consumption utility of -6.514 from staying at the economy hotels compared to staying at the upscale hotels. Also, a consumer's consumption utility would slightly drop after app adoption, but the availability of mobile check-in service at the hotel would add additional benefits. Finally, the monthly discount factor is 0.7731, which is equivalent to a daily discount factor of 0.9917 and is within the same range as the discount factors in the previous literature.

We have two pairs of heterogeneous parameters in our model, and Table 2.4 reports the hyper-parameters in the distributions of the heterogeneous parameters. A consumer would have higher cost of app adoption when there are more competitive hotel apps and travel websites apps in the market. The population means are 2.377 and 2.374, and The standard deviations are 0.014347 and 0.014346. Figure 2.4 exhibits the distributions of the heterogeneous parameters, which indicate that there is substantial variation in the marginal cost of app adoption. The estimates of these parameters show that more competitive hotel apps and travel website apps in the market lead to higher adoption cost for the focal app. The heterogeneous parameters for the outside option has the base coefficients of -0.2727 and -0.2772, suggesting that the utility of the outside option would drop when there are more competitive hotel apps and travel website apps in the market. As shown in Figure 2.4, there is substantial variation in the marginal utility of the outside option.

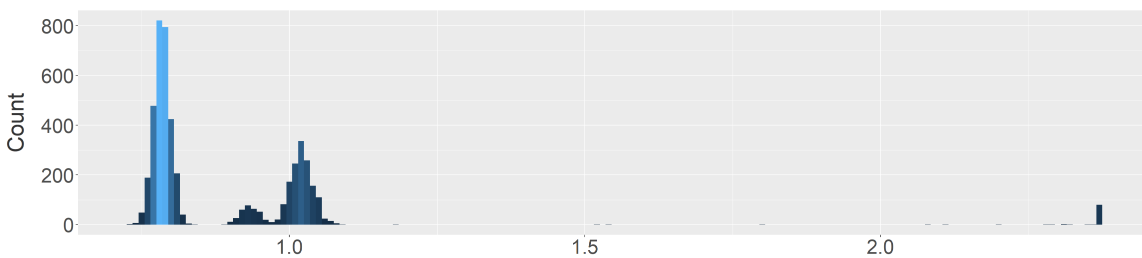
2.5.2 Counterfactual Analysis

We conduct two counterfactual simulations with 3,000 random draws for each analysis, and the results are reported in Table 2.5. First, if the focal firm had not launched its app, the sum of the monthly number of purchasing consumers would drop by 22 percent. In particular, the number of purchasing consumers through the firm-owned channels would decrease by 16 percent while the number of purchasing

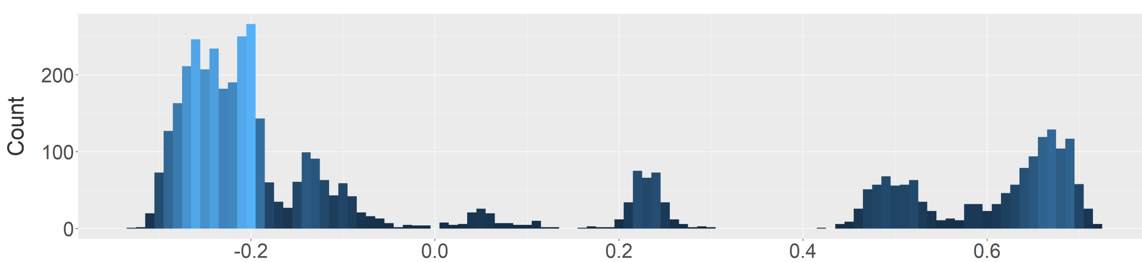
Figure 2.4: The Distributions of the Heterogeneous Parameters



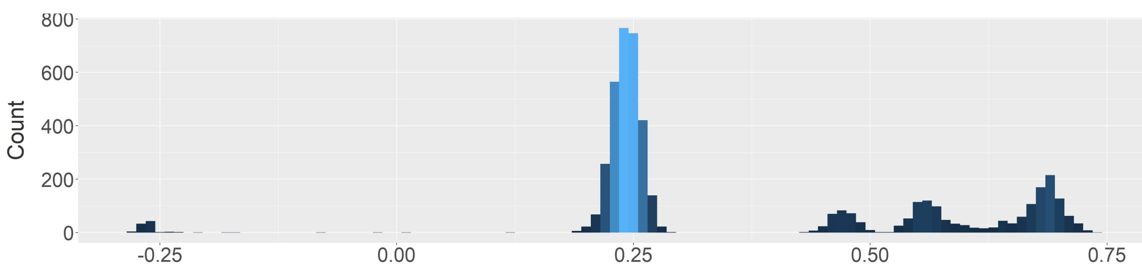
(a) Marginal Cost of App Adoption $c_{i,1}$



(b) Marginal Cost of App Adoption $c_{i,2}$



(c) Marginal Utility of Outside Option $b_{i,1}$



(d) Marginal Utility of Outside Option $b_{i,2}$

consumers through the third-party channels would decrease by 48 percent. It suggests that the introduction of mobile app encourage more consumers to make purchases with the firm.

Table 2.5: Counterfactuals - The Sum of Monthly Number of Purchasing Consumers

	Current	No App Launch		No App Competition	
	Count	**Count**	**Change**	**Count**	**Change**
Total	39,888	31,045	−22.2%	78,767	+97.5%
Firm-Owned Channels	32,190	27,083	−15.9%	70,430	+118.8%
Third-Party Channels	7,698	3,962	−48.5%	8,337	+8.3%

Second, if the competitive hotels and travel websites had not introduced their apps, the sum of the monthly number of purchasing consumers would increase by 98 percent, which would almost double the current number of purchasing consumers. In particular, the number of purchasing consumers through the firm-owned channel would increase by 119 percent and the number of purchasing consumers through the third-party channel would decrease by 8 percent. It suggests that while the firm becomes more attractive as a whole, it is basically driven by the increase through the firm-owned channels. In other words, the competition in the app market hurts the firm-owned channels more than the third-party channels.

2.6 Conclusion

We estimate a two-stage dynamic structural model to investigate an important business strategy – the introduction of mobile apps as a shopping channel. In the first stage, we model consumers' decisions of mobile app adoption based on their app adoption cost and anticipated utility from future app usage. In the second stage, we model consumers' choices of shopping channel, which also depends on their previous app adoption decisions. We find that there are more consumers making reservations with the focal firm after they adopt the app. Furthermore, when there are more competitive apps on the market, the adoption cost of the focal firm's app increases, which leads to the competition effect that consumers being less likely to adopt the focal app if there are more options on the market. Our counterfactual simulations suggest that consumers are less likely to make reservations with the focal firm if the firm had not introduced their apps.

There are several aspects that are not addressed in our analysis, and thus we leave them for future research. First, we do not observe consumers' adoption of other mobile apps in the same product category, and thereby unable to incorporate the competitive factors directly in the model. In the future, if researchers are able observe consumer behavior across different firms, they could model the competition in the mobile app market directly. Second, our research does not include consumers who use both firm-owned shopping channels and third-party shopping channels in each time period. Future research may model this part of consumers as well, which could potentially provide broader insights. Finally, we focus on the discrete choice of shopping channels instead of the amount of consumer spending. It might also be interesting to study the impact of consumers' app adoption on their spending amount.

Chapter 3: Selling the Premium in Freemium

3.1 Introduction

Freemium has become an increasingly popular business model over the past decade. The term "freemium" is a combination of "free" and "premium", which refers to the strategy in which customers can get a basic version of a product or service for free and switch to a premium version with additional features by paying a price. For example, Dropbox offers two gigabytes of free storage capacity in the cloud, generally adequate for text documents. However, if users exceed that storage limit, they have the option to upgrade to one terabyte by paying a monthly subscription fee of $9.99 or an annual subscription fee of $99. According to the latest SEC filings by Dropbox (2018), the freemium model has enabled Dropbox to acquire over 11 million paying users (more than 500 million registered users in total) up to the end of 2017. These paying customers generated approximately $1 billion in revenue for Dropbox in 2017. Other popular examples of the freemium model include voice and video call services by Skype, social networking services by LinkedIn, news content by *The New York Times*, and mobile apps with both free and paid versions. Sometimes, freemium models are also used by firms to get customers into their stores (the "foot-in-the-door" strategy) by giving away free products and cross-selling other offerings from the firm. For example, join.me uses free conference calls to spur demand for its other offerings such as webinars and video conferences.

The appeal of the freemium model for companies is that it can generate high volumes of traffic and provide direct access to potential paying customers without making heavy investment in marketing promotions (Kumar 2014). The freemium strategy serves as a marketing tool that can draw in potential customers through channels such as word of mouth and organic searches. From a customer's perspective, the freemium strategy provides a costless way to sample and experience a new product or service. Companies are able to generate revenues from these customers by selling a premium product to those who value using the product high enough that they are willing to pay for the additional or different functionality. In such a context, one key challenge for companies that utilize the freemium model is how to increase the number of customers purchasing the premium version in the presence of the free version. If a company is unable to convert an adequate fraction of its free customers to paying customers, the freemium model is likely to fail (Kumar 2014).

While extant research has focused on the optimal design of the free product, the pricing of the premium product (e.g., Lee, Kumar, and Gupta 2017), and the conditions for providing free content in a freemium model (e.g., Lambrecht and Misra 2016), such analyses have not studied the premium product line that many implementations of the freemium model involves (e.g., Dropbox offers a premium version for individual consumers and a premium version for businesses). Understanding the impact of such line extensions can be critical for the success of the freemium model. In this research, we focus on the impact of extending the product line to spur demand for the existing premium version and to increase the overall revenues.

Prior research in product lines has shown that extension of a product line is often associated with higher customer utilities as customers are more likely to find a product alternative closer to their respective ideal points (e.g., Moorthy 1984). For example, if the addition of the new alternative is of appropriate (high) quality level, it can lead to higher brand evaluation of the entire product line (Heath, DelVecchio, and McCarthy 2011), which can lead to higher willingness to pay for the premium alternatives in a freemium context leading to higher revenues. However, the freemium context is different from a standard product line context because of the zero price effect (Shampanier, Mazar, and Ariely 2007). This effect predicts that customers perceive the benefits of the free product to be higher than what a cost-benefit model would predict, indicating a stronger preference for the zero priced alternative. Under such conditions, instead of the whole line benefiting from the positive brand evaluations, customers may be disproportionately attracted to the free version. This, in turn, may lead to non-realization of the expected positive impact as predicted by the product line extension literature. As the above example illustrates, the conceptual nuances introduced by the presence of the free version in the freemium model necessitates more than an application of standard product line models to this context and requires a closer examination of the demand outcomes driven by the potential effects of different behavioral mechanisms. This is one of our key objectives in this paper.

The subject of our research is a publisher of scholarly, research-oriented content, the National Academies Press (NAP), which offers free online PDF versions for all of its book titles and sells a paperback version of the same titles for a price. We conducted a randomized field experiment that examined customer downloads and purchases on NAP's website (www.nap.edu) between January 2016 and August 2016. All book titles that were randomly selected in our field experiment have a free PDF version for download and a paperback version available for purchase on the publisher's website before the start of experiment. We manipulated the treatment for book titles by randomly assigning them to one of the following three conditions: (1) a treatment condition under which an additional premium version of ebook format is introduced, (2) a treatment condition under which an additional premium version of hardcover format is introduced, and (3) a control condition under which no additional premium versions are made available.

We apply the difference-in-differences method to the data generated from the field experiment, which covers a seven-month period before the start of the experiment and a seven-month period after the start of the experiment. Overall, we show that the titles accompanied by an additional premium version, either the ebook or the hardcover format, have a higher paperback sales quantity than those in the control condition. The positive impact on paperback sales is stronger for titles that are more popular or lower in price. Additionally, our results reveal that when the ebook price is close to the paperback price, the increase in paperback sales is more pronounced. To further understand the underlying mechanisms driving the results of our randomized field experiment, we analyzed customer choices at the individual level using choice models. Interestingly, our analysis indicates the existence of the compromise effect and the attraction effect in the extended product line setting as the reason for the observed results. Based on our findings we provide specific recommendations on premium versions and prices for managers to improve the effectiveness of their freemium model implementations.

The contributions of this paper are two-fold. First, this study establishes the impact of product line extensions on enhancing the effectiveness of freemium models. Previous literature has discussed how to improve the performance of freemium models by focusing on the design of the free product/service and its price. For example, Lee, Kumar, and Gupta (2017) explore the optimal design of the free product in the context of cloud-based storage service. They estimate a dynamic structural model characterizing consumers' plan choice, usage and referral. Specifically, they analyze the impact of referral incentive on customers' likelihood of upgrading to the premium product and find the referral program contributes to 65% of the value of free consumers. Niculescu and Wu (2014) compare two freemium models in software industry, feature-limited freemium and uniform seeding, to a conventional business model under which software is sold as a bundle without free offers. Their analytical model suggests that the feature-limited freemium strategy outperforms the conventional model when consumers have either relatively low or high prior on premium functionality. Lambrecht and Misra (2016) propose a revenue model for online content providers and suggest it is optimal for firms to offer more free content during periods of high demand but charge for content during periods of low demand. Our paper examines the impact of adding additional versions to the product line and investigates how product line extensions improve the freemium model performance.

Second, our research adds to the substantive findings in the product line literature. Specifically, our study focuses on the conditions under which a product line extension can overcome the potential zero-price effects (Shampanier, Mazar, and Ariely 2007) and result in increasing the demand for the premium versions. In addition, our results show how behavioral mechanisms such as the compromise effect (e.g., Simonson 1989) and the attraction effect (e.g., Huber, Payne, and Puto 1982) can explain the demand patterns seen in a product line extension context, providing evidence for these alternative explanations in our field study. While these effects have been well established in the laboratory setting, this is the first time, to our knowledge, these effects have been empirically identified in a product line setting. In line with the compromise effect, our results suggest that when a higher-quality hardcover version is introduced to the product line, thus making paperback the middle option, the sales of paperback version increase. Also, when an ebook version is introduced to the product line and its price is set close to the paperback version, the sales of paperback version increase because paperback version tends to dominate the high priced ebook version, leading to an attraction effect. Thus, our paper makes substantive contributions to both the freemium model literature and the product line literature. In addition to providing useful managerial implications for the publisher we studied, our research provides better understanding of the behavioral mechanisms leading to the success of freemium models as well as product lines.

3.2 Conceptual Background

The basic question we are seeking to answer in our research is how to increase the demand for the premium version in a freemium model by adding another premium version as a product line extension — a version that could be either of higher quality and higher price than the existing premium version or of lower quality and lower price than the existing premium version. We also seek to identify the underlying mechanism(s) driving the results in such a context. The obvious starting point for such research is the extant work in product line extensions. In what follows, we discuss this literature and the associated conceptual underpinning and argue why a direct application of the results from extant product line research is inadequate in the freemium context.

3.2.1 Product Line Extensions

Focusing on the extant work, addition of products in a product line is often associated with higher customer utilities as customers are more likely to find an option close to their respective ideal points resulting in a larger demand for the firm offering the line. While a monopolist can price these products appropriately and extract all the surplus from customers, in competitive markets the return from a longer product line depends on the pricing power, the costs of producing the extended line and the associated expenses in supporting the line (Moorthy 1984; Kekre and Srinivasan 1990; Bayus and Putsis 1999; and Draganska and Jain 2005). Depending on the distribution of customer valuation for quality and taste preferences, introduction of lower quality/lower priced alternatives in the product line can lead to cannibalization and decreases in revenue and profits (e.g., Desai 2001). Focusing on brand evaluations, Heath, DelVecchio, and McCarthy (2011) find that introduction of higher quality/higher priced alternatives in the line can lead to increases in brand evaluations of larger magnitude than the decreases in brand evaluation when lower quality/lower priced alternatives are introduced. The underlying reason for such effects is that an extended product line provides more options for customers which mitigates the negative evaluation due to the lower quality products, while higher quality products set the overall quality expectations from the brand leading to higher evaluations.

On the other hand, Caldieraro, Kao, and Cunha (2015) show that adding a higher quality/higher priced alternative to the line can lead to decreased overall demand, as the addition of the alternative could lead to consumers' reassessment of their perceptions of the brand in a competitive setting, especially when the attributes of the products match competitor's brand. While these results are mixed, we can, however, expect the positive impact of higher quality/higher priced alternative to exist in a monopolistic setting.

In extending the above findings to the freemium context, we examine two alternative cases: introduction of a higher quality/higher priced alternative, and introduction of a lower quality/lower priced alternative. In the first case, if there are enough customers with ideal points close to the high quality alternative, then we can expect these customers to substitute the existing premium alternative and/or the free alternative with the new premium alternative and thereby increasing revenues (e.g., Tversky 1972; Moorthy 1984). Additionally, one could argue that the introduction of higher quality option can increase the brand evaluation of the freemium brand (e.g., Heath, DelVecchio, and McCarthy 2011), making the entire product line more attractive than before. If this leads to higher willingness to pay for the brand, then the entire product line can benefit through increased demand for the free alternative as well as increased sales of the existing premium and the newly introduced high quality premium alternatives. Analyzing the second case of introducing the lower quality, lower-priced premium alternative as compared to the existing premium alternative, the insights from the product line research would suggest cannibalization of the existing premium alternative by the lower-priced alternative (e.g., Desai 2001). However, the provision of more alternatives could make the product line more attractive and thus some customers could switch from the free alternative to the lower-priced premium alternative (as it could be closer to their ideal point). The net impact of these effects depends on the magnitude of these effects. Revenues may decline if the cannibalization effect on the existing premium alternative outweighs the revenue increase or the other positive impacts generated from the increase in demand for the newly introduced lower-quality, lower-priced premium alternative.

3.2.2 Nuances in the Freemium Context

The freemium context is different from a standard product line context because of the role of zero price of the free alternative influencing customers' choices through the zero price effect (Shampanier, Mazar, and Ariely 2007). The traditional cost-benefit perspective predicts that customers choose the alternative with the highest cost-benefit difference. But the zero price effect predicts that customers perceive the benefits of the free product to be higher than what a cost-benefit model would predict, indicating a stronger preference for the zero priced alternative. Depending on the magnitude of the zero price effect, when a higher quality/higher priced alternative is introduced, the positive brand evaluations suggested by Heath, DelVecchio, and McCarthy (2011) can occur, but instead of the whole line benefiting from it, customers may be disproportionately attracted to the free version, leading to non-realization of the expected positive impact of adding the higher quality/higher-prices alternative as predicted by the product line literature. In the case of adding a lower quality, lower priced alternative to the product line, the existence of zero-price effect could significantly dampen the switch away from the free alternative leading to a net negative effect, and so the results from the product line research may not directly translate to the freemium context.

Additionally, the literature on the anchoring effect (e.g., Simonson and Drolet 2004) would suggest that customers are likely to use the price of the premium alternative as an anchor to evaluate the value of the free alternative, mentally coding the difference between price of the premium and price of the free alternative (zero) as a "gain" in obtaining the free alternative. Upon the introduction of a higher quality, higher priced premium alternative, the anchor could shift from the existing premium alternative to the higher priced premium alternative, leading to a larger gain in obtaining the free alternative and making the free alternative even more attractive. Under this scenario, the introduction of higher quality, higher priced premium alternative could lead to decreased overall revenue as customers could shift from buying premium alternatives to obtaining the free version. This highlights why the freemium context requires a more nuanced analysis than what a direct application of product line research would suggest.

Moreover, there could exist other behavioral effects in the freemium context, hitherto untested in the product line context but could play an important role in a monopolistic setting that we consider. Behavioral research in the context of choice set variations has shown that introducing a new alternative into a choice set can have an impact on the choice probabilities of the original alternatives in two important ways. For example, as per the compromise effect (Simonson 1989), when a higher quality, higher-priced premium alternative is introduced to the choice set, the existing premium alternative could become a compromise option between the higher-quality, higher-priced premium alternative and the free alternative, which could result in customers switching from the free alternative and choosing the compromise option. This could lead to higher sales of the existing premium option and thereby increase revenues - essentially moving customers from the free alternative to the premium alternative. Analyzing the case when a lower-quality, lower-priced alternative (with quality higher than the free alternative) is introduced to the existing product line, the newly introduced alternative could become the compromise option between the existing premium and free alternatives, and its sales could see an increase. If the extent of cannibalization of the existing premium by this alternative is lower than the increase in revenue due to customers switching from the free alternative to the newly introduced alternative, the new impact on revenues could still be positive. On the other hand, if the newly introduced alternative cannibalizes the existing premium alternative to a greater degree than it gains in revenues, total revenue could decrease.

In addition to the compromise effect, this setup could also make salient the attraction effect (Huber, Payne, and Puto 1982; Huber and Puto 1983; Simonson 1989), where the existing premium alternative dominates the higher-quality, higher-priced premium alternative in one case, and dominates the lower-quality, lower-priced alternative in the other case. It is conceivable that if the higher-quality alternative is priced close to the existing premium product, then the newly introduced alternative could dominate the existing premium alternative. The existence of attraction effect would make the dominating alternative (the newly introduced alternative) more attractive and this could lead customers to switch from the free alternative and the existing premium option to the newly introduced alternative, leading to increased revenues. In the second case, if the newly introduced lower-quality alternative is very low on the quality dimension as compared to the existing premium but priced close to the premium option's price, it could be dominated by the existing premium option and could lead to customers switching from the free option to the existing premium option, again leading to overall revenue increase. The zero price effect could coexist with the these behavioral effects of compromise and attraction effect and could dampen the demand for the existing premium alternative under both conditions. It could also have an impact on the utilities determined for the various alternatives in the product line design, even in the absence of other behavioral effects.

In sum, the outcome of including another premium alternative to the offerings in the freemium model on the existing premium alternative's demand and overall revenue depends on the relative strengths of these opposing effects of the underlying mechanisms (see Table 3.1). There is a notable difference between the market expansion effects predicted by the product line literature (e.g., the quality-association effect and the positive variety effect) and the choice set effects by the behavioral literature (the compromise effect and the attraction effect). The market expansion effects predict the brand and the product line as a whole are evaluated higher by customers. However, although more customers could be attracted to the brand, they are much more likely to be attracted to the free alternative due to the zero-price effect. On the contrary, the compromise effect and the attraction effect work at targeting a specific alternative in the choice set. Therefore, even if the compromise effect and the attraction effect could be dampened by the zero-price effect, firms can still focus on making the premium alternative very attractive to realize a positive impact on the demand for the premium alternatives.

Our objective in this paper, therefore, is to determine the outcome of adding the alternatives — either a high-quality, higher-priced alternative or a lower-quality, lower-priced alternative — using a randomized field experiment. We analyze the data on sales at the aggregate title level obtained from the field experiment to determine the outcome and identify the mechanisms that could be at play (see Table 3.1), and analyze data at the individual customer level to confirm specific mechanisms at work that lead to the results.

3.3 Field Experiment

3.3.1 Institutional Setting

National Academies Press (NAP) is a non-profit publisher of scholarly content on a wide range of topics in the areas of pure sciences, social sciences and education. Motivated by its dual mission of dissemination and self-sustainability, NAP is committed to enhancing its reach to the audiences all over the world while remaining self-sufficient. Most of NAP's sales revenue comes from its website, which started selling books in print format in 1996. Around 2003, NAP started selling the book titles in PDF version, which could be directly downloaded from the website. The prices of the PDF versions were set around 75% of the print book prices (see Kannan, Pope,

Table 3.1: **The Expected Impacts of Adding a Premium Alternative on Sales**

	Expected Impact on Sales		
Underlying Mechanisms	**Free Alternative**	**Existing Premium**	**Overall Revenue**
Strategy 1: Adding a higher-quality, higher-priced premium alternative			
Market expansion			
Variety effect (e.g., Berger et al. 2007)	Positive	Positive	Positive
Quality-association effect (e.g., Heath et al. 2011)	Positive	Positive	Positive
Cannibalization effect (e.g., Tversky 1972; Moorthy 1984)	Negative	Negative	Positive
Compromise effect (e.g., Simonson 1989)	Negative	Positive	Positive
Attraction effect (e.g., Huber et al. 1982)	Negative	Negative	Positive
Strategy 2: Adding a lower-quality, lower-priced premium alternative			
Market expansion			
Variety effect (e.g., Berger et al. 2007)	Positive	Positive	Positive
Quality-association effect (e.g., Heath et al. 2011)	Negative	Negative	Negative
Cannibalization effect (e.g., Tversky 1972; Desai 2001)	Negative	Negative	Ambiguous
Compromise effect (e.g., Simonson 1989)	Negative	Negative	Ambiguous
Attraction effect (e.g., Huber et al. 1982)	Negative	Positive	Positive

Note: The zero-price effect will attenuate the negative effect on the free PDF downloads and strengthen the positive effect on the free PDF downloads (Shampanier et al. 2007).

and Jain 2009). All along, NAP also provided all the content of its book titles on the website in a lower-quality PDF format (called “open book”), which could be browsed page-by-page (but cannot be downloaded as a full document). This eliminated any uncertainty about the content of any book a customer purchased. Customers could simply browse the content and then purchase the book in print format and/or PDF format. More recently, under the directive of the board of the National Academies, NAP changed its business model to a freemium model, wherein the PDF format of any book title could be downloaded for free, while a paperback format of the title could be purchased online for a price. The free PDF content attracted significant traffic to the website through organic searches. Most customers tended to sample the open book for content and then download the free content, while a fraction switch to purchasing the premium paperback version. While the website traffic increased due to the freemium model, revenue decreased significantly as compared to the pre-freemium model period (when the PDF version was sold at a price).

Similar to any content seller, NAP’s demand for the titles it sells is characterized by a long tail distribution. A few titles tend to be bestsellers in any given year while majority of titles have low demand, both in terms of free PDF downloads as well as paperback sales. Given the pressures on revenues in the context of the freemium model, NAP contemplated adding additional versions of the book titles to their current offerings of free PDF version and the premium paperback version to increase revenue. Although the free PDF version has an acceptable image resolution, the paperback version has a much higher quality resolution and production quality. To this, NAP wanted to add a hardcover version with superior production quality and a more durable cover and binding. However, the higher cost of producing a hardcover meant that the price for this format would be much higher than that of a paperback. NAP also considered adding an ebook version, with resolution quality somewhat similar to the PDF version but with additional features such as font size customization, notes features, and compatibility with Kindle and iPhone/iPad devices. Since customers' price expectations for electronic formats are generally lower than those of print formats, the ebook's prices are to be set lower than those of the paperback versions. It is in this context that NAP wanted to research the impact of extending the product line of premium products on the revenues of NAP — either through the higher quality/higher priced hardcover option or the lower quality/lower priced ebook option.

The NAP context is similar to the more well-known freemium model implementations (NYTimes, Spotify, Dropbox) on all the salient dimensions. The free alternative or the free PDF content generates a market for the paperback version, which generates the revenues to keep the non-profit self-sustaining, much in the same way free content or free membership or free cloud storage generates revenues for the premium version for NYTimes, Spotify or Dropbox, respectively. The free alternative allows the firm to generate a potentially large number of customers through word of mouth in all these examples including NAP. Spotify and NYTimes earn advertising dollars through the usage of the free service and free content, while NAP earns utility through the dissemination of the content of the books in free PDF versions (which the authors, mainly from academia, value). The freemium models also range from being service subscription (Dropbox) to content subscription (Spotify, NYTimes). Although the content purchase is in product form in the NAP context, the utility for users is derived from the repeated usage of the content of book (which are mainly academic, reference type material) much in the same way Spotify users or NYTimes users derive utility from repeated use of the service. While the time frame of using the free version and switching to the premium version may be longer in the other examples, in the NAP context a customer samples the free content to determine whether to continue using the free version or switch to the premium version in a short time. This is because other freemium models involve experience products, while the value of content at

NAP can be determined quickly.

There are two advantages of conducting this research in the context of NAP. The first advantage is that in the category of books, the product line varies only in its formats while the content remains the same across different formats. Therefore, the free and premium versions will only differ in their formats and prices, which are the attributes that we want to manipulate in the field experiment. Second, NAP can be considered a monopolist in the market for its specialized and authoritative content because it has the exclusive right to publish studies of the National Academies, which frees us from the potential omitted variable problem caused by unobserved competitive factors from other publishers. These two characteristics help ensure a clean experimental setting which improves the validity of our results.

3.3.2 Formats in the Product Line

We now provide a detailed description of the formats to clarify the attributes of the products making up the freemium product line. The free alternative is the PDF version, which is a single PDF file with an image resolution lower than that of the existing premium version (the paperback), but with the advantage of being electronically transferable. The paperback version has all the advantages of a print copy: it can be leafed through quickly and can be stacked on a bookshelf for easy retrieval. The higher-quality, higher-priced alternative is the hardcover version, which has a thicker protective cover, a better quality binding, and acid-free pages, all of which make the hardcover version more durable than the paperback version. The lower-quality, lower-priced alternative is the ebook version, which is optimized for e-reader devices and apps that offer a much better digital reading experience than a PDF. The ebook version allows readers to resize fonts, bookmark pages, make notes, highlight specific passages, and save selected text. The ebook version also automatically scales for the content while the PDF version has a fixed width and height. The paperback and hardcover versions can be viewed as vertically differentiated, as are the PDF version and the ebook version. In comparing the PDF with the paperback on some features such as image quality, the formats are vertically differentiated, the paperback being clearly better, but on other dimensions they could be horizontally differentiated. Nevertheless, customers could choose between the two formats given the strength of their preferences for each and their relative prices. Regardless, the important question in our context is how the addition of the new formats affects the demand for the original existing premium product and overall revenue.

3.3.3 Experimental Design

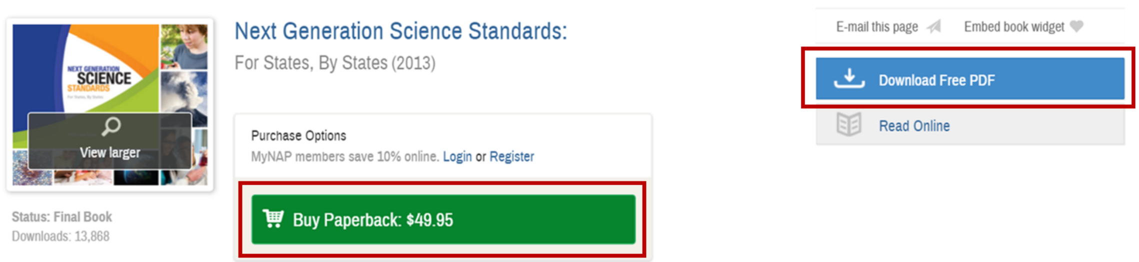
The field experiment includes a total of 2,051 randomly selected book titles that NAP was already selling in paperback version on its website before the start of the experiment, and the content of these titles was also available for free download in PDF version. Given that the demand distribution (PDF downloads as well as paperback sales) is long-tail and NAP wanted the results to be generalizable to all their titles, we sampled titles from the entire distribution including bestsellers. The heterogeneity in the titles is accounted for by our design as described below. Our field experiment started on January 10, 2016, when all new formats were introduced on the same day, and lasted for 32 weeks until August 20, 2016. Our dataset covers the 32 weeks before the experiment and the 32 weeks after it started, totaling 64 weeks.

Randomized block design. Since the prices and the popularity (proxy for potential demand) of the titles vary significantly, we use a restricted randomized design that uses the price of the paperback version as one blocking variable with three levels (low, medium, and high) and title popularity as the other blocking variable with three levels (low, medium and high). This creates nine cells within which the titles are similar on the two blocking variables. According to the distribution of price of paperback version, the price level is defined as “low” if it is among the 25% lowest prices (in our setting this lower quartile has a price $\leqslant$ \$29), as "high" if it is among the 25% highest prices (in our setting this higher quartile has a price $>$ \$47.25), and as "medium" if otherwise (i.e., \$29 $<$ price $\leqslant$ \$47.25). Popularity is measured by the count of page views of a book title's catalog page (webpage where the details of the titles are provided), using the number of views as a metric of popularity among customers. The threshold for each level of popularity is determined in a similar way based on the distribution of the metric: popularity is defined as low if counts $\leqslant 136$ (the bottom 25%); as high if counts > 953 (the top 25%); and as medium if $136 <$ counts $\leqslant 953$ (the middle 50%). To summarize, cells 1-9 contain 269, 208, 38, 218, 557, 250, 40, 250 and 221 titles, respectively.

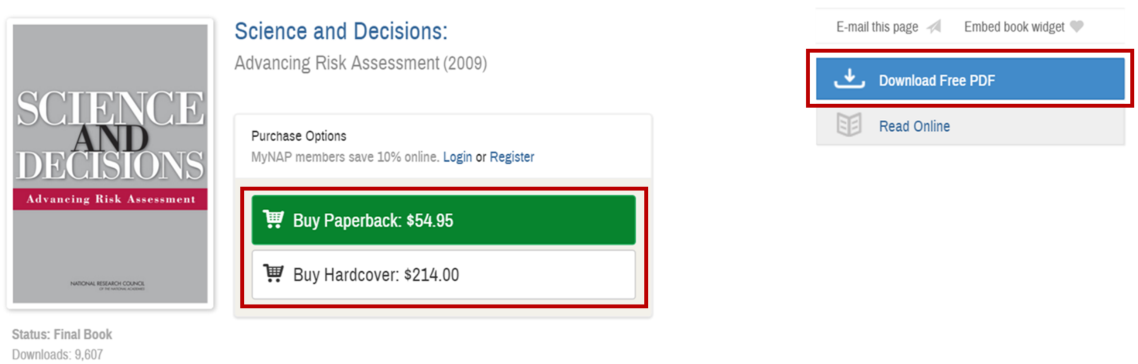
Control and treatment conditions. In each cell, titles are randomly assigned to one of the following three conditions: the treatment condition, TH, where a hardcover version is added to the existing product line; the treatment condition, TE, where an ebook is added to the existing product line; and the control condition, C, where no changes are made to the existing product line. As in Figure 3.1(a), the existing product line at the start of the experiment comprised two versions: the free version (the PDF version for download) and the premium version (the paperback version at a specific price). At the start of the experiment, no changes are made to control condition (C) titles; a hardcover version is added for the TH titles (see Figure 3.1(b)); and an ebook version is added to the existing product line for the TE titles (see Figure 3.1(c)). The presentation order of the new premium version (Figure 3.1) is also randomized within a treatment condition. Overall, 716 titles are assigned to the control condition, 668 titles are assigned to the hardcover condition, and 667 titles are assigned to the ebook condition (see Table C1 in Appendix C).

Price of new premium versions. The prices of titles in paperback version are kept at their existing levels, with the corresponding PDF versions being available for free. One objective of our research in varying the relative prices of the new versions is to

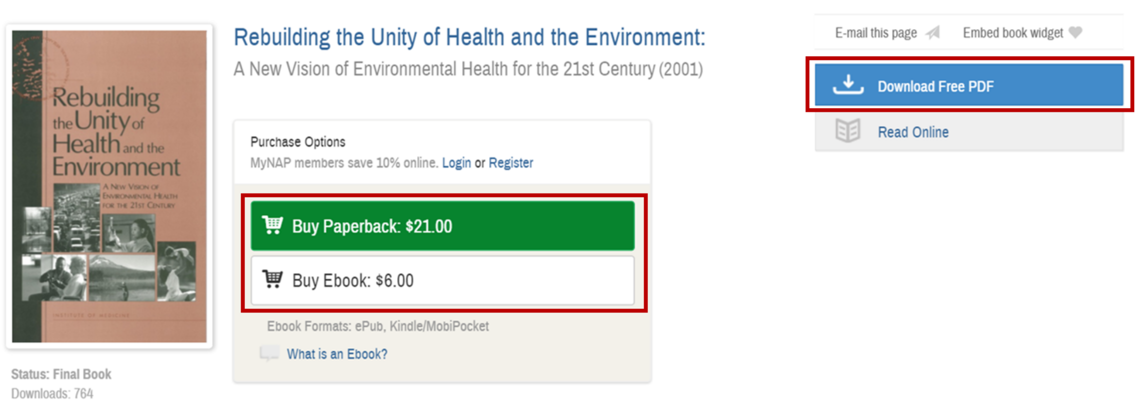
Figure 3.1: The Examples of Titles under Control and Treatment Conditions



(a) The Product Line of All Titles before Experiment Started and Titles under Control Condition (C)



(b) The Product Line of Titles under Hardcover Condition (TH) during the Experimental Period



(c) The Product Line of Titles under Ebook Condition (TE) during the Experimental Period examine when paperback sales increase. For the ebook version, in line with the pricing practices of publishing industry (Trachtenberg 2010; De los Santos and Wildenbeest 2017), the titles are priced lower than the price of paperback version. We generate an equally spaced sequence that consisted of 20 levels ranging from 25% and 80% of the price of the titles in the paperback version. Similarly, hardcover prices are set between the range of 160% and 790% of the paperback version prices, which covers the usual range of hardcover prices encountered in NAP as well as in the publishing industry. The wide range of the specific pricing levels are also chosen to cover all the possibilities that NAP was considering with regard to the pricing strategy of ebook and hardcover versions. The ebook or hardcover prices are displayed in absolute dollar terms on the website. Finally, the prices for paperback, ebook, and hardcover versions of all titles are held constant throughout the duration of the experiment.

3.3.4 Descriptive Statistics

Our dataset contains the characteristics of book titles including the price of different versions, the copyright year, the number of pages, and the dummy variables identifying the different experimental condition. We also observe the sale quantities and the revenues of paperback, ebook and hardcover versions, the number of free PDF downloads and the number of email campaigns.[1] Table 3.2 shows the summary statistics for the data, and Table 3.3 compares the mean value of variables across the control and treatment conditions.

As seen in Table 3.2, there is a positive number of PDF downloads for every title (ranging from 4 to 26,066), while the paperback sales are positive for a much smaller number of titles. This is the characteristic of freemium models, where the demand for free products are much higher than that for the premium version. For example,

[1] NAP uses email campaigns directed at their customers to advertise their titles. Our data show that the number of email campaigns increased somewhat after the start of the experiment but this was mainly in the control condition and not others, thus posing less of a threat to validity of our results.

Table 3.2: **Summary Statistics**

Variables	**Mean**	**SD**	**Min**	**25%**	**75%**	**Max**
Paperback Price ($)	39.28	17.96	5.00	29.00	47.25	475.00
Copyright Year	2008	4	2000	2004	2011	2015
Number of Pages	181	121	20	96	236	1,357
Before Experiment 32 weeks (average per title)						
Free PDF Quantity	188.47	844.25	4	33	149	26,066
Paperback Quantity	1.77	26.96	0	0	0	1,035
Paperback Revenue ($)	47.57	464.91	0	0	0	13,602
Total Revenue ($)	47.57	464.91	0	0	0	13,602
Email Campaigns	.05	.33	0	0	0	8
Experiment Period 32 weeks (average per title)						
Free PDF Quantity	173.32	699.63	5	34	134.5	23,961
Paperback Quantity	1.57	25.71	0	0	0	837
Paperback Revenue ($)	42.38	558.09	0	0	0	21,600.77
Total Revenue ($)	43.69	568.89	0	0	0	22,061.87
Hardcover Quantity	.01	.09	0	0	0	1
Hardcover Revenue ($)	1.19	15.73	0	0	0	357.00
Ebook Quantity	.16	1.06	0	0	0	17
Ebook Revenue ($)	2.82	20.94	0	0	0	461.10
Email Campaigns	.06	.33	0	0	0	6

Table 3.3: **Comparing Mean Values across Conditions**

Variables	All Titles	Hardcover Condition	Ebook Condition	Control Condition
Paperback Price	39.28	40.51	38.86	38.53
Copyright Year	2008	2007	2008	2008
Number of Pages	181.09	187.62	180.19	175.83
Before Experiment 32 weeks (average per title)				
Free PDF Quantity	188.47	166.83	248.08	153.13
Paperback Quantity	1.77	1.58	2.93	0.86
Paperback Revenue ($)	47.57	49.90	67.57	26.77
Total Revenue ($)	47.57	49.90	67.57	26.77
Email Campaigns	0.05	0.04	0.07	0.03
Experiment Period 32 weeks (average per title)				
Free PDF Quantity	173.32	162.36	205.08	153.96
Paperback Quantity	1.57	1.43	2.79	0.57
Paperback Revenue ($)	42.38	41.34	64.80	22.48
Total Revenue ($)	43.69	42.53	67.62	22.48
Email Campaigns	0.06	0.05	0.06	0.06

for Dropbox the percentage of premium users in its customer base is around 4%.

The average number of paperback sales (quantity) per title before the start of experiment is 1.77, and it drops to 1.57 during the experimental period. Our difference-in-differences model will account for the potential decreasing trend over time (due to seasonal effects NAP typically has higher sales in the last four months of a calendar year which is part of our pre-experiment period). Along with the drop in paperback sales quantity, both paperback revenue and total revenues decrease during the duration of the experiment. Also consistent with the seasonal effects, the downloads of free PDF decline from 188.5 to 173.3 over time.

In addition, the hardcover version generates average revenue of $1.19 per title under the hardcover treatment condition. This primarily reflects the very low sales quantity of the hardcover version. The ebook version generates average revenue of $2.82 per title under the ebook treatment condition. We note that there is a great amount of heterogeneity in sales quantity and revenues across book titles, as their standard deviations are much larger than the means, which justifies our decision to use block design by dividing book titles into nine cells to reduce the variance.

Comparing the sales performance across titles under the control and treatment conditions, we highlight several notable patterns in Table 3.3. First, the sales quantity and revenues of paperback version decrease across all three conditions from the pre-experimental period to the post-experimental period. Therefore, it is essential to control for the time trend in our analysis. Second, the total revenue of titles under the ebook condition increases slightly after the start of the experiment. Additionally, free PDF downloads increase slightly for titles under the control condition but decrease for titles under the treatment conditions. This provides model-free evidence that adding a new version (hardcover version or ebook version) may lead to lower downloads of free PDF version.

Finally, there may seem to be differences across treatment and control conditions in terms of quantities downloaded and sold before the start of the experiment. For example, titles under the ebook condition have the highest downloads of PDF version and the highest sales quantity and revenues of paperback version among the three conditions before the start of experiment. However, in conducting a one-way analysis of variances to compare the mean values of the three conditions before the start of experiment, we find no statistical difference in means at a significance level of .05 (see Table C2). This result confirms that our random assignment ensures the titles are similar across different experimental conditions before the experiment started.

3.4 Difference-in-Differences Model

3.4.1 Model Specification

In order to analyze the effect of adding a hardcover version to the freemium product line, we apply the difference-in-differences (DID) method to the data of titles across the two periods of pre-experimental and experimental durations under the control and hardcover conditions. Similarly, we estimate the effect of adding an ebook version by using the data for the titles under the control and ebook conditions. The DID model allows us to account for unobserved factors which are common to all book titles, such as economic shocks, seasonal effects, and firm level marketing campaigns that are common to all book titles. It also eliminates the impact of the fixed effects of the titles as we take differences across pre-experiment and experimental periods. If these two factors are left unaccounted for, they can seriously invalidate the results from any market response model.

Our dependent variable is the sales quantity of the paperback version of a book title, since the paperback version is the only premium (non-free) format present in all conditions and the dominant revenue source. We combine the DID model with a negative binomial distribution as the dependent variable, sales quantity of paperback, follows a long right tail distribution with non-negative values. Using a negative binomial distribution also diminishes the impact of variances in sales quantity across titles, while accounting for the title demand variances being different from the means. For book title i in period t,

$$E\left[Y_{it}=y|\boldsymbol{X}_{it},\epsilon_{it}\right]=\exp\left\{\boldsymbol{X}_{it}\boldsymbol{\beta}+\epsilon_{it}\right\},\quad t=0,1 \tag{3.1}$$

$$\boldsymbol{X}_{it}\boldsymbol{\beta}=\alpha+\theta\cdot\mathrm{Treat}_i+\gamma\cdot\mathrm{Post}_t+\delta\cdot\left(\mathrm{Treat}\times\mathrm{Post}\right)_{it}+\rho\cdot Z_{it}+\boldsymbol{W}_i\boldsymbol{\lambda} \tag{3.2}$$

where Y_{it} is the dependent variable, sales quantity of paperback version. There are two periods: $t=0$ denotes the period before experiment started, and $t=1$ denotes the period of experimental duration. $\mathrm{Treat}_i=1$ if title i is assigned to the treatment condition; $\mathrm{Post}_t=1$ if period t happens after the start of experiment; Z_{it} measures the number of email campaigns for title i in period t; and $\boldsymbol{W}_i$ represents the characteristics of book titles including the copyright year and the number of pages. Since the randomized block design captures the variations on the dimensions of price and popularity, we use the copyright year and the number of pages to control for additional variations among book titles that have not been captured by the randomized block design. The copyright year describes the recency of publication, and the number of pages captures the production cost of the paperback version.[2]

Additionally, ϵ_{it} is an idiosyncratic error term that captures all determinants of Y_{it} that our model omits, with $\exp\left\{\epsilon_{it}\right\}$ following a one-parameter gamma distribution $G\left(\kappa,\kappa\right)$. Furthermore, we allow the error terms of the same title to be correlated before and after the experiment started, i.e., $\mathrm{E}\left(\epsilon_{i0}\epsilon_{i1}\right)\neq 0$, $\forall i$. This corrects for a potential bias in the estimation of standard errors due to serial correlation.

For parameters, θ is the treatment condition specific effect that captures the average permanent differences between the control and treatment conditions. If our random assignment has been successful, the estimate of θ should not be significantly

[2] We also run the DID model with the price of paperback versions as an additional explanatory variable. The results remain qualitatively the same.

different from zero; γ accounts for the time trend during the sample period that is common to all titles; and most importantly, the difference-in-differences estimator is δ, which captures the true effect of introducing the ebook or hardcover version on the sales quantity of the paperback version:

$$\hat{\delta} = \left(\log\left\{\mathrm{E}\left[\bar{Y}_1^{\text{treat}}\right]\right\} - \log\left\{\mathrm{E}\left[\bar{Y}_0^{\text{treat}}\right]\right\}\right) - \left(\log\left\{\mathrm{E}\left[\bar{Y}_1^{\text{control}}\right]\right\} - \log\left\{\mathrm{E}\left[\bar{Y}_0^{\text{control}}\right]\right\}\right) \tag{3.3}$$

While the magnitude of $\hat{\delta}$ does not equal to the marginal effect of the treatment, the sign of δ and its marginal effect are identical. That is, $\hat{\delta}$ being positive indicates that the introduction of a new version has a positive impact on paperback sale quantity.

3.4.2 Main Results

Table 3.4 presents our DID model results estimating the effects of (a) introducing a hardcover version to the product line on the sales quantity of paperback versions, and (b) introducing an ebook version to the product line on the sales quantity of paperback version. In Table 3.4, Columns (1) and (2) show the estimates of coefficients and marginal effects at means (MEM) for titles under the control condition and the hardcover condition; and Columns (3) and (4) show the estimates of coefficients and marginal effects at means (MEM) for titles under the control condition and the ebook condition.

Hardcover condition vs. control condition. First, the estimate for dummy variable Treat_i is insignificant, therefore indicating that titles assigned to the control and hardcover conditions are comparable. This confirms that our random assignment has been successful. Second, the sales quantity of the paperback version shows a decreasing trend over time as the estimate for dummy variable Post_t is negative and statistically significant. This indicates a seasonal effect or decreasing trend for NAP titles in general. More interestingly, the coefficient estimate for Treat× Post

Table 3.4: The Effects of Introducing Hardcover or Ebook Version on Paperback Sales Quantity

	Hardcover vs. Control		Ebook vs. Control	
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Coeff.**	**MEM.**	**Coeff.**	**MEM.**
Treat	$-.511$	$-.338$	$-.483$	$-.383$
	(.380)	(.297)	(.355)	(.313)
Post	$-.734{*}{*}$	$-.495{*}$	$-.550{*}$	$-.441$
	(.306)	(.292)	(.282)	(.274)
Treat$\times$Post	$1.207{*}{*}$	1.156	$1.795{*}{*}{*}$	2.582
	(.551)	(.931)	(.688)	(2.262)
Email Campaigns	$1.649{*}{*}{*}$	$1.089{*}{*}$	$2.878{*}{*}{*}$	$2.282{*}{*}$
	(.502)	(.503)	(.657)	(1.038)
Copyright Year	$.135{*}{*}{*}$	$.0893{*}{*}{*}$	$.111{*}{*}$	$.0877{*}{*}$
	(.0468)	(.0281)	(.0432)	(.0375)
Number of Pages	$.344{*}{*}$	$.227{*}{*}{*}$	.0719	.0570
	(.138)	(.0768)	(.0930)	(.0617)
Constant	$-1.810{*}{*}$	–	$-1.294{*}$	–
	(.872)	–	(.702)	–
Observations	2,768	2,768	2,766	2,766
Log-Likelihood	-2243	–	-2321	–
AIC	4502	–	4658	–
BIC	4549	–	4706	–
Pseudo R^2	.0406	–	.0560	–

MEM: marginal effect at means

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

is positive and significant after controlling for the above effects. This shows that introducing a new premium version, the hardcover version in this case, increases the demand for the paperback version. More precisely, introducing the hardcover version can increase sales of the paperback version by 1.2 copies per title, which can be translated into an increase in total sales of 772 paperback copies. In addition, the introduction of hardcover version is able to offset the negative time effect, as the marginal effect's magnitude of Treat$\times$ Post is greater than that of Post_t. Finally, as expected, marketing titles using email campaigns increases paperback sales; book titles published more recently and with more pages have higher paperback sales.

Ebook condition vs. control condition. When the ebook version is added to the freemium product line, we find similar results. First, there is no significant difference between titles under the control condition and those under the ebook condition. Second, the sales quantity of paperback version decreases over time. Furthermore, introducing the ebook version to the product line approximately triples the average sales quantity of the paperback version. In particular, the sales quantity of paperback version increases by 2.6 copies per title due to the introduction of the ebook version, which can be translated into an increase in total sales of 1,722 paperback copies. In addition, email campaigns for book titles significantly increase paperback sales. And book titles published more recently have higher paperback sales.

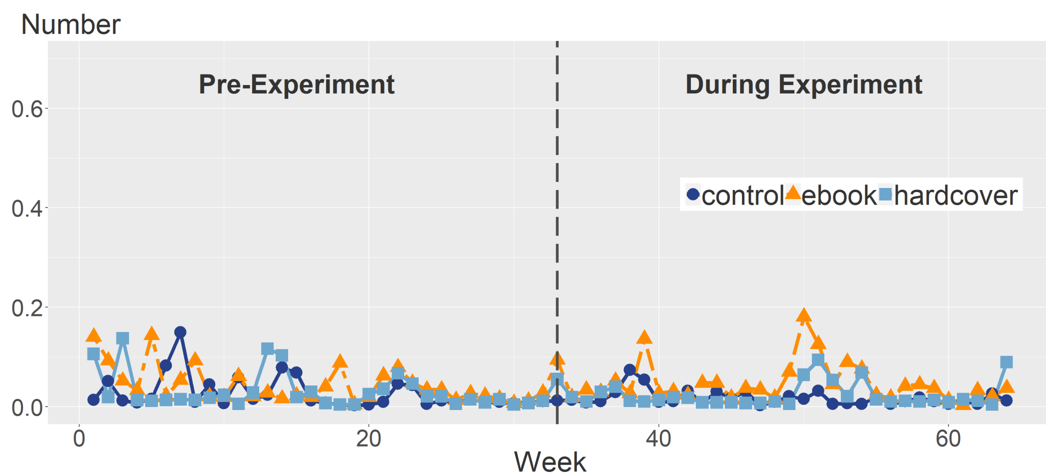
3.4.3 Robustness Checks

As indicated in an earlier section, the demand distribution for NAP's title follows a long tail. Given that a few bestsellers could make up a significant portion of revenue for the firm in a given year, our sampling frame included all those bestsellers to ensure that the results of the analysis was representative of all titles that NAP carries. The use of the DID model already accounts for the fixed effects of each title. In addition, our use of the NBD specification in the DID model ensures that these demand variations do not skew the DID results. Yet, there might still be a concern that the estimated positive effect is driven by a few titles that have the largest quantity of paperback sales (the blockbusters).

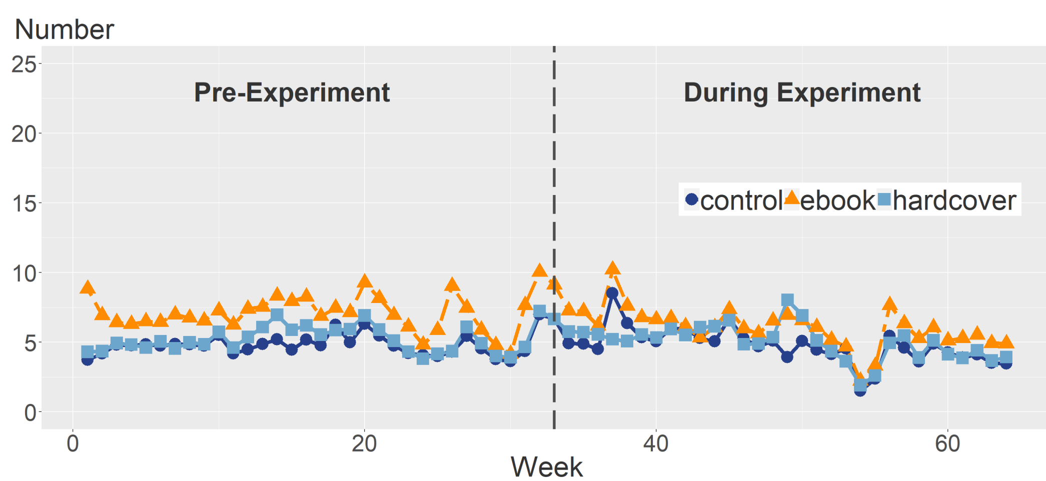
We address this concern by applying our DID analysis to: 1) the sample without blockbuster titles; 2) the 500 sub-samples constructed by randomly drawing 90 percent of all titles; and 3) the 500 bootstrap samples. We find consistent results from all robustness checks. The blockbuster titles are identified based on the distribution of paperback sales quantity (see Table C3 in Appendix C). After removing two blockbuster titles — one title under the hardcover condition and one title under the ebook condition — the mean values across the three conditions become much closer and the standard deviations of titles under the hardcover condition and the ebook condition become much smaller. Furthermore, we evaluate the pre-experiment time trends for the control and treatment customers through a graphical inspection: Figure 3.2 suggests that the time trends of the paperback quantity of titles under the treatment and control conditions were similar before the start of experiment, and thus supports the validity of identification for treatment effect using the DID approach.

After removing the blockbuster titles, we find that adding a hardcover version has a significant positive impact on paperback sales quantity. In particular, introducing the hardcover version can increase sales of the paperback version by .5 copies per title, which can be translated into an increase in total sales of 335 paperback copies. Similarly, we find that introducing the ebook version to the product line can increase sales of the paperback version by .4 copies per title, which can be translated into an increase in total sales of 287 paperback copies (see Table C4, C5 and C6 in Appendix C).

Figure 3.2: The Graphical Test of Common-Trend Assumption (Blockbuster Titles Excluded)



(a) Average Paperback Quantity



(b) Average PDF Quantity

3.4.4 Sub-Sample Analysis

To obtain a better understanding of the influence of introducing new versions to the freemium model, we also analyze the effect of extending the product line on paperback sales quantity using sub-samples defined by paperback price, popularity and copyright year. We find that the positive impact on paperback sales is stronger for titles that are more popular or lower in price. In addition, adding a hardcover version increases the paperback sales of titles published more recently while adding an ebook version increases the paperback sales of titles published earlier (see Table C7, C8 and C9 in Appendix C).

3.4.5 Price of New Premium Versions

We also investigate how price levels of the new premium versions affect the increase in paperback sales quantity. In Table 3.5, the estimates show that introducing the hardcover increases the sales of paperback. The increase is statistically significant when the hardcover price is set at 160%∼230% and 440%∼790% of the paperback price.

In Table 3.6, the results of titles under the ebook and control conditions suggest that the introduction of ebook version increases the sales of paperback. The increase is statistically significant when the ebook price is set at the level of 45%∼55% and 65%∼80% of paperback price. It is interesting to note that the sales quantity of the paperback version increases when the ebook price is high and thus close to the paperback price. We will provide additional insights into this result in the next section where we examine the underlying mechanisms for the results.

3.4.6 The Impact on Overall Revenues

We also use our model to investigate the impact of extending the product line on the total revenues for NAP. Given that the revenues follow a long right tail distribution

Table 3.5: The Influence of Hardcover Pricing on Paperback Sales Quantity

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Variables	**160%-230%**	**230%-300%**	**300%-370%**	**370%-440%**	**440%-790%**
Hardcover	$-1.269***$	$-1.349***$	.125	.294	$-.713$
	(.370)	(.408)	(.433)	(.523)	(.510)
Post	$-.583**$	$-.576**$	$-.572**$	$-.602**$	$-.557**$
	(.264)	(.257)	(.256)	(.263)	(.251)
Hardcover×Post	.800*	.424	.345	1.305	2.109**
	(.419)	(.442)	(.469)	(.815)	(.999)
Email Campaigns	1.710***	.952***	.907***	1.686***	1.056***
	(.395)	(.248)	(.253)	(.607)	(.183)
Copyright Year	.0866	.0828	.114**	.0992*	.0842
	(.0562)	(.0556)	(.0538)	(.0568)	(.0592)
Number of Pages	.148	.162	.137	.165	.124
	(.118)	(.111)	(.116)	(.131)	(.109)
Constant	-1.216	-1.183	-1.385	-1.330	-1.145
	(.849)	(.833)	(.847)	(.880)	(.850)
Observations	2,744	2,744	2,744	2,744	2,744
Log-Likelihood	-6850	-6850	-6850	-6850	-6850
AIC	13780	13780	13780	13780	13780
BIC	14017	14017	14017	14017	14017

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: The Influence of Ebook Pricing on Paperback Sales Quantity

	(1)	(2)	(3)	(4)	(5)
Variables	**25%-35%**	**35%-45%**	**45%-55%**	**55%-65%**	**65%-80%**
Ebook	$-1.100***$	$-.465$	.168	$-.0327$	$-1.305***$
	(.320)	(.461)	(.400)	(.521)	(.368)
Post	$-.570**$	$-.539**$	$-.535**$	$-.579**$	$-.626**$
	(.253)	(.246)	(.267)	(.254)	(.280)
Ebook×Post	.344	.00466	2.592***	.0586	1.031***
	(.357)	(.556)	(.827)	(.616)	(.396)
Email Campaigns	.970***	.865***	2.764***	1.052***	2.694***
	(.257)	(.200)	(.825)	(.280)	(.695)
Copyright Year	.0917	.0841	.0898	.0667	.0948*
	(.0558)	(.0545)	(.0566)	(.0545)	(.0564)
Number of Pages	.147	.105	.0650	.167	.196
	(.114)	(.107)	(.0890)	(.124)	(.123)
Constant	-1.233	-1.110	-1.133	-1.065	-1.357
	(.845)	(.814)	(.802)	(.829)	(.869)
Observations	2,758	2,758	2,758	2,758	2,758
Log-Likelihood	-6891	-6891	-6891	-6891	-6891
AIC	13861	13861	13861	13861	13861
BIC	14098	14098	14098	14098	14098

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

which cannot be appropriately captured by the normal distributed error terms, we ran a DID model with the negative binomial distribution as in Equation (1) and (2). The results in Table 3.7 show that the total revenues increase under the hardcover treatment condition as well as under the ebook treatment condition.[3]

Table 3.7: The DID Estimates for the Effects of Introducing Hardcover or Ebook Version on Total Revenues

	Hardcover	**Ebook**
	(1)	**(2)**
Full Sample	.859**	1.246**
	(.403)	(.514)
Paperback Price: Low	2.427***	5.185***
	(.483)	(1.033)
Paperback Price: Medium	−.148	.183
	(.482)	(.526)
Paperback Price: High	5.185***	−.195
	(1.033)	(.339)
Popularity: Low	−.496	1.052
	(1.514)	(1.361)
Popularity: Medium	−.00523	−.683
	(.509)	(.616)
Popularity: High	.653*	.917**
	(.393)	(.379)
Copyright Year: 2000∼2004	.378	2.046***
	(.624)	(.578)
Copyright Year: 2005∼2010	−.0802	1.399
	(.586)	(.989)
Copyright Year: 2011∼2015	1.172***	.441
	(.407)	(.419)

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

[3] The results are robust when the blockbuster titles are excluded.

3.4.7 Discussions

The DID results show that introducing a new version, either the hardcover or ebook version, to the product line of book titles has a positive impact on the sales quantity of the existing premium version, the paperback version as well as overall revenue. There could be several explanations for the results as seen in Table 3.1. For example, the market expansion during the experiment duration could increase paperback sales. With the control group acting as a control for seasonality, one could argue that the market expansion is due to the addition of new versions (quality-association effect or variety effect) leading to increased traffic, which could affect the demand under the experimental conditions more positively than in the control condition. In fact, the DID analysis on free PDF downloads shows an increase under the treatment conditions of hardcover and ebook, showing clearly that market expansion effects exist. However, this effect could be confounded with the compromise effect in the hardcover case and the attraction effect in the ebook case. Therefore, in order to investigate the underlying behavioral mechanisms cleanly, we must control for this confound formally. In particular, we use the individual level choice analysis which we discuss next.

3.5 Underlying Mechanisms

The increase in demand for the paperback version under the hardcover treatment condition can be explained by the compromise effect (Simonson 1989). The addition of the higher-quality and higher-priced hardcover version to the already existing paperback option renders the paperback version as the compromise option in the choice set. A survey of NAP's customers showed (see Koukova, Kannan, and Kirmani 2012, page 105) that the image quality of the paperback version is generally perceived to be of higher quality than the PDF (the mean ratings of overall quality on a nine point scale with 1 = low and 9 = high were 7.81 (SD = 1.3) for paperback and 6.11 (SD = 1.8) for the PDF version, positioning the paperback version quality between that of the PDF version and the hardcover version, while the price of paperback is also between that of the free PDF and the higher-priced hardcover version. Thus, compromise effect predicts that adding the hardcover version will increase the choice probability of the paperback version as the paperback version becomes a middle option in the product line of hardcover, paperback and PDF versions.

The situation with respect to the ebook treatment condition is slightly more complex. On the one hand, ebooks generally have a lower price than paperback books, and arguably lower quality as perceived by the majority of NAP customers. From this perspective, an ebook would become the compromise option once added to the existing choice set of free PDF version and the more expensive paperback version. Since the ebook version was not offered prior to the experiment, this compromise effect will be confounded with the baseline quality of the ebook, and thus cannot be separately identified. Furthermore and more importantly, this compromise effect does not explain the change in paperback sales, which is positive.

Instead, with regard to the change in sales quantity of paperback titles under the ebook treatment condition, we argue that in specific situations the main underlying driver would be the attraction effect (Huber, Payne, and Puto 1982; Huber and Puto 1983; Simonson 1989). Ebooks are commonly sold at noticeably lower prices than paperback books[4], as a significant number of consumers consider the quality and costs of a paperback title to be higher.[5] In this context and given consumers' expectations, if an ebook is priced too closely to its paperback counterpart, some consumers may

[4] Based on a web survey of prices at Amazon, Barnes & Noble, Books-a-Million and Apple during 2012 and 2013, De los Santos and Wildenbeest (2017) find that ebook prices are around 50% of the list prices of print format.

[5] A survey of 55 NAP customers who had purchased paperback/e-book or downloaded PDF copies in the last one year provided an overall quality rating on a nine point scale with 1 = low and 9 = high of 8.02 (SD = 1.7) for paperback and 7.1 (SD = 2.1) for ebooks, with 40% of customers indicating that paperbacks had a higher overall quality than ebooks. 82% estimated the costs to be higher for paperbacks than for ebooks.

find the ebook price "unfairly high" even if the price is still technically lower than that of the paperback. In this situation, these consumers may perceive the ebook option to be dominated by the existing paperback option. This generates the attraction effect, where the introduction of a dominated option increases the attractiveness of the dominating one. When an ebook is introduced with a relatively high price, then, the sales of the paperback version can increase because of this attraction effect.

Finally, given that one of the options in the product line is free PDF, we expect the presence of zero price effect (Shampanier, Mazar, and Ariely 2007) as the free PDF may play a special role in influencing customers' choices. We cannot identify the zero-price effects from our study. However, since we expect the zero-price effect to have a negative impact on the demand for paperback in both the treatment conditions, thus countering the impacts of the compromise effect and the attraction effect, we argue that our tests in the main results are conservative tests in showing the evidence for these effects. We leave the estimation of zero-price effects in such contexts for future research.

3.5.1 Modeling the Compromise and Attraction Effects

In order to empirically examine the extent of the compromise effect and the attraction effect, we use the individual level purchase data to analyze customers' choice decisions between different versions. We use a multinomial logit model for this analysis. Each choice incidence corresponds to one customer purchasing from a choice set consisting of {PDF, paperback, hardcover} or {PDF, paperback, ebook}. In other words, we analyze the choice among different versions conditional on the customer purchasing or downloading the book. We model this conditional choice for two reasons. First, we do not observe the non-purchase incidences, where a customer browsed through the different versions but decided not to buy. Second, as discussed earlier, when a new version is introduced, there could be a market expansion effect that lifts the utility of all options relative to the outside option. Modeling the conditional choices controls for this market expansion effect, and helps to identify the compromise effect and attraction effect which only changes the relative appeal of different versions.

Overall, there are 742,991 choice incidences, in most of which customers choose the free PDF. Moreover, customers choose the paperback version in 1,024 incidences before the experiment and 1,049 incidences during the experimental period. After the introduction of new versions, customers choose the hardcover version in only six choice incidences while they choose the ebook version in 115 incidences. In addition, we observe 338 choice incidences in which customers purchase multiple copies of the same title in paperback version. There is no incidence in which customers purchase multiple copies of the same title in hardcover, ebook or PDF version. Finally, we only observe two cases where a customer purchases different versions of the same title at the same time. Therefore, for modeling purpose, we consider the purchases for different versions of the same title as separate choice incidences. In each choice incidence, a customer purchases one version of one title.

In choice incidence i across customers and time, the customer's utility of choosing version j, $j \in \{\text{PDF, paperback, hardcover, ebook}\}$ is:

$$u_{ij} = v_{ij} + \epsilon_{ij} = \alpha_j + \gamma \cdot D_{ij}^{\text{treat}} + \delta \cdot \text{price}_{ij} + \epsilon_{ij} \tag{3.4}$$

In this equation, D_{ij}^{treat} is a dummy variable which equals 1 if: 1) the title in choice incidence i is assigned to the treatment condition, 2) the choice incidence i happens after the experiments started, and 3) version j is paperback. All three conditions must be satisfied for the variable to be 1. As a result, γ captures the change in the utility of the paperback version by adding hardcover or ebook version to the product line. In addition, price_{ij} is the purchase price of version j of title at choice incidence i. Finally, ϵ_{ij} is the error term clustered by customers.

For the hardcover treatment group, we expect that the compromise effect exists, and thus we expect γ to be positive. For the ebook treatment group, we expect that the attraction effect exists if the ebook price is close enough to the paperback price to allow the paperback version to dominate the ebook version. In this model setup, the attraction effect for that subset of the book titles will be loaded onto the parameter γ at the entire treatment group level, and thus we also expect γ to be positive. To identify the attraction effect with better precision, we also extend the model to separate the high-price situations of the ebook from the non-high-price situations. For titles under the ebook condition, we further add a dummy variable, $D_{ij}^{\text{high.ebook}}$, which equals 1 if D_{ij}^{ebook} equals 1 and the price level of the ebook version is relatively high. We set a threshold level of 65%, i.e., the dummy variable is set to 1 if the price of ebook version is higher than 65% of the price of paperback price, and 0 otherwise.

$$u_{ij} = v_{ij} + \epsilon_{ij} = \alpha_j + \gamma \cdot D_{ij}^{\text{ebook}} + \lambda \cdot D_{ij}^{\text{high.ebook}} + \delta \cdot \text{price}_{ij} + \epsilon_{ij} \tag{3.5}$$

Customers perceive the ebook to be more similar to the paperback version than the PDF version when the ebook price is high and close to the paperback price, creating dominance over the ebook version by the paperback version. Since we expect the attraction effect to exist, we expect the coefficient of dummy variable D_{ij}^{ebook} to become smaller in magnitude and insignificant after adding the dummy variable $D_{ij}^{\text{high.ebook}}$. Furthermore and more importantly, we expect the coefficient for $D_{ij}^{\text{high.ebook}}$ to be positive and significant.

For each choice incidence, we use free PDF as the baseline version. Assuming ϵ_{ij} follows a type I extreme value distribution, the choice probability function is given by

$$\Pr(y_i = j) = \frac{\exp\{v_{ij}\}}{\sum\limits_{k=1}^{J_i} \exp\{v_{ik}\}} \tag{3.6}$$

where J_i is the number of total versions in choice incidence i. The log-likelihood function is then given by

$$\ln L\,(y|\boldsymbol{x};\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{j=1}^{J_i}\{d_{ij}\cdot \ln \Pr(y_i = j)\} \tag{3.7}$$

where $d_{ij} = 1$ if $y_i = j$ and $d_{ij} = 0$ otherwise; N is the total number of choice incidences; and J_i is the number of product versions in choice incidence i.

3.5.2 Estimation Results

Columns (1) and (2) of Table 3.8 present our estimation results of the multinomial logit model to identify the underlying factors. First, the estimates for dummy variable D^{hard} and D^{ebook} are statistically significant and positive, indicating that paperback version gains more preference from adding a new premium version, which is consistent with our DID estimates. Based on model specification in column (1), introducing a hardcover version to the product line increase the logarithmic of the odds of purchasing a paperback relative to downloading a PDF by 16.2%. Based on model specification in column (2), introducing an ebook version to the product line increase the log odds of purchasing a paperback relative to downloading a PDF by 22.6%. Second, the coefficients for the paperback price are significantly negative, showing that customers are more likely to download a free PDF if the paperback prices are higher (the prices of hardcover and ebook are also higher). In general, the estimation results of the choice model are consistent with results of the aggregate analysis.

Moreover, the results in Column (3) confirm our expectation. On the one hand, the coefficient for D^{ebook} becomes insignificant after adding the new dummy variable and its magnitude changes from .226 (Column (2)) to .0809 (Column (3)). On the other hand, the coefficient for $D^{\text{high.ebook}}$ is positive and significant. Taken together, the results in Columns (2) and (3) are strong evidence that the increase of paperback sales quantity for titles under the ebook condition is due to the attraction effect.

Table 3.8: Multichannel Logit Estimates for Identifying the Underlying Factors

Variables	Hardcover & Control Titles (1)	Ebook & Control Titles Spec. I (2)	Ebook & Control Titles Spec. II (3)	All Titles Spec. I (4)	All Titles Spec. II (5)
D^{hard}	.162**	–	–	.179**	.179**
	(.0751)	–	–	(.0736)	(.0736)
D^{ebook}	–	.226***	.0809	.167***	.0229
	–	(.0634)	(.0786)	(.0629)	(.0783)
$D^{\text{high.ebook}}$	–	–	.344***	–	.340***
	–	–	(.0973)	–	(.0972)
Price	−.0211***	−.0194***	−.0191***	−.0204***	−.0202***
	(.00261)	(.00204)	(.00205)	(.00184)	(.00184)
Constant (Hardcover)	−7.294***	–	–	−7.359***	−7.383***
	(.535)	–	–	(.505)	(.505)
Constant (Ebook)	–	−6.650***	−6.657***	−6.628***	−6.633***
	–	(.106)	(.106)	(.104)	(.104)
Constant (Paperback)	−5.050***	−5.190***	−5.204***	−5.094***	−5.104***
	(.120)	(.0916)	(.0921)	(.0838)	(.0840)
Observations	987,496	1,181,517	1,181,517	1,729,337	1,729,337
Log-Likelihood	−8151	−10565	−10559	−15139	−15133
AIC	16309	21139	21128	30291	30280
BIC	16356	21187	21188	30365	30367

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.5.3 Robustness Checks

Additional analyses support the robustness of these results to alternative specifications of choice model. First, we use a binary logit model to estimate the probability of purchasing a paperback version given the product line of book titles. Second, we use a nested logit model to relax the assumption of independence of irrelevant alternatives. We test two nest structures with data of titles under the ebook and control conditions: (i) free (i.e., PDF) vs. non-free (paperback and ebook); and (ii) digital (PDF and ebook) vs. non-digital (paperback) (see Table C10 and C11 in Appendix C). Finally, we use an alternative specification of multinomial logit model, which allows us to include the data of all titles.

$$u_{ij} = \alpha_j + \gamma_1 \cdot D_{ij}^{\text{hard}} + \gamma_2 \cdot D_{ij}^{\text{ebook}} + \gamma_3 \cdot D_{ij}^{\text{high.ebook}} + \delta \cdot \text{price}_{ij} + \epsilon_{ij} \tag{3.8}$$

where D_{ij}^{hard} is for the hardcover condition and D_{ij}^{ebook} and $D_{ij}^{\text{high.ebook}}$ are for the ebook condition.[6]

In summary, our results consistently demonstrate that introducing a hardcover or ebook version has a positive impact on paperback sales quantity. Moreover, when a hardcover version is added to the product line, the choice probability of the paperback version increases relative to that of a PDF version. Consistent with the compromise effect, the paperback version becomes a compromise option in the choice set and thus becomes more attractive to customers. When an ebook version is added to the product line and its price is set close to the paperback price, the relative choice probability of paperback version to that of a PDF version also increases. It is because the paperback version tends to dominate the ebook version, that this leads to an attraction effect.

3.6 Discussions and Managerial Implications

Our analyses at the aggregate level and the individual customer level have confirmed that there is increased demand for the existing premium alternative as well as an increase in overall revenue when a new premium version is introduced - hardcover

[6] In order to account for heterogeneity, we have also run the random coefficient model allowing the parameter for price_{ij} to be random. The results are robust.

or ebook. The aggregate analysis may seem to imply that the increased demand for the existing premium version, overall revenue and the free version is due to the market expansion effect. However, the individual level analysis makes it clear that the increase in the sales of the existing premium version comes at the expense of the free version – the compromise effect playing a role when the higher-quality, higher priced alternative is introduced, and the attraction effect playing a role when the lower-quality but similarly priced alternative is introduced. The confirmation of these effects in the context of product line extension research is a significant contribution as it highlights conditions and mechanisms under which a product line extension can be successful in increasing revenue. In the case of the freemium model, the results show that such an line extension strategy can even overcome the zero-price effect and lead to higher overall revenues if the new premium alternatives are appropriately priced. This is a significant contribution in the freemium context as this is the first time such strategies have been shown to have positive impact.

In the case of NAP, the results imply that an extended premium product line can be more attractive and can lead customers to the website due to the variety effect and at the same time increase sales of paperback by moving customers from free PDF downloads to the existing premium alternative. Also, our sub-sample analyses using the aggregate data and DID model reveal that the above effects are more pronounced for book titles which are more popular or lower in price. These results make intuitive sense — the lower priced paperbacks are more likely to be within consumers' threshold of willingness-to-pay for the titles and thus the compromise effect or the attraction effect manifest themselves more easily than in the case of higher priced paperbacks, which may be beyond the threshold of willingness-to-pay for many consumers. Similarly, the more popular the titles are, the higher the consumers' willingness-to-pay for the titles is, leading to significant compromise or attraction effects. The sub-sample analyses also highlight that adding a hardcover version has a greater impact on paperback sales of those titles published more recently. Given that the hardcover option becomes more salient for newer titles and the willingness-to-pay could be higher for newer titles, it becomes easier for the compromise effect to manifest itself under these conditions. In the case of older titles, this threshold could be lower and the lower priced ebook version becomes more salient leading to the observed result. Overall, these results help NAP to determine which titles to focus on for its initial roll-out of the additional premium version.

If NAP were to introduce an additional version which version should they introduce and at what price? To help answer this question, we performed a scenario analysis based on the results of the DID models. As seen in Table 3.9, NAP could offer hardcover versions, but given that hardcover versions did not sell well in the experiment and that the impact of increase in demand for the paperback version is lower than when an ebook version is introduced, the hardcover option is not ideal. The ebook version, on the other hand, sells more and when priced closely to the paperback version also has a positive impact on paperback sales. Table 3.9 provides the impact of relative prices of the sales of ebooks as well as the impact on paperback sales. When the ebook version is priced lower, its impact on paperback sales is neutral (or negative), resulting in a trade-off in the pricing decision. Taking the trade-offs into account, we find that pricing the ebook version at 70%-80% of the paperback price is optimal for NAP.

Table 3.9: The Impact of Pricing Level of New Versions on Revenues

Hardcover Condition	**160-220%**	**220-300%**	**300-330%**	**330-360%**	**360-390%**	**390-440%**	**440-790%**
% Revenue (Hardcover)	0.26%	0.25%	0.00%	0.30%	0.00%	1.57%	0.00%
% Revenue (Paperback)	1.38%	3.24%	6.22%	−9.00%	−6.96%	8.90%	3.58%
% Revenue (Total)	1.64%	3.49%	6.22%	−8.71%	−6.96%	10.47%	3.58%

Ebook Condition	**25-34%**	**34-40%**	**40-49%**	**49-54%**	**54-61%**	**61-70%**	**70-80%**
% Revenue (Ebook)	0.44%	0.62%	0.20%	0.64%	0.60%	0.23%	1.45%
% Revenue (Paperback)	1.73%	−3.73%	1.60%	−4.35%	−6.86%	5.53%	21.45%
% Revenue (Total)	2.17%	−3.11%	1.80%	−3.71%	−6.26%	5.75%	22.90%

While the results of our freemium study are set in the context of books, the findings should extend readily to other freemium implementations as we discussed earlier. Streaming video or audio services, such as YouTube or Pandora, are quite similar to the NAP context. Currently, the free version and the premium version of these services are differentiated on the basis of advertisement as an annoyance in the free version, with the quality, in terms of resolution, being almost the same between the two versions. One could conjecture that adding another premium version and differentiating the two versions in terms of resolution or speed (given the relaxation of net neutrality rules) could lead to similar compromise effects. Another context where extending the product line of premium versions could be fruitful is for software and online services, such as Skype or LinkedIn. Firms can differentiate the quality of the premium versions and the free version based on the availability of different functions and features. Finally, the findings would also extend to contexts such as newspapers and other ad-supported content platforms (e.g., Pauwels and Weiss 2008; Kanuri, Mantrala, and Thorson 2017; and Sridhar and Sriram 2015). That is, our findings should apply to instances where premium services dominate the free option on the quality dimension. It should be noted however that in our context, we defined quality in terms of resolution and production quality. This dimension was also very salient for the customer base of NAP consisting mainly of academics and highly educated readers. But this may not be the case for other types of content and other types of customer base. For example, in the case of newspapers, quality dimension based on resolution and production quality may not be a salient differentiator. Instead, quality could be the portability of the content/device, which could enable easy and ubiquitous access. Thus, it is important for firms in different contexts to understand on what dimensions they should vertically differentiate their premium versions based on what dimensions customers consider salient (e.g., Kumar and Shah 2011). The magnitude of demand increase through such product line extensions is, of course, dependent on the particular context of the freemium model, which could be tested using a field experiment. Nevertheless, the take-away from our research should be very encouraging for firms trying to increase revenues using their freemium model implementations.

3.7 Limitations and Conclusions

With firms increasingly resorting to the freemium model to sell their products and services, it is important that research focus on understanding how firms can make them profitable and sustainable. While extant research has focused on the design and pricing issues in such models, it has mainly been in the context of a single premium product. We extend this literature by focusing on the product line problem and use a field experiment to highlight how such strategies can lead to increased sales. From a substantive viewpoint, our results should be very valuable to both academics and practitioners. We not only show how firms can increase their sales in the freemium models, but also uncover possible behavioral mechanisms that contribute to such increases. From the perspective of product line research, our research highlights the role of the behavioral mechanisms such as the compromise effect and attraction effect as alternative explanations for the demand patterns seen in the field experiment.

We, however, need to point out several limitations of our work so that we do not oversell the identification of the behavioral mechanisms. After all, the mechanisms have been teased out of a study where the major focus was on the aggregate data and as such the study was not specifically designed for ruling out other possible effects. For example, extant research (Frederick, Lee, and Baskin 2014) has shown that attraction effect is highly context dependent and may depend on how the items are presented to the customers. Also, we show these effects to be present under conditions of low price and high popularity. Although we provide a possible reason for such boundary conditions, more research is needed to understand why these become boundary conditions. Such research may require field experiments specifically designed for such retail conditions (e.g., Parker and Lehmann 2011).

Another limitation of this research is that we cannot clearly identify to what extent the zero-price effect has dampened the effect of the compromise and the attraction effects. This identification is needed to clearly understand how they affect the demand patterns in a product line setting/ Ideally, future empirical study should separately estimate the zero-price effect. Additionally, there is a possibility that the customers in the field experiment viewed different experimental conditions. So the experimental design could be improved in the future to avoid such problems. Also, our field experiment was conducted for one category. However, additional examinations may extend to a larger set of categories such as those of hedonic or luxury products, which may introduce other moderating effects. We leave these questions for future research. Our effort will hopefully encourage other studies researching these issues.

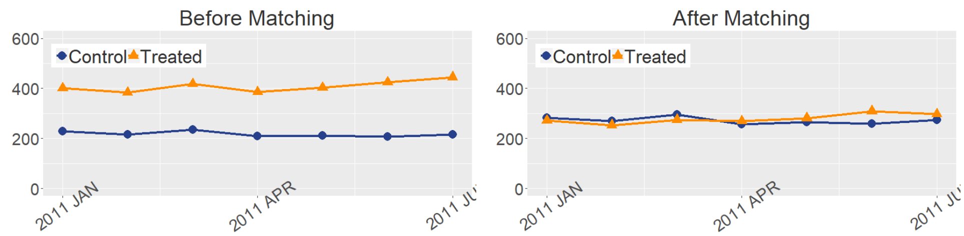
Appendices

Appendix A: Chapter 1

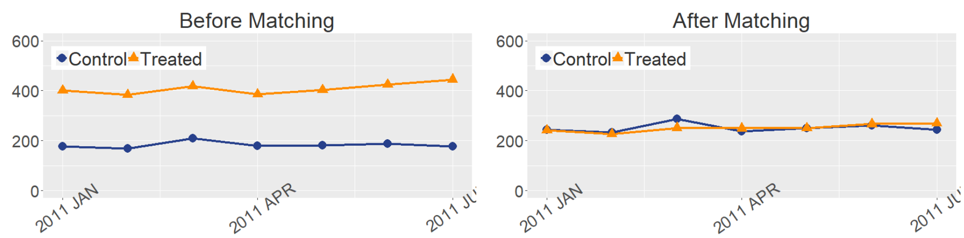
Figure A1: Real Customer Example of Aggregation Bias

Notes: The vertical dashed line denotes the month when the four customers adopted the mobile app.

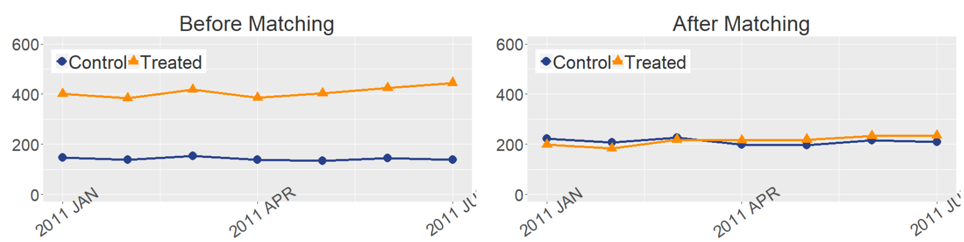
Figure A2: Graphical Test of Common-Trend Assumption



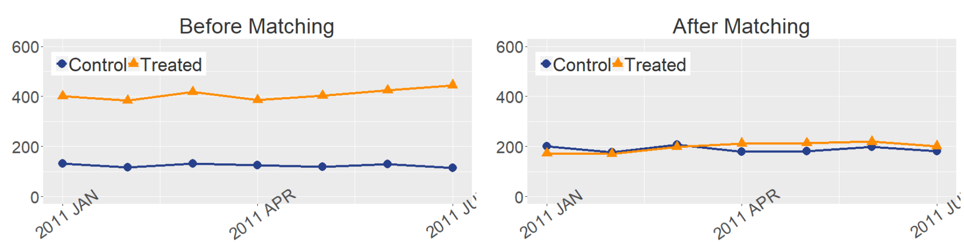
(a) **Cohort 1 vs 3**



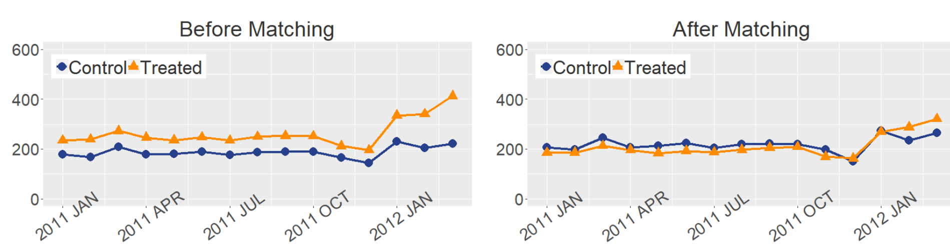
(b) **Cohort 1 vs 4**



(c) **Cohort 1 vs 5**

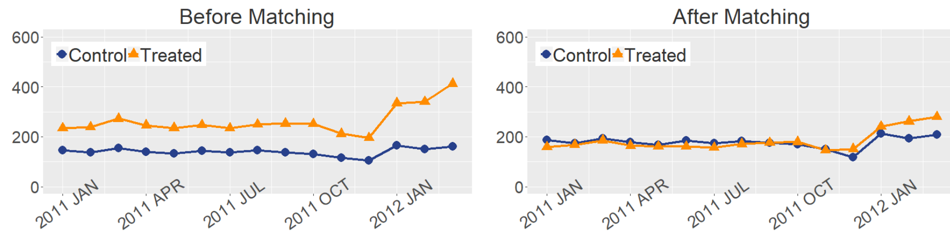


(d) **Cohort 1 vs 6**

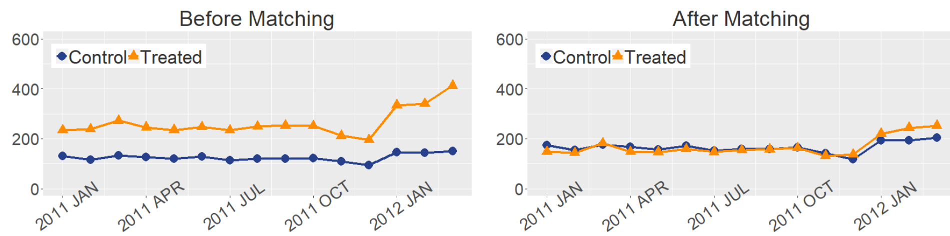


(e) **Cohort 2 vs 4**

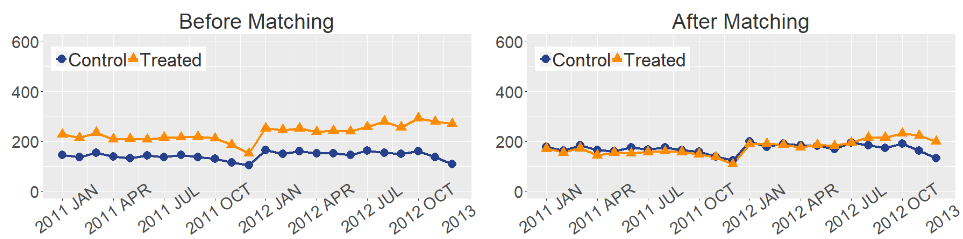
Figure A2: Graphical Test of Common-Trend Assumption (continued)



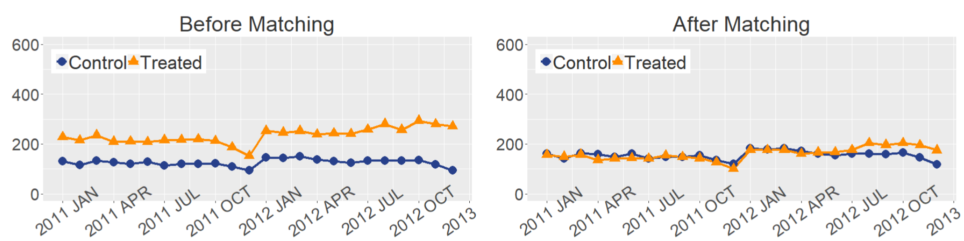
(f) Cohort 2 vs 5



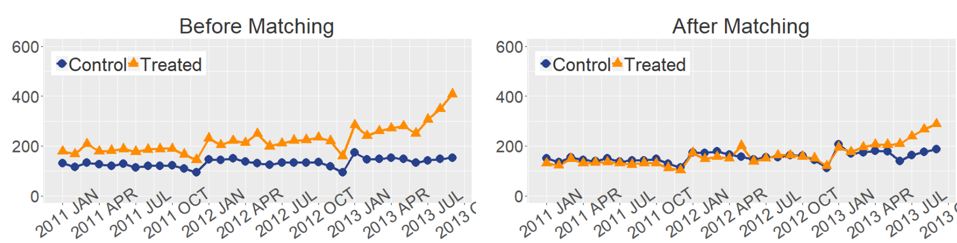
(g) Cohort 2 vs 6



(h) Cohort 3 vs 5



(i) Cohort 3 vs 6



(j) Cohort 4 vs 6

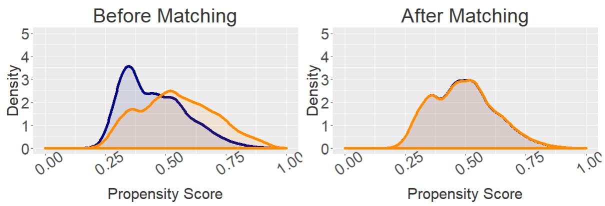
Table A1: Impact of Customer App Adoption on Total Spending by Customer Segment

Segments	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
Elite Customer										
DID	−0.0495*	−0.0467	−0.00211	−0.0404	−0.224***	−0.178***	−0.147***	−0.151***	−0.0916***	−0.219***
DID-Heckman	−0.0498	−0.0414	0.00522	−0.0363	−0.217***	−0.170***	−0.140***	−0.148***	−0.0863**	−0.222***
DID-Matching	−0.0181	−0.00122	−0.00979	−0.0366	−0.194***	−0.144***	−0.135***	−0.109***	−0.0358	−0.175***
Non-Elite Customer										
DID	−0.184***	−0.181***	−0.140***	−0.174***	−0.273***	−0.157***	−0.124***	−0.172***	−0.165***	−0.400***
DID-Heckman	−0.181***	−0.181***	−0.142***	−0.173***	−0.270***	−0.153***	−0.122***	−0.172***	−0.165***	−0.405***
DID-Matching	−0.0296	0.00124	−0.0539	−0.0615	−0.147***	−0.0261	−0.027	−0.122***	−0.135***	−0.317***
Heavy Spender										
DID	−0.0126	−0.000171	−0.0158	−0.0921*	−0.160***	−0.172***	−0.00276	−0.0731	−0.0407	−0.0837*
DID-Heckman	−0.0138	0.00348	−0.00590	−0.0909**	−0.151***	−0.167***	0.000688	−0.0721	−0.0418	−0.0836*
DID-Matching	0.0365	0.0299	0.0104	−0.0619	−0.128***	−0.112**	0.0917	−0.00202	0.0359	−0.0472
Light Spender										
DID	−0.139***	−0.0996**	−0.0993**	−0.0939**	−0.243***	−0.104***	−0.127***	−0.112***	−0.0957***	−0.348***
DID-Heckman	−0.138***	−0.102**	−0.106***	−0.0962**	−0.237***	−0.0997***	−0.125***	−0.112***	−0.0956***	−0.358***
DID-Matching	−0.0908*	0.0238	−0.0173	−0.00609	−0.117***	−0.006	−0.0456	−0.0687	−0.0696*	−0.259***
Digital Channel User										
DID	−0.114***	−0.101***	−0.0312	−0.0677**	−0.287***	−0.208***	−0.170***	−0.208***	−0.193***	−0.331***
DID-Heckman	−0.115***	−0.0999***	−0.0283	−0.0660**	−0.285***	−0.206***	−0.169***	−0.209***	−0.195***	−0.331***
DID-Matching	−0.0929***	−0.0698**	−0.0719**	−0.103***	−0.224***	−0.143***	−0.173***	−0.168***	−0.184***	−0.319***
Non-Digital Channel User										
DID	−0.313***	−0.324***	−0.261***	−0.257***	−0.489***	−0.318***	−0.321***	−0.323***	−0.226***	−0.490***
DID-Heckman	−0.317***	−0.322***	−0.262***	−0.257***	−0.487***	−0.314***	−0.322***	−0.318***	−0.224***	−0.497***
DID-Matching	−0.229**	−0.0959	−0.192**	−0.131	−0.382***	−0.136	−0.266***	−0.258***	−0.12	−0.340***

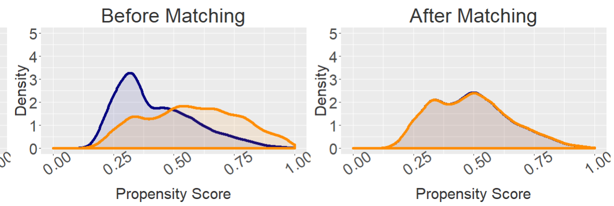
Clustered standard errors are used

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

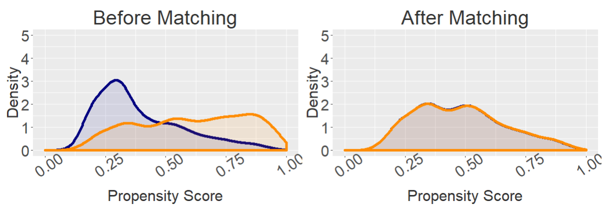
Figure A3: Distribution of Propensity Scores before and after Matching



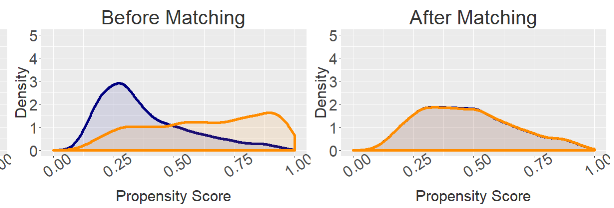
(a) Cohort 1 vs 3



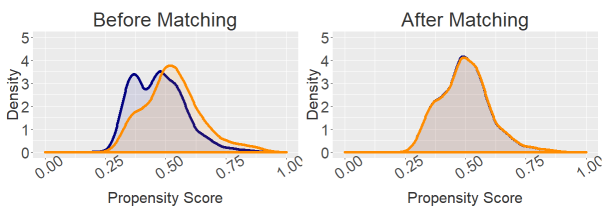
(b) Cohort 1 vs 4



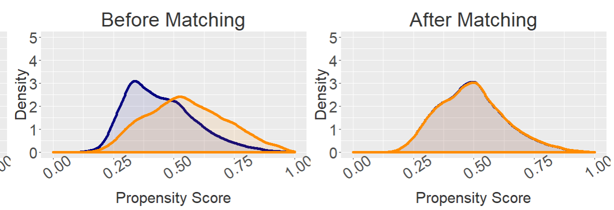
(c) Cohort 1 vs 5



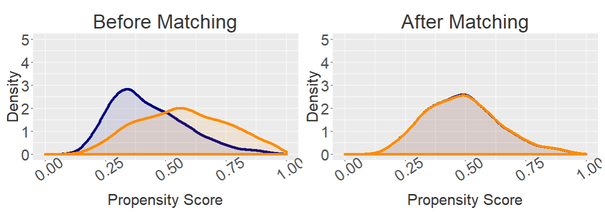
(d) Cohort 1 vs 6



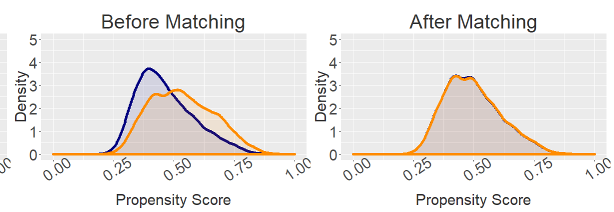
(e) Cohort 2 vs 4



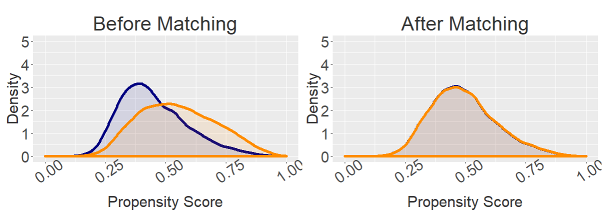
(f) Cohort 2 vs 5



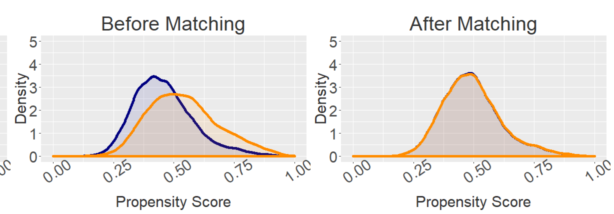
(g) Cohort 2 vs 6



(h) Cohort 3 vs 5



(i) Cohort 3 vs 6



(j) Cohort 4 vs 6

Table A2: Sample Size & Kolmogorov-Smirnov Test

	Size of Each Cohort			K-S Statistic	
Cohort Pair	**Before Matching**	**After Matching**	**Size Change**	**Before Matching**	**After Matching**
Cohort 1 vs 3	10,000	7,461	−25.39%	0.2573***	0.0058
Cohort 1 vs 4	10,000	6,625	−33.75%	0.3403***	0.0054
Cohort 1 vs 5	10,000	5,914	−40.86%	0.4101***	0.0066
Cohort 1 vs 6	10,000	5,552	−44.48%	0.4473***	0.0059
Cohort 2 vs 4	10,000	7,904	−20.96%	0.2125***	0.0061
Cohort 2 vs 5	10,000	7,192	−28.08%	0.2840***	0.0061
Cohort 2 vs 6	10,000	6,698	−33.02%	0.3319***	0.0058
Cohort 3 vs 5	10,000	7,714	−22.86%	0.2318***	0.0049
Cohort 3 vs 6	10,000	7,307	−26.93%	0.2723***	0.0055
Cohort 4 vs 6	10,000	7,516	−24.84%	0.2506***	0.0060

*** p<0.01, ** p<0.05, * p<0.1

Table A3: Probit Regression for App Adoption (Heckman Selection Equation)

Variables	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
Pre-Adoption Purchase										
Mobile Website Usage	0.630***	0.955***	0.914***	1.067***	0.722***	0.747***	0.833***	0.346***	0.489***	0.374***
Booking Window (×30 days)	0.122***	0.120***	0.163***	0.0591***	−0.00287	0.0501*	−0.0542**	0.0921***	−0.0407	0.0536*
ΔSpending (×1,000 $)	0.0348***	0.0567***	0.0627***	0.0851***	0.0797***	0.0878***	0.0899***	0.0586***	0.0710***	0.132***
Number of Room Nights	0.0191***	0.0316***	0.0257***	0.0341***	0.0206***	0.0247***	0.0356***	0.00235	0.0201***	0.0452***
% of Online Spending	0.402***	0.467***	0.482***	0.537***	0.431***	0.479***	0.564***	0.300***	0.437***	0.526***
% of Offline Spending	0.0983***	0.180***	0.124***	0.0713***	0.265***	0.216***	0.123***	0.191***	0.128***	0.118***
% of GDS Spending	0.141***	0.0799***	0.0601**	0.126***	0.121***	0.138***	0.178***	0.131***	0.227***	0.372***
% of OTA Spending	0.102	0.172**	0.360***	0.120	0.248***	0.366***	0.151**	0.355***	0.175**	0.159**
Pre-Adoption Reservation Characteristics										
Luxury Brand	−0.433***	−0.217*	−0.316**	0.00480	0.210	0.0612	0.416**	0.0314	0.301	0.561**
Premium Brand	−0.0658	−0.177**	−0.0299	0.400***	−0.137	0.306***	0.553***	0.102	0.402***	1.464***
Upscale Brand	−0.105	0.0703	0.129	0.397***	0.191*	0.386***	0.480***	0.181	0.265*	0.351**
Economy Brand	−0.0512	0.185**	0.188**	0.530***	0.299***	0.466***	0.667***	0.253*	0.639***	0.605***
Franchised Hotel	0.132**	0.0702	0.0634	0.0698	0.0543	0.214**	0.0615	0.285***	0.0393	0.0765
US Hotel	0.404***	0.227***	0.562***	0.403***	0.230**	0.472***	0.432***	0.326***	0.270**	0.741***
Weekend Stay	0.321***	0.477***	0.349***	0.564***	0.527***	0.486***	0.651***	0.283***	0.583***	0.421***
Holiday Stay	0.178***	0.00871	0.148**	0.0804	0.378***	0.317***	0.358***	0.794***	0.756***	−0.795***
Airport Area	−0.0327	−0.000510	−0.164**	−0.329***	−0.246***	−0.513***	−0.511***	−0.385***	−0.358***	−0.536***
Downtown Area	0.221***	0.183***	0.0890	−0.0805	−0.114	−0.272***	−0.313***	−0.307***	−0.241**	−0.549***
Resort Area	0.0656	0.0651	−0.0101	−0.198***	−0.267***	−0.395***	−0.421***	−0.288***	−0.211*	−0.376***
Suburban Area	−0.124**	−0.171***	0.0858	−0.213***	−0.241***	−0.00457	−0.263***	0.221**	−0.0344	−0.201
Expressway Area	−0.0761	0.0825	−0.200**	−0.217**	0.00370	−0.442***	−0.431***	−0.447***	−0.452***	−0.589***
Metro Area	0.0261	−0.0291	−0.0348	−0.114	−0.220***	−0.353***	−0.345***	−0.365***	−0.394***	−0.455***
Customer Demographics										
Platinum Member	−0.0255	0.240***	0.464***	0.403***	−0.101**	0.0259	−0.0198	0.239***	0.114***	−0.456***
Gold Member	−0.133***	0.150***	0.245***	0.269***	0.00468	0.00238	0.0708*	0.196***	0.210***	−0.374***
Silver Member	−0.131***	0.0347	0.0711**	0.00956	0.0124	−0.0220	−0.0308	0.0315	−0.0145	−0.356***
Basic Member	−0.208***	−0.233***	−0.307***	−0.324***	−0.166***	−0.276***	−0.264***	−0.254***	−0.278***	−0.339***
Age	−0.00834***	−0.00986***	−0.0138***	−0.0189***	−0.00698***	−0.0118***	−0.0177***	−0.00780***	−0.0140***	−0.0125***
Female	−0.155***	−0.252***	−0.320***	−0.275***	−0.0909***	−0.162***	−0.120***	−0.181***	−0.143***	−0.0511***
Constant	−0.0424	−0.0446	0.0864***	0.296***	−0.0493	0.0994***	0.339***	0.0377	0.264***	0.234***

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4: Logit Regression for App Adoption (Propensity Score Matching)

Variables	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
Pre-Adoption Purchase										
Mobile Website Usage	1.055***	1.707***	1.625***	1.967***	1.209***	1.278***	1.490***	0.571***	0.833***	0.620***
Booking Window (×30 days)	0.193***	0.199***	0.271***	0.102**	−0.005	0.082	−0.089	0.155**	−0.069	0.080
ΔSpending (×1,000 $)	0.061***	0.098***	0.113***	0.156***	0.153***	0.177***	0.182***	0.109***	0.136***	0.234***
Number of Room Nights	0.041***	0.060***	0.052***	0.069***	0.036***	0.045***	0.062***	0.006	0.035**	0.075***
% of Online Spending	0.657***	0.763***	0.797***	0.890***	0.690***	0.775***	0.919***	0.485***	0.709***	0.850***
% of Offline Spending	0.158**	0.294***	0.205***	0.118*	0.421***	0.345***	0.200***	0.307***	0.206***	0.189***
% of GDS Spending	0.231***	0.135**	0.097	0.208***	0.189***	0.216***	0.286***	0.213***	0.369***	0.597***
% of OTA Spending	0.167	0.283*	0.594***	0.204	0.394**	0.588***	0.242	0.571***	0.290*	0.258
Pre-Adoption Reservation Characteristics										
Luxury Brand	−0.712**	−0.366	−0.520	0.056	0.329	0.109	0.682	0.050	0.484	0.941
Premium Brand	−0.117	−0.304*	−0.012	0.720***	−0.233	0.526*	0.970***	0.167	0.698**	2.454***
Upscale Brand	−0.181	0.113	0.268	0.727***	0.306	0.669**	0.854***	0.297	0.471	0.564
Economy Brand	−0.085	0.324*	0.398*	1.016***	0.486*	0.824***	1.182***	0.445	1.114***	0.993**
Franchised Hotel	0.210	0.120	0.103	0.081	0.094	0.367*	0.093	0.471**	0.076	0.146
US Hotel	0.643***	0.340**	0.859***	0.569***	0.371*	0.720***	0.631**	0.509*	0.375	1.173***
Weekend Stay	0.510***	0.814***	0.594***	0.978***	0.871***	0.794***	1.091***	0.463***	0.970***	0.720***
Holiday Stay	0.258*	0.002	0.246	0.129	0.626***	0.554**	0.657***	1.296***	1.275***	−1.310***
Airport Area	−0.057	−0.002	−0.289*	−0.546***	−0.409**	−0.878***	−0.852***	−0.634**	−0.586**	−0.868***
Downtown Area	0.354**	0.301**	0.136	−0.125	−0.195	−0.453*	−0.517**	−0.483*	−0.374	−0.873***
Resort Area	0.102	−0.202	−0.226	0.409	−0.684*	−0.952**	−0.040	−0.795*	0.569	−0.459
Suburban Area	0.100	0.108	−0.035	−0.330*	−0.440**	−0.680***	−0.724***	−0.463*	−0.341	−0.578*
Expressway Area	−0.124	0.145	−0.362*	−0.350	0.005	−0.745***	−0.726**	−0.726**	−0.758**	−0.983***
Metro Area	0.048	−0.046	−0.067	−0.163	−0.368*	−0.596***	−0.565**	−0.594**	−0.648**	−0.752**
Customer Demographics										
Platinum Member	−0.060	0.378***	0.763***	0.648***	−0.167*	0.039	−0.042	0.376***	0.173*	−0.745***
Gold Member	−0.227***	0.218***	0.376***	0.403***	0.001	−0.005	0.096	0.305***	0.327***	−0.627***
Silver Member	−0.218***	0.041	0.093	−0.015	0.017	−0.044	−0.065	0.041	−0.037	−0.590***
Basic Member	−0.338***	−0.386***	−0.509***	−0.545***	−0.266***	−0.447***	−0.436***	−0.413***	−0.453***	−0.557***
Age	−0.014***	−0.016***	−0.023***	−0.032***	−0.011***	−0.019***	−0.029***	−0.013***	−0.023***	−0.020***
Female	−0.252***	−0.414***	−0.531***	−0.460***	−0.146***	−0.264***	−0.201***	−0.293***	−0.233***	−0.082**
Constant	−0.064	−0.069	0.161**	0.529***	−0.081	0.168**	0.578***	0.064	0.438***	0.379***

*** p<0.01, ** p<0.05, * p<0.1

Table A5: Treatment Effects with Alternative Time Frame

	(1) Cohort 2 vs 4	(2) Cohort 2 vs 5	(3) Cohort 2 vs 6	(4) Cohort 3 vs 5	(5) Cohort 3 vs 6	(6) Cohort 4 vs 6
Time Frame of Two Years						
All Channels						
DID	−0.340∗∗∗	−0.225∗∗∗	−0.209∗∗∗	−0.263∗∗∗	−0.196∗∗∗	−0.421∗∗∗
DID-Heckman	−0.336∗∗∗	−0.218∗∗∗	−0.200∗∗∗	−0.263∗∗∗	−0.192∗∗∗	−0.420∗∗∗
DID-Matching	−0.272∗∗∗	−0.201∗∗∗	−0.217∗∗∗	−0.216∗∗∗	−0.176∗∗∗	−0.363∗∗∗
Inner Channels						
DID	−0.212∗∗∗	−0.0926∗∗∗	−0.0756∗∗∗	−0.134∗∗∗	−0.0568∗∗	−0.234∗∗∗
DID-Heckman	−0.203∗∗∗	−0.0740∗∗∗	−0.0557∗∗	−0.124∗∗∗	−0.0437∗	−0.232∗∗∗
DID-Matching	−0.102∗∗∗	−0.0228	−0.0310	−0.0760∗∗∗	−0.0248	−0.176∗∗∗
Share of Inner Channels						
DID	0.0192∗∗∗	0.0232∗∗∗	0.0227∗∗∗	0.0159∗∗∗	0.0283∗∗∗	0.0136∗∗∗
DID-Heckman	0.0204∗∗∗	0.0262∗∗∗	0.0262∗∗∗	0.0193∗∗∗	0.0321∗∗∗	0.0135∗∗
DID-Matching	0.0426∗∗∗	0.0455∗∗∗	0.0437∗∗∗	0.0268∗∗∗	0.0376∗∗∗	0.0252∗∗∗
Outer Channels						
DID	−0.247∗∗∗	−0.203∗∗∗	−0.192∗∗∗	−0.199∗∗∗	−0.200∗∗∗	−0.215∗∗∗
DID-Heckman	−0.254∗∗∗	−0.218∗∗∗	−0.205∗∗∗	−0.216∗∗∗	−0.217∗∗∗	−0.216∗∗∗
DID-Matching	−0.308∗∗∗	−0.306∗∗∗	−0.277∗∗∗	−0.223∗∗∗	−0.232∗∗∗	−0.247∗∗∗
Treatment Window of Five Months						
All Channels						
DID	−0.171∗∗∗	−0.0943∗∗∗	−0.0824∗∗∗	−0.190∗∗∗	−0.140∗∗∗	−0.195∗∗∗
DID-Heckman	−0.169∗∗∗	−0.0927∗∗∗	−0.0794∗∗∗	−0.189∗∗∗	−0.136∗∗∗	−0.196∗∗∗
DID-Matching	−0.146∗∗∗	−0.0837∗∗∗	−0.0949∗∗∗	−0.160∗∗∗	−0.131∗∗∗	−0.156∗∗∗
Inner Channels						
DID	−0.0584∗∗	−0.00029	0.024	−0.0741∗∗∗	0.00388	−0.0378
DID-Heckman	−0.0507∗	0.0105	0.0348	−0.0681∗∗	0.0117	−0.0365
DID-Matching	−0.00946	0.0443	0.0517∗	−0.0501∗	0.00789	−0.0111
Share of Inner Channels						
DID	0.0226∗∗∗	0.0164∗∗∗	0.0204∗∗∗	0.0135∗∗	0.0277∗∗∗	0.0141∗∗
DID-Heckman	0.0247∗∗∗	0.0197∗∗∗	0.0234∗∗∗	0.0155∗∗∗	0.0305∗∗∗	0.0143∗∗
DID-Matching	0.0392∗∗∗	0.0357∗∗∗	0.0341∗∗∗	0.0196∗∗∗	0.0326∗∗∗	0.0231∗∗∗
Outer Channels						
DID	−0.126∗∗∗	−0.109∗∗∗	−0.107∗∗∗	−0.125∗∗∗	−0.142∗∗∗	−0.121∗∗∗
DID-Heckman	−0.133∗∗∗	−0.118∗∗∗	−0.114∗∗∗	−0.133∗∗∗	−0.150∗∗∗	−0.124∗∗∗
DID-Matching	−0.177∗∗∗	−0.185∗∗∗	−0.147∗∗∗	−0.129∗∗∗	−0.171∗∗∗	−0.145∗∗∗

Clustered standard errors are used

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A6: Treatment Effects with Other Testing Methods

	(1) Cohort 1 vs 3	(2) Cohort 1 vs 4	(3) Cohort 1 vs 5	(4) Cohort 1 vs 6	(5) Cohort 2 vs 4	(6) Cohort 2 vs 5	(7) Cohort 2 vs 6	(8) Cohort 3 vs 5	(9) Cohort 3 vs 6	(10) Cohort 4 vs 6
Alternative Measurements of Customer Purchase: Number of Room Nights										
DID	−0.0565***	−0.0784***	−0.0807***	−0.0839***	−0.0341***	−0.0244***	−0.0162***	−0.0194***	−0.0105*	−0.0738***
DID-Heckman	−0.0573***	−0.0773***	−0.0784***	−0.0825***	−0.0331***	−0.0240***	−0.0156**	−0.0197***	−0.0106*	−0.0741***
DID-Matching	−0.0204***	−0.0345***	−0.0346***	−0.0376***	−0.0321***	−0.0297***	−0.0193***	−0.0181***	−0.0102	−0.0486***
Alternative Measurements of Customer Purchase: Number of Bookings										
DID	−0.00185*	−0.00207**	$4.31E-05$	−0.000387	−0.00323***	−0.00139	−0.00159*	−0.00359***	−0.00301***	−0.00626***
DID-Heckman	−0.00187*	−0.00203*	0.000197	−0.00027	−0.00316***	−0.00128	−0.00151	−0.00347***	−0.00286***	−0.00623***
DID-Matching	−0.00200*	−0.00168	−0.000167	−0.00108	−0.00234**	−0.000481	−0.00109	−0.00260**	−0.00156	−0.00484***
Top 50 Percent Active App Users										
DID	0.000183	−0.00773	0.0627*	0.0304	−0.198***	−0.153***	−0.0877**	−0.184***	−0.122***	−0.264***
DID-Heckman	0.00118	−0.00569	0.0665*	0.0354	−0.195***	−0.150***	−0.0760*	−0.186***	−0.109***	−0.260***
DID-Matching	−0.00768	0.0208	0.0111	−0.000735	−0.147***	−0.144***	−0.0466	−0.192***	−0.113**	−0.217***
App Adopters without Spending Spike prior to Adoption										
DID	−0.0958***	−0.0796***	−0.0106	−0.0501*	−0.219***	−0.121***	−0.0782***	−0.167***	−0.124***	−0.279***
DID-Heckman	−0.0959***	−0.0793***	−0.0108	−0.0498*	−0.218***	−0.122***	−0.0804***	−0.167***	−0.122***	−0.279***
DID-Matching	−0.118***	−0.0761**	−0.0502	−0.110***	−0.196***	−0.124***	−0.128***	−0.139***	−0.0835***	−0.267***
Alternative Model Specification										
Panel DID	−0.0447***	−0.0496***	−0.0554***	−0.0548***	−0.0114	0.00518	0.0068	0.00282	0.00396	−0.0376***
Alternative Random Samples										
DID	−0.0527*	−0.110***	−0.121***	−0.0849***	−0.0853***	−0.0639*	−0.0529	−0.153***	−0.174***	−0.133***
DID-Heckman	−0.0554*	−0.111***	−0.122***	−0.0880**	−0.0857**	−0.0641*	−0.0523	−0.152***	−0.171***	−0.138***
DID-Matching	−0.0475	−0.0882**	−0.117***	−0.0858*	−0.102***	−0.0705*	−0.0759*	−0.176***	−0.215***	−0.108***
Alternative Treatment Cohorts										
DID	−0.214***	−0.232***	−0.217***	−0.188***	−0.299***	−0.226***	−0.237***	−0.299***	−0.233***	−0.339***
DID-Heckman	−0.216***	−0.230***	−0.213***	−0.185***	−0.296***	−0.222***	−0.234***	−0.298***	−0.230***	−0.338***
DID-Matching	−0.148***	−0.207***	−0.197***	−0.192***	−0.241***	−0.192***	−0.219***	−0.276***	−0.216***	−0.300***

Clustered standard errors are used

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Survey Questions

1. Are you above 18 years old?

   Options: Yes; No

2. Do you use a smart phone?

   Options: Yes; No

3. Do you have any hotel app (e.g., Hilton, Marriott and Starwood) in your phone?

   Options: Yes; No

4. What type of smart phone do you primarily use? (please check one)

   Options: iPhone; Android; Other (please specify)

5. Which hotel apps do you have in your phone? (please check all that apply)

   Options: Hilton; Marriott; Starwood; Hyatt; IHG; Wyndham; Choice; Best Western; La Quinta; Other (please specify)

6. Among the hotel apps in your phone, which one did you adopt/download MOST RECENTLY?

   Options: Hilton; Marriott; Starwood; Hyatt; IHG; Wyndham; Choice; Best Western; La Quinta; Other (please specify)

7. Among the hotel apps in your phone, which one did you adopt/download THE EARLIEST?

   Options: Hilton; Marriott; Starwood; Hyatt; IHG; Wyndham; Choice; Best Western; La Quinta; Other (please specify)

8. Which travel website apps do you have in your phone? (please check all that apply)

   Options: Expedia; Booking.com; Kayak; Orbitz; Priceline; CheapInn; Travelocity; Hotels.com; Hotwire; TripAdvisor; CheapTickets; HotelTonight; Trivago; Agoda; Other (please specify); None

9. Among the travel website apps in your phone, which one did you adopt/download MOST RECENTLY?
   Options: Expedia; Booking.com; Kayak; Orbitz; Priceline; CheapInn; Travelocity; Hotels.com; Hotwire; TripAdvisor; CheapTickets; HotelTonight; Trivago; Agoda; Other (please specify); None

10. Among the travel website apps in your phone, which one did you adopt/download THE EARLIEST?
    Options: Expedia; Booking.com; Kayak; Orbitz; Priceline; CheapInn; Travelocity; Hotels.com; Hotwire; TripAdvisor; CheapTickets; HotelTonight; Trivago; Agoda; Other (please specify); None

Questions 11 to 15 are about the mobile app by hotel X. If the respondent adopted the focal app (according to question 5), hotel X is the focal hotel. If the respondent did not adopt the focal app, hotel X is the first hotel app she or he adopted (according to question 7).

11. When did you adopt hotel X's app?
    Options: 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017

12. How often do you use the following functions in hotel X's app?
    Functions: Search for hotels; Make reservations; Manage reservations; Manage rewards account; Call customer service; Mobile check-in/out; Mobile service request
    Options: Never; Very rarely; Rarely; Occasionally; Frequently; Very Frequently

13. How do you agree with the following statements about hotel X's app's quality?
    Statements: Hotel X's app never fails; Hotel X's is fast; It is easy to find the information I am looking for in hotel X's app; If you are paying attention, please

select Strongly Disagree; The design of hotel X's app is attractive; Overall, I am very satisfied with hotel X's app quality

Options: Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree

14. How is the quality of hotel X's app compared with other hotel and travel website apps you have?

    Options: Among the best; Good; Average; Bad; Among the worst

15. What is your level of hotel X membership?

    Options: the name of each membership level varies across hotels.

16. How many nights do you stay at hotel in a year?

    Options: 0; 1-3; 4-9; 10-15; 16-30; 31-60; 61-90; 91 and more

17. In terms of the total number of nights at hotel in a year, what percentage do you spend for business trip?

    Option: a slider between 0 and 100.

18. Are you a member of the following hotels' loyalty program? (please check all that apply)

    Options: Hilton; Marriott; Starwood; Hyatt; IHG; Wyndham; Choice; Best Western; La Quinta; Other (please specify)

19. Are you male or female?

    Options: Female; Male

20. How old are you?

    Options: 18-24; 25-34; 35-44; 45-54; 55-64; 65 and above

DID Analysis of Customer App Engagement

We sketch out in this appendix the methodology we follow to obtain an unbiased estimate of the impact of using mobile check-in feature in the app on customers' spending. We focus below on customers who have adopted the mobile app and use its search feature.

Let N_{treat} be the number of customers of who self-select themselves to use the mobile check-in feature (the treatment group), N_{control1} be the number of customers who have the option to use the mobile check-in feature but do not use it (the control group 1), and N_{control2} be the number of customers who did not have the option to use the mobile check-in feature (the control group 2). A point estimate for the proportion of customers who self-select themselves to use the mobile check-in feature would be

$$p = \frac{N_{\text{treat}}}{N_{\text{treat}} + N_{\text{control1}}} \tag{3.9}$$

We report the number of customers in each group and the point estimate of "p" in Table A7.

Table A7: The Number of Customers in Each Group

Value	Among Customers Who Made Bookings After App Adoption	Among Customers Who Made Mobile Bookings After App Adoption
N_{treat}	10,099	5,196
N_{control1}	13,389	4,968
N_{control2}	8,718	2,553
p	0.430	0.511

If we perform a difference-in-differences (DID) analysis using the treatment group and the control group 1, then the DID estimate will capture the impact of the treatment group customers using the mobile check-in feature on their spending plus the self-selection bias; that is

$$\beta_1^{DID} = \beta_{mc} + \Delta_{ss} \tag{3.10}$$

We need to isolate the self-selection bias Δ_{ss} to get the unbiased estimate of mobile check-in.

Customers in the control group 2 did not have the option to use the mobile check-in feature. These customers consist of those who would have self-selected themselves had they the option to mobile check-in and those who would not have used the mobile check-in feature anyway. While we do not know the identities of each of these two groups of customers, the point estimate "p" from above can be used to estimate the proportion of those who would have self-selected to use the mobile check-in had it been available.

If we perform a DID analysis using the treatment group and the control group 2, then DID estimate is

$$\beta_2^{DID} = \beta_{mc} + \Delta_{ss} - p \cdot \Delta_{ss} \tag{3.11}$$

As a result, the difference between β_1^{DID} and β_2^{DID} equals $p \cdot \Delta_{ss}$, and thus we can isolate the effect of self-selection bias and get the unbiased estimate of β_{mc}:

$$\beta_{mc} = \beta_1^{DID} - (\beta_1^{DID} - \beta_2^{DID})/p \tag{3.12}$$

In practice, we use a boot-strapping methodology to obtain the estimate β_2^{DID}. We randomly draw 85 percent of customers in the control group 2 and run the DID analysis for 200 times. Thus, we obtain a distribution for the estimate β_{mc}. We have implicitly assumed that the decision to introduce mobile check-in by a hotel is random. We checked for this assumption in our analysis and found them to be valid.

We ran a logistic regression model to explain a hotel property's decision to adopt the mobile check-in feature as a function of the characteristics of the property such as franchise, airport area, downtown area, suburban area, and expressway area. None of these characteristics were significant in explaining the decision to adopt mobile check-in feature.

Table A8: The Impact of App Engagement on Customer Spending

	Among Customers Who Made Bookings After App Adoption		Among Customers Who Made Mobile Bookings After App Adoption	
	Treatment Group vs. Control Group 1	Treatment Group vs. Control Group 2	Treatment Group vs. Control Group 1	Treatment Group vs. Control Group 2
Variables	**(1)**	**(2)**	**(3)**	**(4)**
Check-In×Post	0.0138**	0.0272***	0.0217**	0.0351***
	(0.00692)	[0.000794]	(0.00963)	[0.00132]
Check-In	0.0565***	0.186***	0.0477***	0.179***
	(0.00605)	[0.00168]	(0.00885)	[0.00314]
Constant	0.688***	0.433***	0.727***	0.534***
	(0.0219)	[0.00454]	(0.0315)	[0.00788]
Reservation Characteristics	Yes	Yes	Yes	Yes
Customer Demographics	Yes	Yes	Yes	Yes
Month effect	Yes	Yes	Yes	Yes
Cohort effect	Yes	Yes	Yes	Yes
Number of Customers	23,488	18,817	10,164	7,749
Observations	589,982	448,890	283,897	208,935

Standard errors in parentheses
Standard deviations in brakets
*** p<0.01, ** p<0.05, * p<0.1

Table A8 reports the estimation results of the DID model of the treatment group and the control group 1 in Column 1 and 3. For the DID model of the treatment group and the control group 2, the table reports the mean value and the standard deviation of the 200 estimates in Column 2 and 4. Column 1 and 2 are results for customers who made bookings through any channel after app adoption, and Column 3 and 4 are results for customers who made bookings through the mobile channel after app adoption.

Appendix B: Chapter 2

Notations

Variables

- $d_i = \left\{ d^a_{it}, \left\{ d^p_{ijt} \right\}_{j=0}^{2}, \forall t \right\}$: A vector of observed app adoption and channel choice decisions for consumer i
- $d = \left\{ d^a_{it}, \left\{ d^p_{ijt} \right\}_{j=0}^{2}, \forall i, t \right\}$: A vector of observed app adoption and channel choice decisions

Parameters

- θ^r_h: Draw of hyperparameters (population mean) in iteration r
- $\sigma^r_{\theta_h}$: Draw of hyperparameters (population standard deviation) in iteration r
- θ^{*r}_i: Candidate parameter value specific to consumer i in iteration r
- θ^r_i: Accepted parameter value specific to consumer i in iteration r
- θ^{*r}_c: Candidate parameter value common across consumers in iteration r
- θ^r_c: Accepted parameter value common across consumers in iteration r

Equations

- $\widetilde{w}^r\left(x^r; \theta^{*r}_c, \theta^{*r}_i\right)$: Consumer i's pseudo-choice specific value function for choice d in iteration r
- $\widetilde{W}^r\left(x^r; \theta^{*r}_c, \theta^{*r}_i\right) = E_\epsilon \max_d \left\{ \widetilde{w}^r\left(x^r; \theta^{*r}_c, \theta^{*r}_i\right) + \epsilon \right\}$: Consumer i's pseudo-E_ϵ max function in iteration r
- $H^r = \left\{ \theta^{*l}_c, \left\{ \theta^{*r}_i, \widetilde{W}^l\left(x^l; \theta^{*l}_c, \theta^{*r}_i\right) \right\}_{i=1}^{I} \right\}_{l=r-N}^{r-1}$: Set of past pseudo-$E_\epsilon$ max function used for approximating the expected future value functions in iteration r
- $\widetilde{E}^r_{x'} W\left(x'; \theta^{*r}_c, \theta^{*r}_i\right)$: Pseudo-expected value function for consumer i in iteration r

$$\widetilde{E}^r_{x'} W\left(x'; \theta^{*r}_c, \theta^{*r}_i\right) = \sum_{l=r-N}^{r-1} \widetilde{W}^l\left(x^l; \theta^{*l}_c, \theta^{*r}_i\right) \frac{K_h(\theta^{*l}_c, \theta^{*r}_c) K_h(\theta^{*l}_i, \theta^{*r}_i)}{\sum_{k=r-N}^{r-1} K_h(\theta^{*k}_c, \theta^{*r}_c) K_h(\theta^{*k}_i, \theta^{*r}_i)}$$

where $K_h(\cdot)$ is the Gaussian kernel density with bandwidth h.

- $\widetilde{L}_i^r\left(d_i|x_i;\theta_c^{*r},\theta_i^{*r}\right)$: Pseudo-likelihood for consumer i conditional on H^r, x_i and $(\theta_c^{*r},\theta_i^{*r})$
- $\widetilde{L}^r\left(d|x;\theta_c^{*r},\{\theta_i^{*r}\}_{i=1}^I\right)$: Joint pseudo-likelihood conditional on H^r, x and $(\theta_c^{*r},\{\theta_i^{*r}\}_{i=1}^I)$

IJC Algorithm

Step I: Read data

Step II: Set starting parameter value, $\theta_h^0 = \left(c^0, \sigma_c^0, b^0, \sigma_b^0\right)$ and $\theta_c^0 = \left(c_0^0, \left\{\alpha_j^0, \lambda_j^0\right\}_{j=1}^2, \nu^0, \beta^0\right)$

Step III: Draw $\theta_i^0 \sim N\left(\theta^0, (\sigma_\theta^0)^2\right)$, $\forall i$

Step IV: Set $\widetilde{E}_{x'}^0 W\left(x'; \theta_c^0, \theta_i^0\right) = 0$, $\forall i$ and compute $\widetilde{L}_i^0\left(d_i | x_i; \theta_c^0, \theta_i^0\right)$, $\forall i$

Step V: Outer loop - Start MCMC Iterations and set $r = 1$

Step 1: Draw θ_h^r from the posterior distribution (normal):

$$\theta_h^r \sim N\left(\frac{1}{I}\sum_{i=1}^{I}\theta_i^{r-1}, \left(\frac{\sigma_{\theta_h}^{r-1}}{\sqrt{I}}\right)^2\right)$$

Step 2: Draw $\sigma_{\theta_h}^r$ from the posterior distribution (inverse-gamma):

$$\sigma_{\theta_h}^r \sim \left[IG\left(\frac{1+I}{2}, \frac{1+\sum_{i=1}^{I}\left(\theta_i^{r-1} - \theta_h^r\right)^2}{2}\right)\right]^{0.5}$$ with shape and rate parameters

Step 3: Consumer loop - Draw θ_i^r and set $i = 1$

Step i: Draw a candidate parameter vector, θ_i^{*r}

Step ii: Inner Loop - Compute $\widetilde{E}_{x_i'}^r W\left(x_i'; \theta_c^{r-1}, \theta_i^{*r}\right)$ and $\widetilde{E}_{x_i'}^r W\left(x_i'; \theta_c^{r-1}, \theta_i^{r-1}\right)$

Step iii: Compute $\widetilde{L}_i^r\left(d_i | x_i; \theta_c^{r-1}, \theta_i^{*r}\right)$ and $\widetilde{L}_i^r\left(d_i | x_i; \theta_c^{r-1}, \theta_i^{r-1}\right)$

Step iv: Compute the acceptance probability, λ

Step v: Draw $u \sim U[0,1]$. If $u < \lambda$, set $\theta_i^r = \theta_i^{*r}$; if $u \geqslant \lambda$, set $\theta_i^r = \theta_i^{r-1}$

Step 4: Draw a candidate parameter vector, θ_c^{*r}

Step 5: Inner Loop - Compute $\widetilde{E}_{x'}^r W\left(x'; \theta_c^{*r}, \theta_i^r\right)$ and $\widetilde{E}_{x'}^r W\left(x'; \theta_c^{r-1}, \theta_i^r\right)$, $\forall i$

Step 6: Compute $\widetilde{L}^r\left(d | x; \theta_c^{*r}, \{\theta_i^r\}_{i=1}^I\right)$ and $\widetilde{L}^r\left(d | x; \theta_c^{r-1}, \{\theta_i^r\}_{i=1}^I\right)$.

Compute and store $\widetilde{W}^r\left(x; \theta_c^{*r}, \theta_i^r\right)$, $\forall i$.

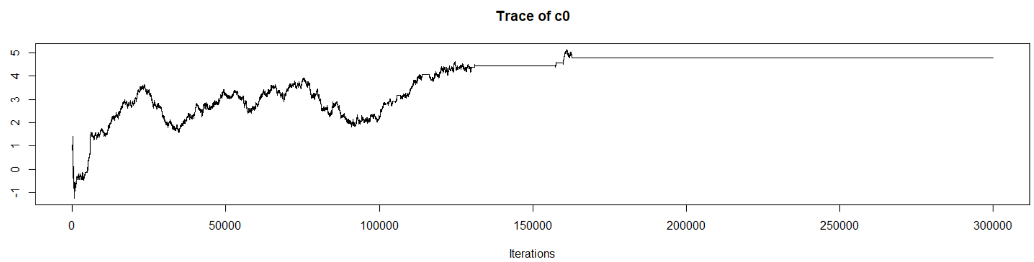
Step 7: Compute the acceptance probability, λ

Step 8: Draw $u \sim U[0,1]$. If $u < \lambda$, set $\theta_c^r = \theta_c^{*r}$; if $u \geqslant \lambda$, set $\theta_c^r = \theta_c^{r-1}$

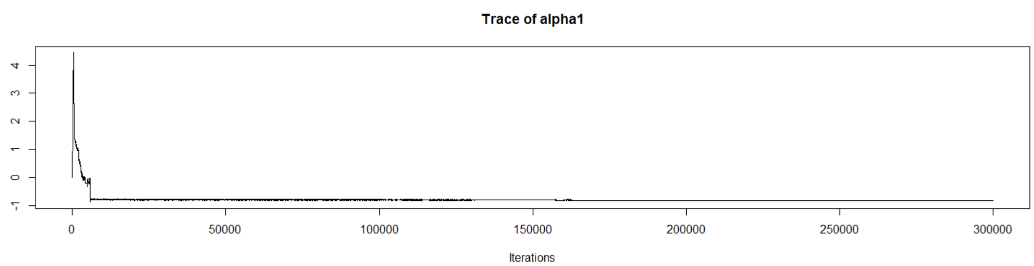
Step VI: End if r reaches the total number of MCMC iterations

Trace Plots

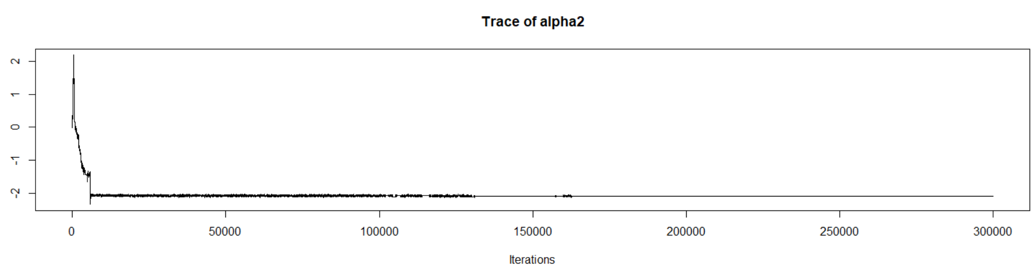
Figure B1: The Trace Plots of the Common Parameters



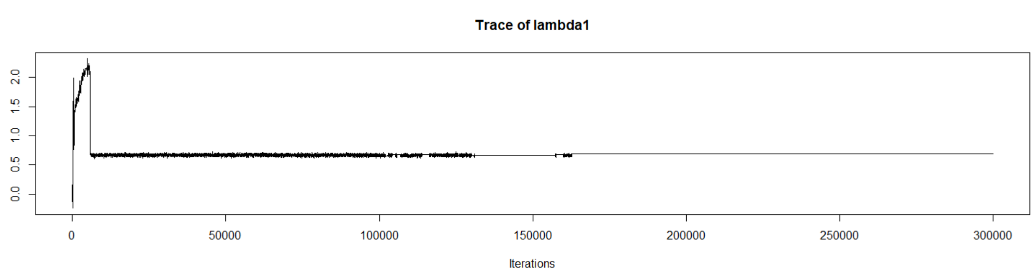
(a)



(b)

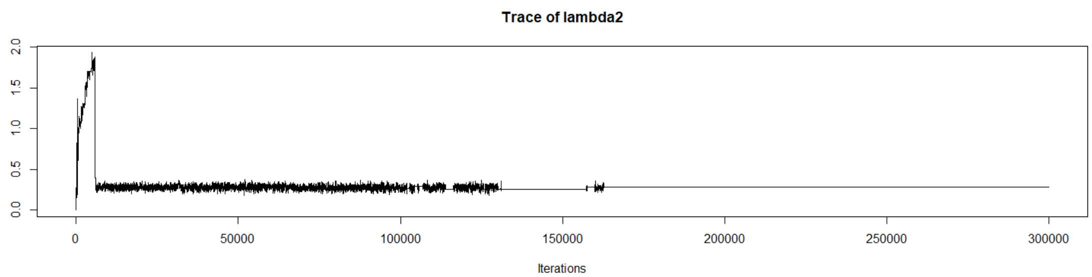


(c)

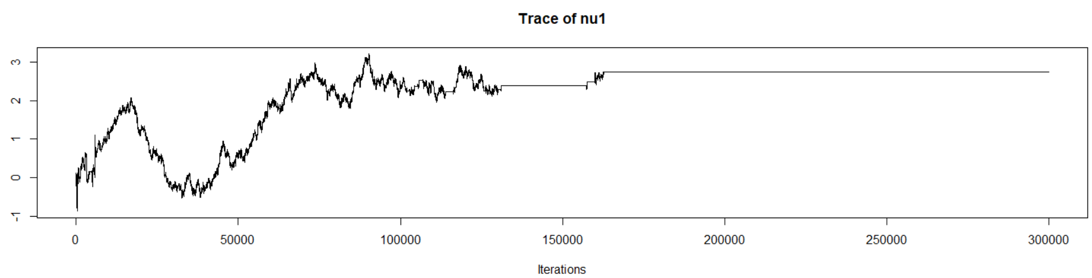


(d)

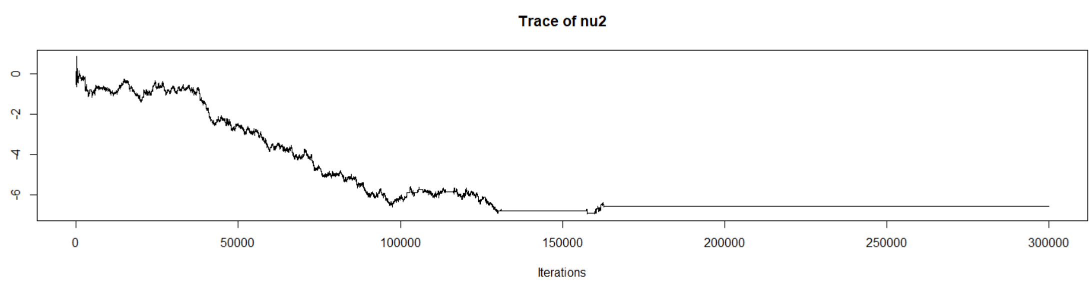
Figure B1: The Trace Plots of the Common Parameters (continued)



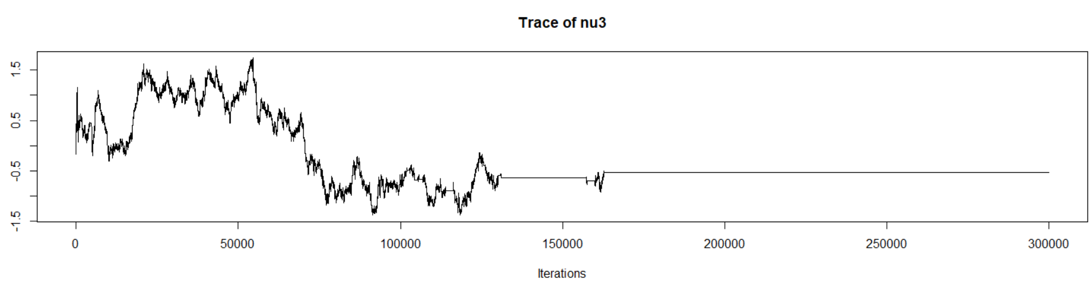
(e)



(f)

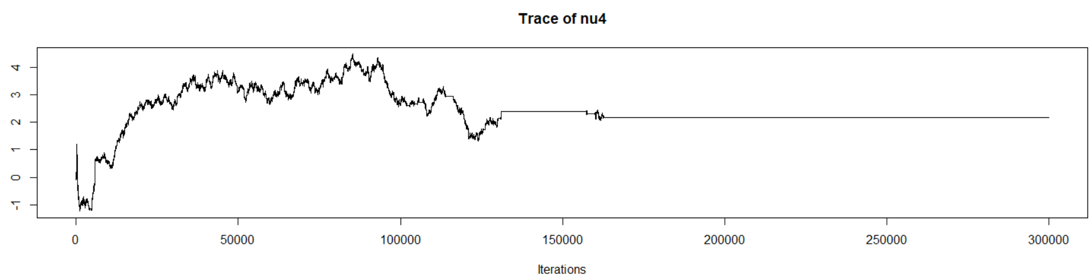


(g)

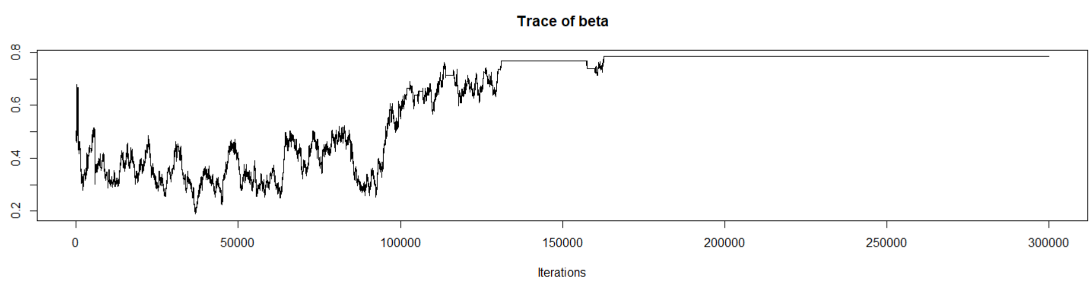


(h)

Figure B1: The Trace Plots of the Common Parameters (continued)



(i)



(j)

Appendix C: Chapter 3

Table C1: The Number of Titles in Each Cell and Experimental Condition

	Cell									
	1	2	3	4	5	6	7	8	9	Total
Control (C)	95	70	14	74	191	88	17	89	78	716
Hardcover (TH)	87	69	12	72	183	81	12	80	72	668
Ebook (TE)	87	69	12	72	183	81	11	81	71	667

Table C2: ANOVA: Comparing Mean Values across Conditions

Variables	**F-value**	**p-value**
Free PDF Quantity	2.513	0.081
Paperback Quantity	1.042	0.353
Paperback Revenue ($)	1.343	0.261

Table C3: The Distribution of Positive Paperback Sales Quantity

	Titles	Mean	Std.	Min	20%	40%	60%	80%	90%	Max
All Titles (before experiment)										
Control Condition	127	5	14	1	1	1	2	4	7	106
Hardcover Condition	112	9	41	1	1	1	2	3	10	406
Ebook Condition	122	16	101	1	1	1	2	3	10	1,035
All Titles (during experiment)										
Control Condition	133	3	5	1	1	1	2	4	6	41
Hardcover Condition	117	8	41	1	1	1	1	5	9	435
Ebook Condition	125	15	95	1	1	1	1	3	4	837
Blockbuster Titles Excluded (before experiment)										
Control Condition	127	5	14	1	1	1	2	4	7	106
Hardcover Condition	111	6	16	1	1	1	2	3	9	116
Ebook Condition	121	8	40	1	1	1	2	3	9	434
Blockbuster Titles Excluded (during experiment)										
Control Condition	133	3	5	1	1	1	2	4	6	41
Hardcover Condition	116	4	10	1	1	1	1	4	8	73
Ebook Condition	124	8	60	1	1	1	1	3	4	668

Table C4: The Effects of Introducing Hardcover or Ebook Version on Paperback Sales Quantity (Blockbuster Titles Excluded)

	Hardcover vs. Control		Ebook vs. Control	
	(1)	(2)	(3)	(4)
Variables	Coeff.	MEM.	Coeff.	MEM.
Treat	−.547	−.307	−.854**	−.436*
	(.379)	(.244)	(.388)	(.256)
Post	−.676**	−.386*	−.701**	−.360*
	(.292)	(.219)	(.311)	(.205)
Treat × Post	.726*	.502	.700*	.430
	(.424)	(.396)	(.397)	(.329)
Email Campaigns	1.111***	.622***	2.019***	1.015***
	(.207)	(.165)	(.515)	(.363)
Copyright Year	.117***	.0657***	.0920**	.0462***
	(.0426)	(.0189)	(.0426)	(.0173)
Number of Pages	.294**	.165***	.335***	.169***
	(.119)	(.0501)	(.106)	(.0394)
Constant	−1.619**	–	−1.508**	–
	(.794)	–	(.764)	–
Observations	2,766	2,766	2,764	2,764
Log-Likelihood	−2186	–	−2192	–
AIC	4388	–	4401	–
BIC	4435	–	4448	–
Pseudo R^2	.0328	–	.0709	–

MEM: marginal effect at means

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C5: The Effects of Introducing Hardcover or Ebook Version on Paperback Sales Quantity (90% Random Samples)

	Hardcover vs. Control		Ebook vs. Control	
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Coeff.**	**MEM.**	**Coeff.**	**MEM.**
Treat	−.500	−.334	−.498	−.385
	(.192)	(.128)	(.178)	(.133)
Post	−.715	−.488	−.549	−.430
	(.218)	(.154)	(.173)	(.137)
Treat×Post	1.197	1.188	1.713	2.570
	(.299)	(.392)	(.514)	(.959)
Email Campaigns	1.643	1.080	2.875	2.217
	(.227)	(.169)	(.408)	(.385)
Copyright Year	.136	.089	.110	.085
	(.0189)	(.0101)	(.0177)	(.0166)
Number of Pages	.352	.230	.097	.069
	(.0520)	(.0249)	(.0707)	(.0341)
Constant	−1.854	–	−1.341	–
	(.377)	–	(.292)	–
Observations	2,491	2,491	2,490	2,490

MEM: marginal effect at means

Standard deviations in parentheses

Table C6: The Effects of Introducing Hardcover or Ebook Version on Paperback Sales Quantity (Bootstrap Samples)

	Hardcover vs. Control		Ebook vs. Control	
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Coeff.**	**MEM.**	**Coeff.**	**MEM.**
Treat	−.511	−.338	−.483	−.383
	(.410)	(.321)	(.425)	(.341)
Post	−.734**	−.495	−.550*	−.441
	(.335)	(.316)	(.328)	(.303)
Treat×Post	1.207**	1.156	1.795**	2.582
	(.599)	(1.010)	(.825)	(2.752)
Email Campaigns	1.649***	1.089**	2.878***	2.282*
	(.497)	(.522)	(.826)	(1.302)
Copyright Year	.135**	.0893***	.111**	.0877*
	(.0534)	(.0300)	(.0506)	(.0478)
Number of Pages	.344**	.227**	.0719	.0570
	(.167)	(.0903)	(.167)	(.116)
Constant	−1.810*	–	−1.294	–
	(.993)	–	(.822)	–
Observations	2,768	2,768	2,766	2,766

MEM: marginal effect at means

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

DID Analyses on the Sub-Samples

Paperback Price. Using the same blocking design used in the previous section, we divide our sample of 2,051 titles into three sub-samples based upon price level of paperback version: low (price $\leqslant$ \$29), medium (\$29 $<$ price $\leqslant$ \$47.25) and high (price $>$ \$47.25). As presented in Column (1)$\sim$(3) of Table C7, extending the product line by adding a hardcover version has a significant positive impact on paperback sales quantity for the sub-sample of paperback titles with low price. The results in Column (4)$\sim$(6) of Table C7 show that introducing the ebook version increases the paperback sales of low priced titles. However, adding a new version does not have a significant impact on book titles with medium prices and high prices.

Popularity. Next, we define sub-samples based on popularity, as measured by the count of number of page views of a title's catalog. In particular, popularity is defined as low if count $\leqslant 136$; as medium if $136 <$ count $\leqslant 953$; and as high if count > 953. As shown in Table C8, the coefficient estimates of Treat$\times$Post for highly popular titles are positive and statistically significant. On the contrary, introducing a hardcover or ebook version to titles that are not very popular does not have a significant effect.

Copyright Year. Additionally, the results in Table C9 suggest that adding a hardcover version to the product line increases paperback sales of book titles published between 2011 and 2015, but does not have a significant impact on titles published earlier than 2010. It could be that customers have a higher willingness to pay for newer titles. In contrast, introducing the ebook version has a positive impact on the sale of paperback titles published before 2010 but does not significantly affect titles published more recently.

Table C7: The DID Estimation Results for Sub-Sample based on Paperback Price

	Hardcover vs. Control			Ebook vs. Control		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Variables	**Low Price**	**Medium Price**	**High Price**	**Low Price**	**Medium Price**	**High Price**
Treat	−3.896∗∗∗	.0218	.415	−3.110∗∗∗	−.246	−.111
	(.810)	(.363)	(.470)	(.654)	(.401)	(.325)
Post	−1.782∗∗∗	.0532	−.276	−1.738∗∗∗	−.0181	−.276
	(.328)	(.386)	(.237)	(.294)	(.371)	(.232)
Treat×Post	2.558∗∗∗	−.152	−.00515	5.729∗∗∗	−.0203	−.256
	(.526)	(.513)	(.512)	(1.057)	(.535)	(.362)
Email Campaigns	2.202∗∗∗	1.159∗∗∗	.634∗∗	6.423∗∗∗	1.443∗∗∗	.400∗∗∗
	(.552)	(.361)	(.318)	(1.084)	(.381)	(.122)
Copyright Year	.0761	.218∗∗∗	.138∗∗∗	.0878	.183∗∗∗	.171∗∗∗
	(.0884)	(.0418)	(.0528)	(.113)	(.0359)	(.0337)
Number of Pages	2.609∗∗∗	.868∗∗∗	.271∗∗∗	1.169∗	1.034∗∗∗	.341∗∗∗
	(.536)	(.219)	(.104)	(.597)	(.206)	(.0669)
Constant	−1.646∗	−4.015∗∗∗	−2.348∗∗∗	−1.000	−3.934∗∗∗	−2.852∗∗∗
	(.964)	(.660)	(.721)	(1.346)	(.561)	(.485)
Observations	2,768	2,768	2,768	2,766	2,766	2,766
Log-Likelihood	−2123	−2123	−2123	−2113	−2113	−2113
AIC	4294	4294	4294	4274	4274	4274
BIC	4436	4436	4436	4417	4417	4417

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C8: The DID Estimation Results for Sub-Sample based on Popularity

	Hardcover			Ebook		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Low	Medium	High	Low	Medium	High
Variables	Popularity	Popularity	Popularity	Popularity	Popularity	Popularity
Treat	$-.162$	$-.785**$	.406	$-.196$	.118	.190
	(.682)	(.340)	(.422)	(.657)	(.518)	(.495)
Post	-1.609	.326	$-.725***$	-1.609	.308	$-.634**$
	(1.096)	(.440)	(.241)	(1.096)	(.460)	(.290)
Treat×Post	.921	.0274	.970*	1.345	$-.889$	1.414***
	(1.397)	(.524)	(.512)	(1.335)	(.634)	(.442)
Email Campaigns	$-12.02***$	.295	.926***	–	.445*	1.603***
	(1.227)	(.370)	(.310)	–	(.269)	(.421)
Copyright Year	.0995**	.0696	$-.0376$	.147*	.0538	$-.108**$
	(.0470)	(.0469)	(.0441)	(.0758)	(.0463)	(.0536)
Number of Pages	.231	.00587	.0210	.569*	.194**	$-.0912$
	(.286)	(.0824)	(.0730)	(.333)	(.0780)	(.0910)
Constant	$-4.379***$	$-1.974***$	1.287*	$-5.084***$	$-2.199***$	2.174***
	(.595)	(.487)	(.688)	(.856)	(.472)	(.776)
Observations	2,768	2,768	2,768	2,766	2,766	2,766
Log-Likelihood	-1974	-1974	-1974	-851.3	-851.3	-851.3
AIC	3995	3995	3995	1747	1747	1747
BIC	4137	4137	4137	1877	1877	1877

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C9: The DID Estimation Results for Sub-Sample Based on Copyright Year

	Hardcover			Ebook		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Variables	**2000∼2004**	**2005∼2010**	**2011∼2015**	**2000∼2004**	**2005∼2010**	**2011∼2015**
Treat	−1.357∗∗	.793	−.204	−2.333∗∗∗	.646	−.928∗∗
	(.630)	(.519)	(.424)	(.659)	(.471)	(.418)
Post	−.699	.328	−.831∗∗	−.658	.273	−.989∗∗
	(.464)	(.247)	(.360)	(.458)	(.233)	(.403)
Treat×Post	.552	−.0953	1.327∗∗∗	2.005∗∗∗	2.019∗	.740
	(.528)	(.713)	(.460)	(.588)	(1.063)	(.495)
Email Campaigns	2.296∗∗∗	1.432∗∗∗	1.182∗∗∗	3.834∗∗∗	4.636∗∗∗	1.673∗∗∗
	(.641)	(.420)	(.337)	(.886)	(1.070)	(.566)
Copyright Year	−.0769	.0988	−.123	−.118	.189∗∗	.0717
	(.188)	(.0828)	(.189)	(.171)	(.0754)	(.157)
Number of Pages	.216	.662∗∗∗	.420∗∗	.180	.352∗∗∗	.446∗∗∗
	(.206)	(.120)	(.182)	(.197)	(.0927)	(.137)
Constant	−.452	−3.747∗∗∗	1.502	−.329	−3.781∗∗∗	−.945
	(1.163)	(.737)	(2.538)	(1.179)	(.633)	(2.341)
Observations	2,768	2,768	2,768	2,766	2,766	2,766
Log-Likelihood	−2165	−2165	−2165	−2208	−2208	−2208
AIC	4378	4378	4378	4464	4464	4464
BIC	4520	4520	4520	4607	4607	4607

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C10: Robustness Check: Binary Logit Model

Variables	Hardcover & Control Titles (1)	Ebook & Control Titles Spec. I (2)	Ebook & Control Titles Spec. II (3)	All Titles Spec. I (4)	All Titles Spec. II (5)
D^{hard}	.162**	–	–	.179**	.179**
	(.0752)	–	–	(.0736)	(.0736)
D^{ebook}	–	.229***	.0826	.168***	.0236
	–	(.0634)	(.0786)	(.0629)	(.0783)
$D^{high.ebook}$	–	–	.347***	–	.341***
	–	–	(.0973)	–	(.0972)
Price	−.0220***	−.0183***	−.0179***	−.0201***	−.0199***
	(.00249)	(.00207)	(.00207)	(.00180)	(.00180)
Constant	−5.015***	−5.234***	−5.249***	−5.106***	−5.116***
	(.114)	(.0935)	(.0941)	(.0826)	(.0829)
Observations	439,617	522,204	522,204	741,983	741,983
Log-Likelihood	−8084	−9645	−9638	−14153	−14147
AIC	16174	19295	19284	28314	28303
BIC	16207	19329	19329	28360	28361

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C11: Robustness Check: Nested Logit Model

	Free vs. Non-Free		Digital vs. Non-Digital	
	Spec. I	**Spec. II**	**Spec. I**	**Spec. II**
Variables	**(1)**	**(2)**	**(3)**	**(4)**
D^{ebook}	$.387{*}{*}{*}$	$.307{*}{*}{*}$	$.228{*}{*}{*}$	$.0819$
	$(.0611)$	$(.0855)$	$(.0634)$	$(.0786)$
$D^{high.ebook}$	$-$	$.157$	$-$	$.347{*}{*}{*}$
	$-$	$(.104)$	$-$	$(.0973)$
Price	$-.0181{*}{*}{*}$	$-.0184{*}{*}{*}$	$-.0183{*}{*}{*}$	$-.0179{*}{*}{*}$
	$(.00208)$	$(.00203)$	$(.00207)$	$(.00208)$
Constant (Ebook)	$-5.523{*}{*}{*}$	$-5.649{*}{*}{*}$	$-1.931{*}{*}{*}$	$-1.893{*}{*}{*}$
	$(.0804)$	$(.147)$	$(.485)$	$(.477)$
Constant (Paperback)	$-5.243{*}{*}{*}$	$-5.233{*}{*}{*}$	$-5.234{*}{*}{*}$	$-5.249{*}{*}{*}$
	$(.0939)$	$(.0920)$	$(.0935)$	$(.0941)$
inclusive value τ	$.271{*}{*}{*}$	$.349{*}{*}{*}$	$.326{*}{*}{*}$	$.319{*}{*}{*}$
	$(.0502)$	$(.0907)$	$(.0721)$	$(.0711)$
Observations	$1,181,517$	$1,181,517$	$1,181,517$	$1,181,517$
Log-Likelihood	-10554	-10552	-10559	-10552
AIC	21118	21118	21130	21118
BIC	21178	21202	21202	21202

Clustered standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Bibliography

Andrews, Michelle, Xueming Luo, Zheng Fang, and Anindya Ghose (2015). Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness. *Marketing Science 35*(2), 218–233.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ansari, Asim, Carl F Mela, and Scott A Neslin (2008). Customer Channel Migration. *Journal of Marketing Research 45*(1), 60–76.

Athey, Susan and Imbens, Guido W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica 74*(2), 431–497.

Bart, Yakov, Andrew T Stephen, and Miklos Sarvary (2014). Which Products are Best Suited to Mobile Advertising? A Field Study of Mobile Display Advertising Effects on Consumer Attitudes and Intentions. *Journal of Marketing Research 51*(3), 270–285.

Bayus, Barry L. and Jr. William P. Putsis (1999). Product Proliferation: An Empirical Analysis of Product Line Determinants and Market Outcomes. *Marketing Science 18*(2), 137–153.

Berger, Jonah, Michaela Draganska, and Itamar Simonson (2007). The Influence of Product Variety on Brand Perception and Choice. *Marketing Science 26*(4), 460–472.

Boyd, D. Eric, Rebecca Slotegraaf, and P. K. Kannan (2017). Marketing Technology and Accountability: Understanding the Impact of Branded Apps on Firm Value. *Working Paper*, James Madison University.

Burtch, Gordon and Yili Hong (2014). What Happens When Word of Mouth Goes Mobile? *Working Paper*, University of Minnesota.

Caldieraro, Fabio, Ling-Jing Kao, and Jr. Marcus Cunha (2015). Harmful Upward Line Extensions: Can the Launch of Premium Products Result in Competitive Disadvantages? *Journal of Marketing 79*(6), 50–70.

Chevalier, Judith A and Dina Mayzlin (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research 43*(3), 345–354.

Ching, Andrew T, Susumu Imai, Masakazu Ishihara, and Neelam Jain (2012). A Practitioner's Guide to Bayesian Estimation of Discrete Choice Dynamic Programming Models. *Quantitative Marketing and Economics 10*(2), 151–196.

Court, David, Dave Elzinga, Bo Finneman, and Jesko Perrey (2017). The New Battleground for Marketing-Led Growth. *McKinsey Quarterly*.

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*(5), 710–725.

De los Santos, Babur and Matthijs R. Wildenbeest (2017). E-book Pricing and Vertical Restraints. *Quantitative Marketing and Economics 15*(2), 85–122.

Desai, Preyas S. (2001). Quality Segmentation in Spatial Markets: When Does Cannibalization Affect Product Line Design? *Marketing Science 20*(3), 265–283.

Dinner, Isaac M., Harald J. Van Heerde, and Scott Neslin (2018). Too Far to Walk: Using a Retailer Mobile App to Engage Distant Customers. *Working Paper*, The University of North Carolina at Chapel Hill.

Draganska, Michaela and Dipak C. Jain (2005). Product-Line Length as a Competitive Tool. *Journal of Economics & Management Strategy 14*(1), 1–28.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite (2003). Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy 111*(3), 555–588.

Dropbox (February 23, 2018). Form S-1 Registration Statement.

eMarketer (2016, June). Shoppers Are Downloading More Mobile Retail Apps. eMarketer.

Erdem, Tülin 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 M, Zheng Fang, and Xueming Luo (2015). Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions. *Journal of Marketing Research 52*(5), 726–735.

Frederick, Shane, Leonard Lee, and Ernest Baskin (2014). The Limits of Attraction. *Journal of Marketing Research 51*(4), 487–507.

Ghose, Anindya (2017). *Tap: Unlocking the Mobile Economy*. MIT Press.

Ghose, Anindya and Sang Pil Han (2011). An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet. *Management Science 57*(9), 1671–1691.

Gill, Manpreet, Shrihari Sridhar, and Rajdeep Grewal (2017). Return on Engagement Initiatives (RoEI): A Study of a Business-to-Business Mobile App. *Journal of Marketing 81*(4), 44–66.

Goldfarb, Avi and Catherine E Tucker (2011). Privacy Regulation and Online Advertising. *Management Science 57*(1), 57–71.

Gönül, Füsun and Kannan Srinivasan (1996). Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model. *Marketing Science 15*(3), 262–279.

Gowrisankaran, Gautam and Marc Rysman (2012). Dynamics of Consumer Demand for New Durable Goods. *Journal of Political Economy 120*(6), 1173–1219.

Greene, William (2008). Functional Forms for the Negative Binomial Model for Count Data. *Economics Letters 99*(3), 585–590.

Halbheer, Daniel, Florian Stahl, Oded Koenigsberg, and Donald R. Lehmann (2014). Choosing a Digital Content Strategy: How Much Should be Free? *International Journal of Research in Marketing 31*(2), 192–206.

Heath, Timothy B., Devon DelVecchio, and Michael S. McCarthy (2011). The Asymmetric Effects of Extending Brands to Lower and Higher Quality. *Journal of Marketing 75*(4), 3–20.

Heckman, James J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

Hendel, Igal and Aviv Nevo (2006). Measuring the Implications of Sales and Consumer Inventory Behavior. *Econometrica 74*(6), 1637–1673.

Huang, Lei, Xianghua Lu, and Sulin Ba (2016). An Empirical Study of the Cross-Channel Effects Between Web and Mobile Shopping Channels. *Information & Management 53*(2), 265–278.

Huber, Joel, John W Payne, and Christopher Puto (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research 9*(1), 90–98.

Huber, Joel and Christopher Puto (1983). Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects. *Journal of Consumer Research 10*(1), 31–44.

Imai, Susumu, Neelam Jain, and Andrew Ching (2009). Bayesian Estimation of Dynamic Discrete Choice Models. *Econometrica 77*(6), 1865–1899.

Jumio (2015, August). Jumio Mobile Consumer Insights Study. Jumio.

Kannan, P. K., Barbara Kline Pope, and Sanjay Jain (2009). Pricing Digital Content Product Lines: A Model and Application for the National Academies Press. *Marketing Science 28*(4), 620–636.

Kanuri, Vamsi K., Murali K. Mantrala, and Esther Thorson (2017). Optimizing a Menu of Multiformat Subscription Plans for Ad-Supported Media Platforms. *Journal of Marketing 81*(2), 45–63.

Kekre, Sunder and Kannan Srinivasan (1990). Broader Product Line: A Necessity to Achieve Success? *Management Science 36*(10), 1216–1232.

Kim, Su Jung, Rebecca Jen-Hui Wang, and Edward C. Malthouse (2015). The Effects of Adopting and Using a Brand's Mobile Application on Customers' Subsequent Purchase Behavior. *Journal of Interactive Marketing 31*, 28–41.

Kivetz, Ran, Oded Netzer, and V. Srinivasan (2004). Alternative Models for Capturing the Compromise Effect. *Journal of Marketing Research 41*(3), 237–257.

Konuş, Umut, Scott A. Neslin, and Peter C. Verhoef (2014). The Effect of Search Channel Elimination on Purchase Incidence, Order Size and Channel Choice. *International Journal of Research in Marketing 31*(1), 49–64.

Konuş, Umut, Peter C. Verhoef, and Scott A. Neslin (2008). Multichannel Shopper Segments and Their Covariates. *Journal of Retailing 84*(4), 398–413.

Koukova, Nevena T., P. K. Kannan, and Amna Kirmani (2012). Multiformat Digital Products: How Design Attributes Interact with Usage Situations to Determine Choice. *Journal of Marketing Research 49*(1), 100–114.

Kumar, Vineet (2014). Making "Freemium" Work: Many Start-ups Fail to Recognize the Challenges of This Popular Business Model. *Harvard Business Review 92*(5), 27–29.

Kumar, V. and Denish Shah (2011). Practice Prize Paper—Uncovering Implicit Consumer Needs for Determining Explicit Product Positioning: Growing Prudential Annuities' Variable Annuity Sales. *Marketing Science 30*(4), 595–603.

Lambrecht, Anja and Kanishka Misra (2016). Fee or Free: When Should Firms Charge for Online Content? *Management Science 63*(4), 1150–1165.

Lee, Clarence, Vineet Kumar, and Sunil Gupta (2017). Designing Freemium: Strategic Balancing of Growth and Monetization. Working Paper, Cornell University.

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, Jing, Umut Konuş, Koen Pauwels, and Fred Langerak (2015). The Hare and the Tortoise: Do Earlier Adopters of Online Channels Purchase More? *Journal of Retailing 91*(2), 289–308.

Lipsman, Andrew and Adam Lella (2017, August). The 2017 U.S. Mobile App Report. comScore.

Luce, R. Ducan (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and Sons.

Luo, Xueming, Michelle Andrews, Zheng Fang, and Chee Wei Phang (2013). Mobile Targeting. *Management Science 60*(7), 1738–1756.

Montaguti, Elisa, Scott A. Neslin, and Sara Valentini (2015). Can Marketing Campaigns Induce Multichannel Buying and More Profitable Customers? A Field Experiment. *Marketing Science 35*(2), 201–217.

Moorthy, K. Sridhar (1984). Market Segmentation, Self-Selection, and Product Line Design. *Marketing Science 3*(4), 288–307.

Narang, Unnati and Venkatesh Shankar (2019). The Effects of Mobile Apps on Shopper Purchases and Product Returns. *Working Paper*, Texas A&M University.

Neslin, Scott A, Kinshuk Jerath, Anand Bodapati, Eric T Bradlow, John Deighton, Sonja Gensler, Leonard Lee, Elisa Montaguti, Rahul Telang, Raj Venkatesan, et al. (2014). The Interrelationships between Brand and Channel Choice. *Marketing Letters 25*(3), 319–330.

Niculescu, Marius F and Dong Jun Wu (2014). Economics of Free under Perpetual Licensing: Implications for the Software Industry. *Information Systems Research 25*(1), 173–199.

Parker, Jeffrey R. and Donald R. Lehmann (2011). When shelf-based scarcity impacts consumer preferences. *Journal of Retailing 87*(2), 142–155.

Pauwels, Koen and Allen Weiss (2008). Moving from Free to Fee: How Online Firms Market to Change Their Business Model Successfully. *Journal of Marketing 72*(3), 14–31.

Pires, Tiago (2016). Costly Search and Consideration Sets in Storable Goods Markets. *Quantitative Marketing and Economics 14*(3), 157–193.

Ratneshwar, Srinivasan, Allan D. Shocker, and David W. Stewart (1987). Toward Understanding the Attraction Effect: The Implications of Product Stimulus Meaningfulness and Familiarity. *Journal of Consumer Research 13*(4), 520–533.

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.

Rust, John (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Seiler, Stephan (2013). The Impact of Search Costs on Consumer Behavior: A Dynamic Approach. *Quantitative Marketing and Economics 11*(2), 155–203.

Shampanier, Kristina, Nina Mazar, and Dan Ariely (2007). Zero as a Special Price: The True Value of Free Products. *Marketing Science 26*(6), 742–757.

Shankar, Venkatesh, Alladi Venkatesh, Charles Hofacker, and Prasad Naik (2010). Mobile Marketing in the Retailing Environment: Current Insights and Future Research Avenues. *Journal of Interactive Marketing 24*(2), 111–120.

Simonson, Itamar (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research 26*(2), 158–174.

Simonson, Itamar and Aimee Drolet (2004). Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research 31*(3), 681–690.

Song, Inseong and Pradeep K Chintagunta (2003). A Micromodel of New Product Adoption with Heterogeneous and Forward-Looking Consumers: Application to the Digital Camera Category. *Quantitative Marketing and Economics 1*(4), 371–407.

Sridhar, Shrihari and Srinivasaraghavan Sriram (2015). Is Online Newspaper Advertising Cannibalizing Print Advertising? *Quantitative Marketing and Economics 13*(4), 283–318.

Trachtenberg, Jeffery A. (2010, September 26). Authors Feel Pinch in Age of E-Books. *The Wall Street Journal*.

Tversky, Amos (1972). Elimination by Aspects: A Theory of Choice. *Psychological Review 79*(4), 281.

Tversky, Amos and Daniel Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Venkatesan, Rajkumar, Vipin Kumar, and Nalini Ravishanker (2007). Multichannel Shopping: Causes and Consequences. *Journal of Marketing 71*(2), 114–132.

Verhoef, Peter C., Scott A. Neslin, and Bjorn Vroomen (2007, Jun). Multichannel Customer Management: Understanding the Research-Shopper Phenomenon. *International Journal of Research in Marketing 24*(2), 129–148.

Wang, Rebecca Jen-Hui, Edward C Malthouse, and Lakshman Krishnamurthi (2015). On the Go: How Mobile Shopping Affects Customer Purchase Behavior. *Journal of Retailing 91*(2), 217–234.