Carnegie Mellon Tepper School of Business

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
INDUSTRIAL ADMINISTRATION
(MARKETING)

Titled
"MARKETING SOLUTIONS ENABLED BY BIG DATA USING STRUCTURAL MODELS"

Presented by
Xiao Liu

Accepted by

Kannan Srinivasan	4/22/15
Co-Chair: Prof. Kannan Srinivasan	**Date**
Alan Montgomery	4/22/15
Co-Chair: Prof. Alan Montgomery	**Date**

Approved by The Dean

Robert M. Dammon	4/23/15
Dean Robert M. Dammon	**Date**

Marketing Solutions Enabled by Big Data Using Structural Models

By

Xiao Liu

A dissertation submitted to

Tepper School of Business, Carnegie Mellon University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Industrial Administration (Marketing)

Doctoral Committee:

Professor Kannan Srinivasan (Co-Chair)

Professor Alan Montgomery (Co-Chair)

Professor Baohong Sun

Professor Param Vir Singh

Professor Burton Hollifield

April 22 2015

Abstract

Digital marketing has brought in enormous capture of consumer data. In quantitative marketing, researchers have adopted structural models to explain the (dynamic) decision process and incentives of consumers. However, due to computational burdens, structural models have rarely been applied to big data. Machine learning algorithms that perform well on big data, however, hardly have any theoretical support. Equipped with both economics perspective and machine leaning techniques, in this dissertation, I aim to combine rigorous economic theory with machine learning methods and apply them to data on a large scale.

First of all, I leverage structural models to understand the behaviors of consumers, firms and the entire market. In my first essay "An Empirical Analysis of Consumer Purchase Behavior of Base Products and Add-ons Given Compatibility Constraints", I model consumers as forward-looking utility maximizers who consider the product bundle of base products and add-ons jointly. I derive the add-on-to-base effect from the solution of the dynamic programming problem and then quantified it by model estimates. The underlying theory of switching cost caused by incompatibility constraints helps me rationalize the high market share of Sony's despite its high price. Without this theoretical foundation, I could not explain the data puzzle and might have fallen into false causality.

Doing empirical research in this time of an explosion of the quantity and quality of data is fortuitous. With the help of "Big Data", I am able to explore new areas of research, like social media. In my second essay "A Structured Analysis of Unstructured Big Data Leveraging Cloud Computing", I conduct analysis on a staggering volume of nearly two billion tweets to predict TV show ratings. Different from the traditional economics data that can be easily stored in a spreadsheet, the unstructured format and sheer volume of the tweet data challenges traditional information extraction and selection methods. But after careful sifting the data through a combination of methods from cloud computing, machine learning and text mining, the rich content information imbedded in the text provides us with much better explanatory and predictive power.

The beauty of structural models comes with a cost of computational burden as manifested in estimating dynamic choice models. The problem of curse of dimensionality is exacerbated when I estimate the model at the individual level on a large sample, as I do in the third essay "Overhaul Overdraft Fees: Creating Pricing and Product Design Strategies with Big Data". In this project I build a model where I assume consumers perform financial planning and spend their money rationally subject to monitoring cost and heavy discounting tendency. As consumers exhibit significantly different behavior patterns, I need to estimate the model at the individual level in order to design targeted strategies. But to do this with the standard estimation methods takes prohibitively large amount of time. To solve this problem I employ a new parallel computing algorithm to conduct parallel MCMC to speed up the estimation. This facilitates the marriage of applying the structural model on Big Data.

Acknowledgements

I would like to thank my co-advisors, Professor Kannan Srinivasan and Professor Alan Montgomery for their tremendous help, nurture and support. They set up an example for me to be passionate and rigorous with research. Every time after our conversation, I feel sobered with brilliant insights and inspired with full-hearted motivation. They strengthened my conviction to be a scholar to discover and contribute knowledge to the public.

I am sincerely obliged to Professor Baohong Sun. It was she who brought be to CMU and taught me how to combine consumer behavior theory with empirical modeling techniques. She urged me to read plenty of classical marketing papers, she supported me to go to the top-tier conferences since my first year in the PhD program and she started the first project with me. These experiences all helped me lay a solid foundation for the future research.

I also want to thank Professor Joachim Vosgerau who gave me a hand when I had a hard time. His integrity and generosity helped me survive the hardship and gain confidence to overcome difficulties and move forward.

I feel incredibly fortunate to have worked closely with Professor Param Vir Singh who challenged me to create works that are rich in managerial implications. My thanks also go to Professor Burton Hollifield who encouraged me to work in the area of marketing-finance interface.

I benefited enormously from a number of faculty members and alumni of CMU: Professor Guofang Huang, Professor Nitin Metha (University of Toronto), Professor Kevin Chung (University of Wisconsin-Madison), Professor Baojun Jiang (Washington University St. Louis), Professor Jeff Galak, Professor Chris Olivola, Professor Kinshuk Jerath (Columbia University), Professor Peter Boatwright, Professor Carey Morewedge, Professor Kaifu Zhang, Professor George Loewenstein, Professor Rahul Telang, Professor Beibei Li, Professor Vibhanshu Abhishek, Professor Vishal Singh (New York University), Professor Vineet Kumar (Yale University), Professor Timothy Derdenger, Professor Yan Huang (University of Michigan), Professor Yingda Lu (Rensselaer Polytechnic Institute), Professor Helen Yangfang Zhou (Singapore Management University), Professor Yaroslav Kryukov, Professor Joachim Ryohei Groeger, Professor Bo Zhou (University of Maryland), Professor Liye Ma (University of Maryland), Professor Young Eun Huh, Professor Amin Sayedi (University of Washington), Willie Neiswanger, Professor Juanjuan Zhang (MIT), Professor Carl Mela (Duke University), Professor Jack Xinlei Chen (CKGSB), Professor Jinhong Xie (University of Florida), Professor Wei Zhang (Iowa State University), Professor Yuhuang Zheng and many others.

I especially thank Professor Dipankar Chakravarti and Mrs. Sharmila Chakravarti for their generosity to provide the me with a fellowship to support by dissertation.

Thanks to Lawrence Rapp for the continuous help.

I am grateful to have many friends who made my Phd studies a colorful period of life. Special thanks go to Qianyun Xie, Zichao Yang, Ryan Tae Wook Kim, Ying Xu, Jiaxu Wang, Jing Gong, Yang Yang, Sijia Wang, Ronghuo Zheng, Ying Zhang and Miao Fan. I also thank Francisco Cisternas for being there all the time.

Most importantly, I'll give my deepest thanks to my parents who make me a better person every day. This dissertation is dedicated to my mother Mrs. Chunling Xiao and my father Mr. Yongchang Liu.

Contents

Introduction

Digital marketing has brought in enormous capture of consumer data. In quantitative marketing, researchers have adopted structural models to explain the (dynamic) decision process and incentives of consumers. However, due to computational burdens, structural models have rarely been applied to big data. Machine learning algorithms that perform well on big data, however, hardly have any theoretical support. Equipped with both economics perspective and machine leaning techniques, in this dissertation, I aim to combine rigorous economic theory with machine learning methods and apply them to data on a large scale.

First of all, I am deeply rooted in economic theory in understanding the behaviors of consumers, firms and the entire market. That's why I always have impetus to employ economic theories to explain the data at hand. In my first essay "An Empirical Analysis of Consumer Purchase Behavior of Base Products and Add-ons Given Compatibility Constraints", I model consumers as forward-looking utility maximizers who consider the product bundle of base products and add-ons jointly. I derive the add-on-to-base effect from the solution of the dynamic programming problem and then quantified it by model estimates. The underlying theory of switching cost caused by incompatibility constraints helps me rationalize the high market share of Sony's despite its high price. Without this theoretical foundation, I could not explain the data puzzle and might have fallen into false causality.

Doing empirical research in this time of an explosion of the quantity and quality of data is fortuitous. As two of my favorite economists, Liran Einav and Jonathan Levin have said, "Large-scale administrative datasets and proprietary private sector data can greatly improve the way we measure, track and describe economic activity." With the help of "Big Data", I am able to explore new areas of research, like social media. In my second essay "A Structured Analysis of Unstructured Big Data Leveraging Cloud Computing", I conduct analysis on a staggering volume of nearly two billion tweets to predict TV show ratings. Different from the traditional economics data that can be easily stored in a spreadsheet, the unstructured format and sheer volume of the tweet data challenges traditional information extraction and selection methods. But after careful sifting the data through a combination of methods from cloud computing, machine learning and text mining, the rich content information imbedded in the text provides us with much better explanatory and predictive power. This project has encouraged me to advocate for collecting and using Big Data to conduct marketing research. Armed with state-of-the-art data processing and analyzing techniques, I am anxious to tackle the complex real-world problems with Big Data to offer new insights on consumer behavior, firm strategies and public policy.

The beauty of structural models comes with a cost of computational burden as manifested in estimating dynamic choice models. When estimating the Base and Add-on model, I encounter the problem of curse of dimensionality because there were too many dimensions of the state variables. The problem is exacerbated when I tried to estimate the model at the individual level on a large sample, as I do in the third essay "Overhaul Overdraft Fees:

Creating Pricing and Product Design Strategies with Big Data". In this project I build a model where I assume consumers perform financial planning and spend their money rationally subject to monitoring cost and heavy discounting tendency. As consumers exhibit significantly different behavior patterns, I need to estimate the model at the individual level in order to design targeted strategies. But to do this with the standard estimation methods takes prohibitively large amount of time. To solve this problem I employ a new parallel computing algorithm to conduct parallel MCMC to speed up the estimation. This facilitates the marriage of applying the structural model on Big Data.

In summary, I have done research in areas spanning high-tech marketing, social media and consumer financial decision making using structural models and "Big Data". I hope these studies help researchers and marketers create better marketing solutions enabled by Big Data using structural models.

1 Chapter 1
An Empirical Analysis of Consumer Purchase Behavior of Base Products and Add-ons Given Compatibility Constraints

Xiao Liu, Timothy Derdenger, and Baohong Sun[1]

Abstract

Despite the common practice of multiple standards in the high-technology product industry, there is a lack of knowledge on how compatibility between base products and add-ons affects consumer purchase decisions at the brand and/or standard level. We recognize the existence of compatibility constraints and develop a dynamic model in which a consumer makes periodic purchase decisions on whether to adopt/replace a base and/or an add-on product. Dynamic and interactive inventory effects are included by allowing consumers to account for the long-term financial implications when planning to switch to a base product that is incompatible with their inventory of add-ons. Applying the model to the consumer purchase history of digital cameras and memory cards from 1998 to 2004, we demonstrate that the inventory of add-ons significantly affects the purchase of base products. This "lock-in" effect is enhanced when future prices of add-ons decrease. Interestingly, it is more costly for consumers to switch from Sony to other brands than vice versa. In four policy simulations, we explore the impact of alternative pricing and compatibility policies. For example, if Sony did not create its proprietary Memory Stick, the market share of its cameras would have been reduced by 6 percentage points.

Keywords: Compatibility and standard, base product, add-on product, dynamic structural model, product adoption, product line pricing

[1] Xiao Liu is a PhD student of Marketing at Tepper School of Business of Carnegie Mellon University. Email: xiaoliu@andrew.cmu.edu. Timothy Derdenger is an Assistant Professor of Marketing and Strategy of the same institute. Email: derdenge@andrew.cmu.edu. Baohong Sun is Dean's Distinguished Chair Professor of Marketing at Cheung Kong Graduate School of Business. Email: bhsun@ckgsb.edu.cn

1.1 Introduction

In high-tech markets, products often consist of a basic component, the base product (e.g. cameras), and some accessories, the add-ons (e.g. memory cards). The add-on product market is non-trivial. In 2006, the Consumer Electronics Association estimated that buyers spent an average of 15 percent of the cost of a primary consumer electronics device on compatible accessories.[2] In the automobile industry, the market size of add-ons is $216 billion and has been growing at an annual rate of 8 percent since 2000.[3] These statistics imply that manufacturers are increasingly relying on selling add-ons to raise profits and, more importantly, manufacturers are eager to explore how to design add-on products in order to boost product line total sales. One common strategy is to create a proprietary standard for the add-ons to "lock in" consumers to the base product.

The compatibility creates inter-dependence between the base products and add-ons, which makes consumers' purchase/upgrade/replacement decisions inter-connected both across time and across categories. For example, in the video game industry, many games are tied to only one type of console, e.g. Xbox or PlayStation. Game players accumulate many games over time. At the time when a console needs to be replaced, the game players usually prefer to stay with the same standard or brand of console because they want to continue playing their games (in inventory) and in this way they avoid re-purchasing all the games in inventory in order to achieve the same entertainment value offered by the old console and the games in possession. In a similar fashion, when making a smartphone replacement decision, an Android user might be hesitant to switch to an iPhone because of all the apps purchased. It is interesting and important to understand how the past ownership of add-ons affects consumer choices of brands when multiple standards exist.

There are many existing studies on cross-category purchases of durable goods at a category level. Yet, most of them do not recognize the existence of multiple standards that are regularly observed in practice. Moreover, among the few studies that focus on the interdependence of base and add-on products, many only recognize the simultaneous ownership of base and add-on products and not the potential impact past ownership of add-ons may have on future base product adoption. Hence, there is a need to understand the impact of proprietary standards or incompatibility between base and add-on products on consumers' purchase decisions of these products. In this paper, we evaluate the dynamic impact of add-on products on a consumer's choice of a base product by deconstructing the impact into the following key issues: First, how do the prices/qualities of add-ons affect consumer choice of a base product at the brand or standard level? Second, does the inventory of add-ons affect the purchase of compatible and non-compatible base products; in other words, does a compatibility requirement of add-on products create a cost of switching that locks consumers into the compatible base product? Third, how do consumer expectations about future price/quality trends of add-ons moderate the effect of inventory

---

[2] http://www.letsgodigital.org/en/13653/camera_accessories/

[3] http://www.autodealermonthly.com/64/2523/ARTICLE/The-Value-Of-The-Accessory-Market.aspx

of add-ons on the purchase of compatible base products? Fourth, what is the monetary equivalent for consumers to switch to a non-compatible brand? Finally, is it a profitable strategy to leverage the cost of switching created by incompatibility of add-ons with base products?

This paper provides a framework to explicitly model consumer brand and standard choices of base and add-on products and investigate the dynamic dependence between two product categories, when multiple standards exist. Distinguishing ourselves from prior literature, we model compatibility as the link between the base product and the add-ons which makes consumers' purchase decisions dynamically interdependent. Moreover, we allow forward-looking consumers to consider the inventory of their add-ons when determining the purchase of base and add-on products. Our dynamic structural model further characterizes two new inter-temporal trade-offs of consumers simultaneously: cross-category price effect and the cross-category dynamic inventory effect. Specifically, a consumer may sacrifice the current utility (lower price) of the base product for future benefit (lower price/higher quality) of the add-ons. Besides, a forward-looking consumer may sacrifice the gain from switching to a cheaper but incompatible base product in exchange for the stream of future consumption utilities from the inventory of add-ons by continuing with a compatible base product. We name the later effect the "Add-on-to-base effect", which captures the notion that the more add-ons a consumer has accumulated, the less the consumer is willing to switch to other incompatible base products because the forward-looking consumer wants to continue enjoying the benefit of the add-ons and avoid the cost of replenishing these inventory in the future.

We apply the model to a unique panel data with 828 households and their purchase history of digital camera and memory card from December 1998 to November 2004. During the six year observation period, manufacturers of digital cameras developed (at least) three standards to take advantage of the exclusivity benefit of compatibility: Memory Stick (Standard 1) for Sony, xD Card (Standard 2) for Olympus and Fujifilm and SD Card (Standard 3) for Canon, Kodak, Nikon and HP. The unique structure of this industry provides an ideal opportunity to examine brand competition in the face of standard compatibility constraints.

We find strong empirical evidence of an "add-on-to-base effect." Consumers are indeed locked in by the utility compatible add-ons provide. Interestingly, the cost to switch is asymmetric: it takes more for Standard 2 and Standard 3 to steal Sony consumers ($23.06 and $21.59) than for Sony to steal the consumers from the other two standards ($8.48 and $15.50). The structural model further permits us to investigate the interaction between the "cross-category price effect" and the "add-on-to-base effect." We show that the add-on-to-base effect is enhanced when future prices of add-ons are lower, (i.e. when the expected future price of a memory card decreases) the purchase probability of the compatible camera increases given the same amount of add-ons purchased before.

With the use of several counterfactual simulations we also discover that if manufacturers of Standard 2 cameras lower their camera prices during the initial two year period, they can benefit more from the dynamic "add-on-to-base effect" and increase camera sales by roughly 20%. Moreover, when incompatibility is removed among standards, the manufacturer of a premium memory card cannot reap the profit from camera transactions. For instance, if Sony did not create its proprietary memory card standard, the market share of its cameras would have been reduced by 6 percentage points. Our third finding determines that the sales of Standard 3 cameras increase significantly if they adopt an adapter that makes their cameras compatible with the memory cards of Sony's. And finally when a firm's brand equity (approximated by the intrinsic brand preference) falls below the industry average, incompatibility damages its market share.

We contribute theoretically and substantively to the literature on cross-category purchases of durable goods. Our model advances the literature by endogenizing purchase quantity of add-ons and allowing consumers to consider their inventory of the add-ons when making brand/standard choice decisions. Built at the standard level, our model is the first to allow forward-planning consumers to consider the enhanced consumption value that increases with the number of add-on products at possession and the financial cost of replacing them if she chose to switch to an incompatible base product., Substantively, our results reveal interesting brand level asymmetric competition patterns that can explain the puzzle that high price/low quality products receive high demand. Moreover, our policy simulations offer insights on how compatibility constraints affect consumer demand and firm pricing strategies.

1.2 Literature Review

Our paper stems from three streams of literature: durable goods adoption and replacement decision-making; multi-category purchase analysis; and compatibility. Recent years have seen an increase in research on empirical examination of durable goods adoption and replacement decision-making. This stream of research focuses on how consumers take the price and quality evolvement process into account to make long-term purchase decisions. For example, Melnikov (2013) constructs a dynamic model that describes consumers' adoption of differentiated durable products as an optimal stopping problem and finds evidence of forward-looking behavior. Song and Chintagunta (2003) and Gowrisankaran and Rysman (2012) further incorporate consumer heterogeneity. Nair (2007) studies the optimal pricing strategy for a firm to sell video-games to forward-looking consumers who strategically delay purchases to benefit from lower prices in the future. Gordon (2009) models both product adoption and replacement process. However, this stream of research focuses on a single product category and does not examine multi-category purchases.

There also have emerged a few papers investigating the complementarity relationship between products in different categories. Seetharaman et al. (2005) provides an excellent review of models of multi-category choice behavior, including three outcomes: purchase incidence, brand choice, and quantity consideration. Sriram, Chintagunta and Agarwal (2009), Gentzkow (2007) as well as Liu, Chintagunta and Zhu (2010) present a framework to measure the complementarity effect which is the additional per-period utility derived from owning products of both categories. Hartmann and Nair (2010) study how expectations about the future prices of the aftermarket goods influence initial purchase of the primary good. Though recognizing the complementary relationship between base products and their add-ons, these models use a time-invariant constant term to capture the relationship, not fully capturing the dynamic (especially inventory) impact of one category on the other. In contrast, our paper not only relaxes the assumption that all the add-ons in inventory will be discarded, but also allows the add-on-to-base effect to vary across standard. These model advancements allow us to investigate how previous investment in add-ons affects a consumer replacement choice of base products that are of different standards.

Finally, our paper is related to the literature on the compatibility problem of base products and add-ons. Standard economics literature, mostly analytical works, claims that if products are incompatible, cost of switchings bind customers to vendors. Such cost of switchings not only involve direct efficiency losses but also soften competition and magnify incumbency advantages (see Farrell and Klemperer (2005) for a review). Therefore, consumers as well as economists favor compatibility, or in other words standardization (see Farrell and Simcoe (2012) for benefits of compatibility). However, on the supply side, firms have incentives to create incompatibility constraints. Matutes and Regibeau (1988) used a "mix and match" model to show that compatibility leads to higher pricing. Katz and Shapiro (1985) found that firms with good reputations or large existing networks tend to be against compatibility while firms with weak reputation tended to favor product compatibility. Focusing on the supply side, these models have a simplistic specification of consumer demand hence they are not able to capture consumers' dynamic decision making process.[4] Our paper takes a different approach and focuses on the rich characterization of consumers' inter-temporal tradeoffs. Additionally, our policy simulations reinforce and extend the findings in this analytical literature.

1.3 Industry Background and Data Description

1.3.1 Digital Camera and Memory Card Industries

Since 1994, the digital camera industry has seen constant technology improvements: higher pixel counts, larger sensors, shorter shutter lag, smaller and lighter bodies, and more optical zoom options. The market also saw a substantial increase in models and brands, with Canon, Casio, Fujifilm, Kodak, Nikon, Olympus, and Sony as the leading players. As digital cameras began taking higher quality pictures, consumers demanded larger memory devices to store photos. It was in this memory card territory that competition increased; manufactures

[4] They focus on deriving the optimal strategy for the firms.

developed multiple standards to take advantage of the exclusivity benefit of incompatibility. Table 1 is the adoption timeline of memory cards for different manufacturers.

Table 1 Memory Card Timeline[5]

<table>
<tr><th></th><th>Std. 1(SON)</th><th>Std. 2(OLY/FUJ)</th><th>Std. 3(KOD/CAN/HP/NIK)</th></tr>
<tr><td>1996</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2">PCMCIA</td></tr>
<tr><td>1997</td></tr>
<tr><td>1998</td><td>DISK</td><td rowspan="4">SM</td><td rowspan="3">CF</td></tr>
<tr><td>1999</td><td>DISK/MS</td></tr>
<tr><td>2000</td><td rowspan="5">MS</td></tr>
<tr><td>2001</td><td>CF/SD</td></tr>
<tr><td>2002</td><td>SM/XD</td><td rowspan="3">SD</td></tr>
<tr><td>2003</td><td rowspan="2">XD</td></tr>
<tr><td>2004</td></tr>
</table>

As shown in Table 1, accompanying Sony's first digital camera was a 3.5" floppy disk storage device. The desire for smaller cards led Sony to invest R&D resources to create its own memory card format, the Memory Stick, which was launched in October 1998. After its introduction, from 1999 to 2001, Memory Stick embraced a market expansion from 7% to 25%.[6] Meanwhile, from 1998 to 2004, the market share of Sony's cameras increased from 17% to 23%.[7] Since then, Sony has been using its proprietary standard and its extensions, such as Memory Stick PRO, Memory Stick Duo, and Memory Stick PRO Duo as its compatible storage device.

Olympus and Fujifilm, on the other hand, employed SmartMedia cards for their first few cameras and in July 2002, they jointly invented another standard, the xD card,[8] as the primary storage device to compete with Sony.

The success of the Sony Memory Stick also motivated SanDisk, Matsushita, and Toshiba to develop and market the SD (Secure Digital) memory card.[9] Early samples of the SD card became available in the first quarter of 2000. Later, in March 2003, SanDisk Corporation announced the introduction of the miniSD, a variant of the original SD Card. Because SD cards are ultra-compact, reliable, interoperable, and easy to use, many of leading digital camera manufacturers, including Canon, Kodak, Nikon, and HP, all of which originally used the CompactFlash card format, switched to SD cards in their consumer product lines in 2002.

---

[5]DISK: 3.5 floppy disk, MS: Memory Stick, SM: SmartMedia card, XD: xD card, CF: CompactFlash, SD: SD card

[6] http://news.cnet.com/2100-1040-268460.html

[7] http://www.pcworld.com/article/114711/sony_unveils_digicams_photo_printer.html

[8] http://en.wikipedia.org/wiki/XD-Picture_Card

[9] http://en.wikipedia.org/wiki/Secure_Digital

To summarize, we categorize memory cards into three standards.[10] Disk and Memory Stick (MS) are labeled as "Standard 1" and only Sony cameras are compatible with the Standard 1 cards. SmartMedia cards (SM) and xD cards (XD) are grouped as "Standard 2" with Olympus and Fujifilm cameras compatible with Standard 2 cards. CompactFlash (CF) and SD cards are called "Standard 3" cards with Kodak, Canon, HP, and Nikon cameras all adopting Standard 3 memory cards. We are going to use Standard 1, 2 and 3 in the Model section (section 1.4) to avoid confusion.

1.3.2 Data Description

The data is an individual level scanner panel provided by an anonymous major electronic retailer in the United States. Our sample consists of the complete purchase records of 828 randomly selected households that purchased at least one camera in six years, from December 1998 to November 2004. The transaction record includes detailed information about purchases of products, such as brand name, product type, price paid, time and location of purchases. In addition, we collect information on digital cameras at the brand level from a camera database website that tracks detailed information of all camera models.[11] The quality information on memory cards is obtained from annual reports of major memory card manufacturers at the standard level.[12] Following Song and Chintagunta (2003), we use effective pixels (in megapixels) as a proxy of camera quality because it is the most important factor in determining the performance. The quality of a memory card is measured by capacity (in megabytes). One limitation of the data is that we only know brand information, rather than product specifications such as the model name of each camera and the size of each memory card. Therefore we calculate the average quality of a brand from all models on the market and use this as a proxy. We assume no introduction of new formats and hence the choice set of consumers is identical during our observation period.

Table 2A. Summary of Purchase Incidences of Cameras and Memory Cards

Camera Purchases			**Memory Purchases**		
Brand	**Frequency**	**Percentage**	**Standard**	**Frequency**	**Percentage**
Sony	295	27.86%	**1 (Sony)**	309	29.63%
Olympus	172	16.24%	**2 (Olympus, Fuji)**	241	23.11%
Fuji	81	7.65%	**3 (Kodak, Canon, HP, Nikon)**	493	47.27%
Kodak	212	20.02%			
Canon	114	10.76%			
HP	89	8.40%			

[10] We are able to group two formats (for example, Disk and Memory Stick) as the same standard because when the new Memory Stick was launched, Sony's cameras were designed to use both formats (e.g. Sony's Cyber-shot DSC-D700). Moreover, adapters existed to transfer data from both formats of memory cards to the computer. In the case of Sony, although there were constantly new introductions (later versions of Memory Stick) to the market, i.e. Memory Stick Select and Memory Stick Pro, most devices that use the original Memory Sticks support both the original and PRO Sticks since both formats have identical form factors[10].

[11] www.dpreview.com/products

[12] www.dpreview.com/products

Nikon	96	9.07%			

Table 2B. Total Purchase Incidences

Camera\Memory	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
1	17	621	56	5	3	0	702
	2.05%	75.00%	6.76%	0.60%	0.36%	0.00%	84.78%
2	11	8	6	2	0	0	27
	1.33%	0.97%	0.72%	0.24%	0.00%	0.00%	3.26%
3	4	10	26	47	5	1	93
	0.48%	1.21%	3.14%	5.68%	0.60%	0.12%	11.23%
4	0	0	0	0	1	5	6
	0.00%	0.00%	0.00%	0.00%	0.12%	0.60%	0.72%
Total	32	639	88	54	9	6	828
	3.86%	77.17%	10.63%	6.52%	1.09%	0.72%	100.00%

Table 2C. Summary Statistics of Price and Quality

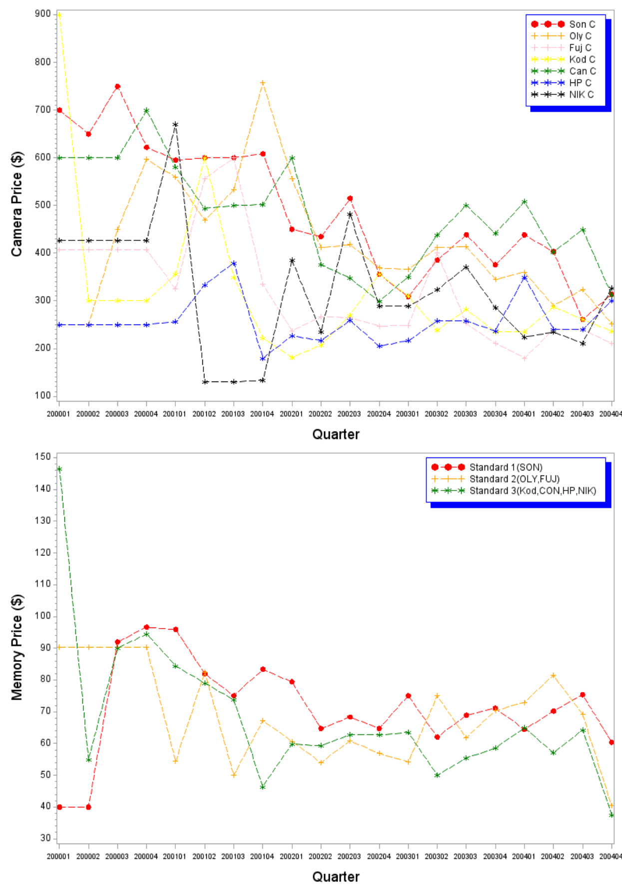
	Sony	**Olympus**	**Fuji**	**Kodak**	**Canon**	**HP**	**Nikon**	**M 1**	**M 2**	**M 3**
Price	521.577	429.172	339.028	387.213	504.043	256.239	342.537	65.182	72.989	62.230
Quality	3.898	3.895	3.547	3.889	4.082	3.316	4.444	3.058	2.900	3.089

We prepare the data in the time frequency of a quarter because consumers seldom purchase cameras and memory cards more frequently than that. During the six-year sample period, the 828 households made 1059 transactions of cameras and 1043 purchases of memory cards. Table 2A presents market shares of different brands of cameras and memory cards. In the digital camera market, Sony had the largest market share of 27.86%. Olympus and Fujifilm together took up 23.89%, and left the remaining 48.15% to other brands. Consistently, Standard 1 memory cards (compatible with Sony cameras) had a market share of 29.63%, Standard 2 memory cards (compatible with Olympus and Fujifilm cameras) had a market share of 23.11% and Standard 3 memory cards (compatible with Kodak, Canon, HP, and Nikon cameras) occupied 47.27%.

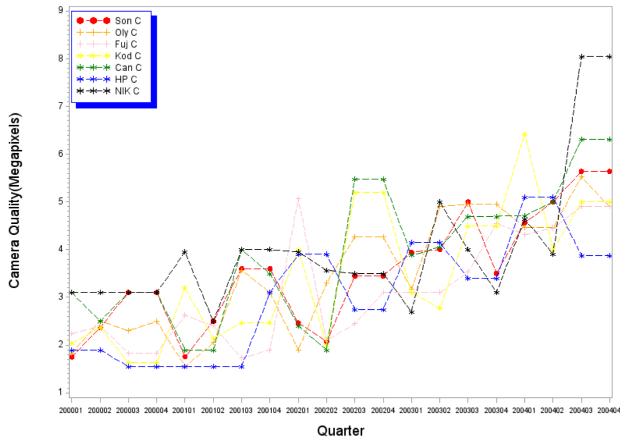
Table 2B reports the total purchase incidences for 828 consumers. 15.22% of consumers had replaced cameras while 84.78% of consumers purchased one camera; 18.97% of consumers purchased more than one memory card. The maximum number of camera purchase incidences is three and the maximum number of memory card purchase incidences is four. These numbers are consistent with the nature of cameras and memory cards as durable goods.

Table 2C reports the summary statistics of price and quality information. On average, the price of the Sony cameras is the highest and that of the HP cameras is the lowest. Interestingly, the quality measure is not quite aligned with price as Nikon, rather than Sony, has the highest average quality. For memory card, Standard 2 is the highest priced with lowest average quality while Standard 3 charges the lowest price with the highest average quality. The misalignment of price and quality triggers our interest to investigate their impacts on sales.

Figures 1A and 1B. Original Price Trend of Camera and Memory Card by Quarter



Figures 2A and 2B. Quality Trend of Camera and Memory Card



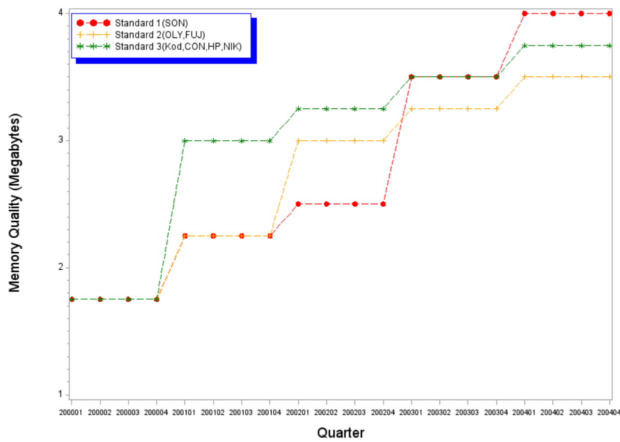


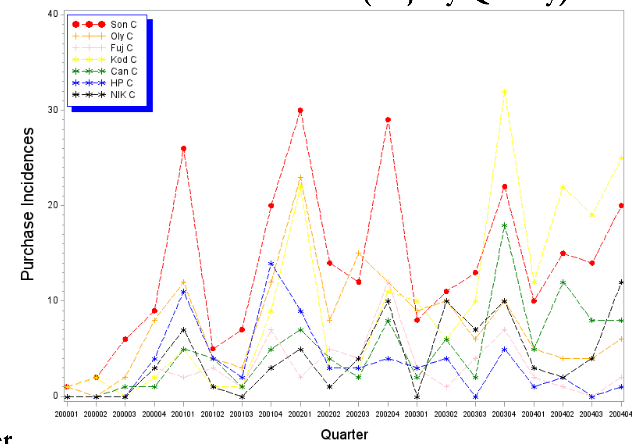
Figure 1A and 1B exhibit the price trend of cameras and memory cards. We find that the price of Sony cameras decreased over time. Prices of Olympus and Fujifilm cameras increased in 2000 and 2001 and then decreased for the rest of the sample periods. Prices of Kodak, Cannon, Nikon and HP decreased at the beginning and stabilized (or slightly increased for Kodak) later on. In terms of memory cards, standard 1 almost always had the highest price except after 2002, when the price of standard 2 caught up. Standard 3 charged a lower price than standard 2 after the second quarter of 2002 and stayed with the lowest price among the three standards.

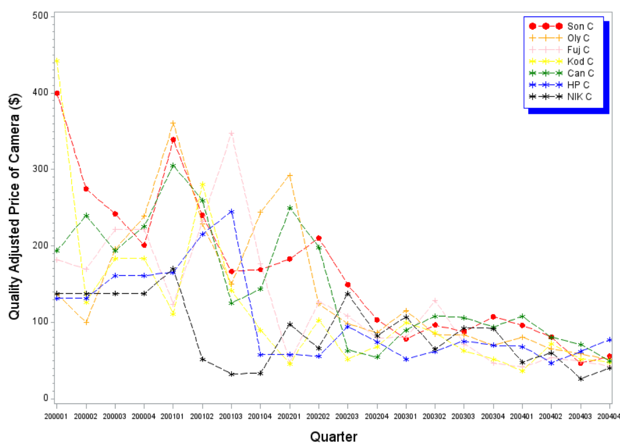
Figure 2A and 2B show the corresponding quality trend of camera and memory card. During our sample period, technology improved dramatically and all products saw a significant quality upgrade. Interestingly, there's no clear quality differentiation among brands of cameras, in other words, no brand had the dominating quality position throughout time. In addition, although Sony set a relatively high price in both the camera and memory card market, it doesn't have clear quality advantage over its competitors. Recall that Table 2A reveals Sony as the camera market leader, which cannot be explained by Sony's high price and low quality. This intriguing puzzle motivates us to investigate the dynamic impact of add-on products on a consumer's choice of base products by modeling the micro-foundations of consumer decisions.

1.3.3 Model Free Evidence of Cross-Category Inter-temporal Dependence

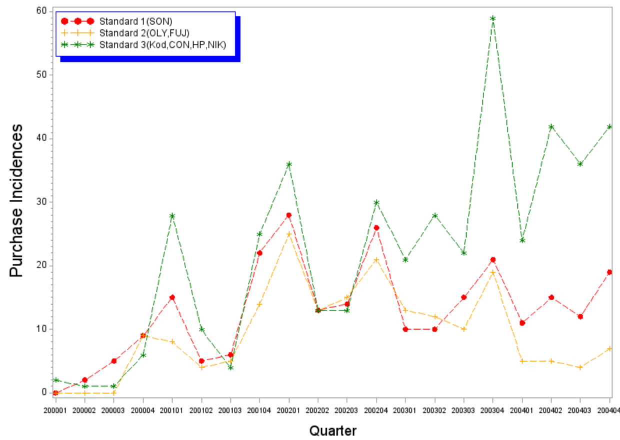
Cross-Category Price/Quality Effect

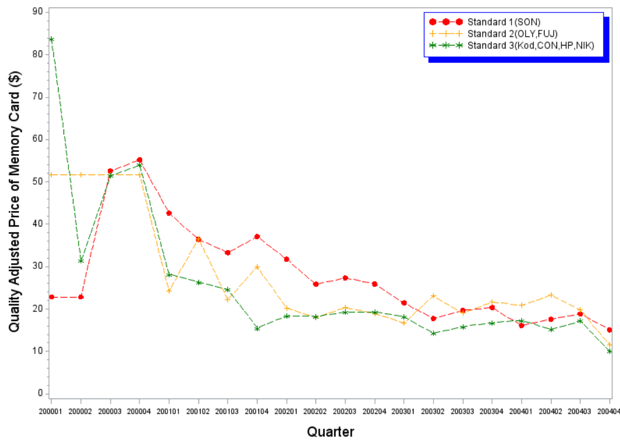
Figures 3A and 3B. Purchase Incidences and Price (Adj. by Quality) Trend of Camera by Quarter





Figures 4A and 4B. Purchase Incidences and Price (Adjusted by Quality) Trend of Memory Card by Quarter





For technology goods like cameras and memory cards, prices highly depend on features of the model. Prices alone don't provide the true nature of the product, thus we need to use quality-adjusted price. Figure 3A illustrates how demand of cameras evolved over time whereas Figure 3B shows the quality-adjusted price trends for each camera brand. We also present demand and quality-adjusted price trends of memory cards in Figures 4A and 4B. There are two interesting findings to note from these figures.

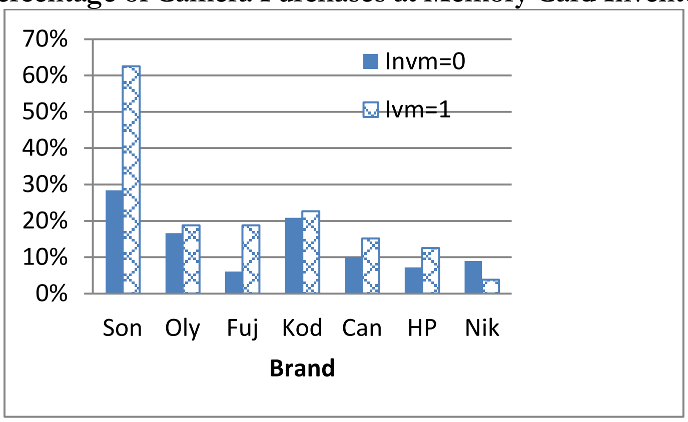
First, the price gap between Sony and most Standard 3 cameras became smaller over time--Sony's price fell over the entire time period whereas Canon's fell at the beginning but slightly increased for the remainder of the data period when it became higher than Sony's price. Yet, Canon's sales of cameras increased faster than those of Sony's. So price of the camera market alone cannot explain the demand pattern. But when we look at the memory card market, Standard 3 always charged the lowest price. We conjecture that it's because when making a purchase decision of cameras, consumers not only consider price/quality of cameras but also the add-on, memory cards. Lower future prices of memory cards decrease the total financial cost for consumers and become the driver of consumers' choice of cameras.

Second, although Olympus and Fuji cut camera price aggressively after the second half of 2002, it did not help save their camera sales. Perhaps such a limited response is a consequence of the price of memory cards compatible with their cameras rising and of forward-planning customers realizing the long-term financial burden of having such a bundle of products.

The above two points provide some evidence for a "cross-category dynamic price effect". More specifically, if consumers anticipate the price of future add-on products as rising, they will switch brands in the base product category to minimize the total financial burden of the product portfolio.

Add-on-to-base Effect (Cross-category Dynamic Inventory Effect)

Figure 5. Percentage of Camera Purchases at Memory Card Inventory at 0 vs. 1



Recall that in section 1.3.2, we discovered a puzzle about Sony's high market share versus high price and low quality. We conjecture that in addition to the above memory card price effect on camera purchase incidence, perhaps inventory of memory cards also plays an important role in camera purchases, what we call the "add-on-to-base effect". We assume memory cards do not become obsolete, and with this a consumer who owns a memory card should be more reluctant to switch to a camera that is incompatible with her existing stock of memory inventory. On the other hand, a consumer who has zero stock of memory card inventory is not "locked-in" to a particular camera brand. Figure 5 illustrates the purchase incidences for each camera brand conditional on consumer inventory levels of compatible memory cards. We see that for all camera brands, purchase incidence increases as the inventory level of compatible memory card increase. This is particularly true for Sony: it appears that a Sony consumer is locked-in and perhaps faces higher cost of switchings or add-on-to-base effects associated with existing memory card inventory than consumers who own other standards.

In summary, the data pattern shows the cross-category inter-temporal interdependence between purchases of base and add-on products. It is evident that forward-planning consumers take into account the price/quality of add-ons as well as financial implications of discarding their existing add-ons when comparing long-term utilities of alternative choice sequences. In the next session, we develop a model to explicitly describe this decision process.

1.4 Model

Consumer purchase behavior of high-tech and durable goods is distinguished from that of consumer package goods on several fronts. Existing literature has established that prices of technology products decline while qualities improve over time. Accordingly, our model of consumer adoption of products accounts for the fact consumers anticipate these future price and quality trajectories while deciding when to purchase. Moreover, since base products are durable in nature and add-on products can be purchased subsequently, consumers tend to look into the future when making purchasing decisions (Nair, Chintagunta and Dube 2004, Derdenger and Kumar 2013). The forward-looking behavior of consumers and the issue of compatibility between camera and memory cards, imply that a consumer's decision of purchasing the base product depends on the anticipated purchase(s) of the add-on products. Therefore, the purchase decision for the base product would depend not only on the expected price and quality trajectories of that product, but also on the anticipated price and quality of the add-on product. To approximate a consumer's decision process that accounts for the above characteristics, we develop a model of consumers' joint purchase (adoption and replacement) decisions of base and add-on products as a dynamic optimization problem under price and quality uncertainty.

1.4.1 Assumption

In light of the data on hand and the specific industry we study, we make several assumptions regarding consumer behavior for model parsimony. First, we assume consumers can buy only at the focal store. This assumption may seem quite restrictive as consumers may often purchase at several electronic stores. However, this concern is mitigated in our sample for several reasons. One, all households are holders of loyalty cards of the store so they are frequent buyers in the store. Two, the focal store offers a "low-price guarantee" that will beat any price from a local store stocking the same item. This highly competitive pricing strategy provides a large disincentive for these loyal consumers to purchase at rival stores. Three, we delete households that only purchased memory cards from the store. These consumers are more likely to purchase from multiple stores rather than only one. Lastly, our observed data pattern from this store is representative of the industry average (See Appendix A1.3). For example, we observe a camera replacement cycle of 4.67 years while the industry average is 4.30 years with a standard deviation of 2.28 years. Above all, we acknowledge the data limitation and only claim that we provide insight on loyal consumers' brand and standard choice behavior within a store. Store competition is beyond our scope of this research. Our second assumption treats a consumer who buys multiple memory cards on a single trip as only one purchase incidence. This assumption is reasonable because only a very small portion (0.6%) of the purchases in our sample involves multiple items.[13] Third, we assume there is no resale market for cameras and a discarded camera cannot be exchanged for its residual value. This implicitly assumes consumers only derive utility from their most recently purchased camera. Finally, we assume that consumers keep all memory cards i.e. memory cards are accumulated not replaced. Past research of durable good replacement purchases ignores the quantity of memory card in inventory, which is equivalent to assuming consumers discard all the add-on products that they purchased before and ignore them when making decisions on base product replacement choices (it is fine if they do not study standard choice). In contrast, we relax the assumption and allow inventory to be cumulative. We rely on the estimated coefficient (γ to be explained later) to tell us the extent to which the memory cards in inventory affect consumers' decision to buy a new camera.

1.4.2 Consumer Choices and Flow Utility

Our model follows the large literature pertaining to choice models (Guadagni and Little 1983). The per-period utility for consumer i $(i = 1,2, \dots, I)$ who makes purchase decisions of both the base product (camera of brand c) and the add-on (memory card of standard m) jointly at time period t $(t = 1,2,..,T)$ can be decomposed into a deterministic part $\overline{U}_{it}^{c,m}$ and an idiosyncratic error term $\varepsilon_{it}^{c,m}$.

$$U_{it}^{c,m} = \overline{U}_{it}^{c,m} + \varepsilon_{it}^{c,m} \tag{1}$$

The deterministic part of the per-period utility $\overline{U}_{it}^{c,m}$ has three elements 1) consumption utility from using the camera (with the internal memory or small external memory card that comes with the camera purchase) 2) enhanced consumption utility that is associated with the additional number of add-ons, and 3) cost of purchasing/replacing the products. We adopt an additive utility specification which follows a large body of literature of complementary good[14] and multi-category purchases[15]. We also conduct robustness test with other utility specifications, like multiplicative utility. These alternative utility functions and estimation results are exhibited in Appendix A1.4.

So in specific,

$$\overline{U}_{it}^{c,m} = \underbrace{U_{it}^{c}}_{camera\ consumption} + \underbrace{U_{it}^{m} * I(c \sim m)}_{memory\ consumption} - \underbrace{\lambda_i * (PC_t^c + PM_t^m)}_{financial\ cost} \tag{2}$$

1) Consumption Utility of Camera

---

[13] When multiple memory cards are bought, we treat each item as a separate purchase incidence. The state variable, inventory of the memory cards is cumulated by the number of items bought. We delete the consumer who purchased multiple cameras because this might be a case of several family members adopt together. We only examine the behavior of an individual consumer.

[14] Cheng and Nahm 2007, Chen and Nalebuff 2006, Nair, Chintagunta and Dube 2004, Lee 2013, Derdenger and Kumar 2013, Derdenger 2014.

[15] Sriram, Chintagunta and Agarwal 2009 and Liu, Chintagunta and Zhu 2010, Manchanda et al. 1999, Seetharaman et al. 1999, Russell and Petersen 2000, Chung and Rao 2003, Wedel and Zhang 2004, Seetharaman et al. 2005, Song and Chintagunta 2006, Gentzkow 2007.

Since most camera models have internal memories[1] with an average size of 16.2 MB or come with a free small-size (for example 16 MB[1]) memory card at the time of the purchase, the cameras can function by themselves.

$$U_{it}^{c} = \alpha_i^c + \phi_i QC_t^c + SD_i * I(c = \bar{c}) \tag{3}$$

As shown in equation (3), the consumption utility of camera (U_{it}^{c}) is summarized by the brand-specific constant (α_i^c), quality ($\phi_i QC_t^c$) and state dependence ($SD_i * I(c = \bar{c})$). The first term α_i^c is the brand-specific fixed effect, which represents a persistent form of product differentiation that captures the household's intrinsic brand preferences of camera brand c. The second term QC_t^c is the quality of the camera c at its purchase time t. Quality is measured by megapixels as in Song and Chintagunta (2003). The coefficient ϕ_i is the marginal utility for a single unit of quality increment. If $\phi_i > 0$, a consumer obtains higher utility from a high quality product than from a low quality product. The next term $SD_i * I(c = \bar{c})$ denotes state dependence (Dubé, Hitsch and Rossi 2010), that is if the consumer considers purchasing a camera (brand c) of the same brand as the one she already owns (brand $\bar{c}$), she can receive an extra utility SD_i compared to other brand choices. If $SD_i > 0$, then the model predicts persistency in camera brand choices. More specifically, if a household adopts brand c, the probability of a repeat purchase of brand c is higher than prior to this purchase: the conditional choice probability of repeat-purchasing exceeds the marginal choice probability. There could be different behavioral mechanisms that generate state dependence. One is that consumers have become loyal to the brand because of their past user experiences, thus incurring a psychological cost of switching to choose other brands. Or a consumer is going through a learning process that she finds high match value from the brand. The purpose of this paper is not trying to differentiate these explanations. We will simply capture the "state dependence" effect as it is.

2) Consumption Utility of Add-on Memory Cards

The add-on memory cards refer to the additional memory cards that do not come with the camera. These additional memory cards enhance the utility of the camera by adding more storage space, offering flexibility of taking more and larger size (more pixels) pictures (e.g. on a trip), shooting videos, decreasing the frequency of transferring pictures to computers, hedging failure of other memory cards, etc. It is the additional investment on add-ons that does not come with the base product camera.

Note that add-on products only provide consumption value to the compatible base product. The indicator $I(c \sim m)$ in equation (2) denotes that only compatible memory cards can enhance the utility of a camera. From the data description section (section 1.3.2) we know that $(c = 1) \sim (m = 1)$[16]; $(c = 2,3) \sim (m = 2)$; $(c = 4,5,6,7) \sim (m = 3)$.

[16] $(c = 1) \sim (m = 1)$ means that Sony camera is compatible with standard 1 memory card (Memory Stick)

$$U_{it}^{m} = \underbrace{u_{it}^{m}}_{new\ purchase} + \gamma_i^m * \underbrace{u_{it}^{INVM}}_{inventory} \tag{4}$$

$$u_{it}^{m} = \theta_i^m + \psi_i * QM_t^m \tag{4-1}$$

$$u_{it}^{INVM} = \sum_{k=1}^{INVM_{it-1}^{m}} u_{iPT_k}^{m} \tag{4-2}$$

According to our assumption, consumption utility from the memory card (U_{it}^{m}) includes the utility of the currently purchased memory card (u_{it}^{m}) as well as utility from all memory cards in inventory (u_{it}^{INVM}), as shown in equation (4). (The coefficient $\gamma_i^m$is going to be explained later in Case 1.) For a newly purchased memory card of standard m, the consumption utility (equation 4-1) is summarized by θ_i^m, the standard-specific fixed effect and quality, QM_t^m, measured by megabytes. The coefficient ψ_i is consumer i's sensitivity to memory card quality or storage capacity. Utility of memory cards in inventory (equation 4-2) sums the consumption utility of each item k from 1 to $INVM_{it-1}^{m}$, the total number of memory cards (standard m) purchased by consumer i up to $t-1$. Note that we don't allow depreciation, but with quality $QM_{PT_k}^{m}$ being associated with the purchase time PT_k, memory cards purchased prior to period t naturally provide less utility compared to a current card due to the lower quality in the past.

3) Cost of Purchasing

Finally, the cost of purchasing is the sum of PC_t^c, price for camera of brand c, and PM_t^m, the price for memory card of standard m., The coefficient λ_i is the price sensitivity.

We allow consumers to choose from multiple brands of cameras ($c \in \{0,1,2,\dots,C\}$)and multiple standards of memory cards $m \in \{0,1,2,\dots,M\}$ where 0 denotes no purchase, and C/M is the total number of camera brands/memory card standards. In our data, $c = 1,2,3,4,5,6,7$ represents Sony, Olympus, Fujifilm, Kodak, Canon, HP and Nikon respectively while $m = 1,2,3$ corresponds to Standard 1 (Memory Stick), Standard 2 (SmartMedia/xD card), Standard 3 (CompactFlash/SD card) respectively. Thus, during each time period, a consumer faces altogether 18 choice alternatives,[17] which can be classified into four types of choice alternatives: (1) purchase (adopt or replace) only a camera of brand c, (2) purchase only a memory card of standard m , (3) purchase one camera and one memory card together and (4) purchase neither product.

Case 1: Purchase Camera Only

[17] Utility functions for each of the 18 choice alternatives of this full model are shown in Table A1.1 of the Appendix.

When the consumer only purchases a camera but no memory cards, she obtains consumption utility from the camera and pays for the purchase. In addition, she obtains enhanced utility associated with the compatible add-ons in inventory. So for $c \in \{1,2, \ldots, C\}$

$$\bar{U}_{it}^{c,0} = \underbrace{U_{it}^{c}}_{camera\ consumption} + \underbrace{\gamma_i^m * u_{it}^{INVM} * I(c \sim m)}_{memory\ consumption(inventory)} - \underbrace{\lambda_i * PC_t^c}_{financial\ cost}{}^{18} \qquad (5)$$

As you can see from the second term in equation (5), the more memory cards (larger $INVM_{it}^{m}$) a consumer has, and/or the better quality the memory cards (larger $QM_{PT_k}^{m}$) are, the higher the utility a consumer can derive from the inventory. A positive coefficient γ_i^m corresponds to memory cards in inventory having an impact on a consumer's decision to purchase a new camera. We call this effect the "add-on-to-base effect" which links a consumer purchase decision of a camera and that of memory cards into a single framework, i.e., a forward-looking consumer who makes a purchase decision of a camera at time *t* will consider not only the extra utility generated by the compatible memory cards in inventory, but also the effect of future price and quality of the memory cards. Without this term, the purchase decisions of the two categories will be separated. Note however, if a consumer chooses a camera that is incompatible with the memory cards in inventory, she cannot continue deriving utilities from these memory cards. She instead loses the utility provided by these memory cards and must re-invest on more memory cards to enhance the consumption value of the camera in the future.

Fundamentally different from existing literature on cross-category purchases of durable goods, our approach allows us to investigate the dynamic and interdependent consumer decision process: First, we recognize the compatibility at the brand (for camera) and standard level (for memory cards). This allows us to study how brand choices of base products are driven by past, current, and future choices of standard of the add-on products. Second, we allow the add-on-to-base effect to depend on the number and quality of the add-on products owned. Therefore, the add-on-to-base effect can vary across time and affect the inter-temporal decision-making of forward-looking consumers--since the more compatible memory cards that are accumulated, the higher the per-period add-on-to-base effect. This implies that the accumulation of add-on products creates a higher cost of switching for consumers to abandon the compatible base product. Note that our definition of the add-on-to base effect is consistent with the literature that recognizes complementarity between product categories (Sriram, Chintagunta and Agarwal (2009) and Liu, Chintagunta and Zhu (2010)). However, previous models define the complementary term as time-invariant and only at the category level. We advance the literature by making it time varying and standard specific. In specific, the add-on-to-base effect coefficient (γ^m) is standard specific (superscript m). This supports our model free evidence in section 1.3.3 that the effect of inventory of memory cards on camera purchases varies for different standards of memory

[18] When there's no purchase of memory card, the consumption utility from new purchase is zero. Besides, $PM_t^0 = 0$, the cost is zero.

cards. We hope to compare the magnitude of different add-on-to-base effects in order to explain the observed data pattern and conduct meaningful counterfactual analysis.

Case 2: Purchase Memory Only

When a consumer buys only a memory card, she must have owned a compatible camera. So her utility originates from consumption utility of using the camera in inventory and enhanced utility that comes from the additional purchases of memory cards net the cost of purchasing. So for $m = 1,2,3$ the utility function takes the form

$$\bar{U}_{it}^{0,m} = \underbrace{U_{it}^{\bar{c}}}_{camera\ consumption} + \underbrace{\left(\underbrace{u_{it}^{m}}_{new\ purchase} + \gamma_i^m * \underbrace{u_{it}^{INVM}}_{inventory} \right) * I(\bar{c} \sim m)}_{memory\ consumption} - \underbrace{\lambda_i * PM_t^m}_{financial\ cost} \text{[19]} \tag{6}$$

where $\bar{c}$ is the brand of the camera in inventory. Note that when the consumer only purchases the memory card, she will definitely pick the standard that is compatible with the camera in inventory ($I(\bar{c} \sim m) = 1$) because other memory card standards cannot be used with the camera at hand.

Case 3: Camera & Memory

When a consumer simultaneously purchases camera $c(> 0)$ and memory card $m(> 0)$, they must be compatible (no consumer purchased incompatible base product and add-on at the same time in our data).

$$\bar{U}_{it}^{c,m} = \underbrace{U_{it}^{c}}_{camera\ consumption} + \underbrace{U_{it}^{m} * I(c \sim m)}_{memory\ consumption} - \underbrace{\lambda_i * (PC_t^c + PM_t^m)}_{financial\ cost} \tag{7}$$

The utility function is a combination of those in the previous two cases.

Case 4: No Purchase

If a consumer does not own a camera and she decides not to make a purchase of any product at time *t*, we normalize the utility to zero.

$$\bar{U}_{it}^{0,0} = 0\ .$$

However, if the consumer owns a camera and decides not to replace it with a new one, she continues to receive utility from the camera and the compatible memory cards in inventory (if there is any) without paying additional cost. Thus, the utility function has two components: possession of a camera $\bar{c}$, and the add-on-to-base effect provided by inventory of compatible memory cards.

[19] When there's no purchase of camera, $PC_t^0 = 0$, cost of camera is zero.

$$\bar{U}_{it}^{0,0} = \underbrace{U_{it}^{\bar{c}}}_{camera\ consumption} + \underbrace{\gamma_i^m * u_{it}^{INVM} * I(\bar{c} \sim m)}_{memory\ consumption(inventory)} \tag{8}$$

1.4.3 State Transitions

Inventory Process

According to our assumptions, a consumer uses only the latest purchased camera. So the camera inventory variable, $\bar{c}$, is just an indicator of the brand of camera owned at time t. When the consumer buys a new camera c, its inventory switches from $\bar{c}$ to c. When no camera is purchased at time t, the inventory remains the same as in the last period. So the inventory process for cameras is (after dropping the consumer index i)

$$\bar{c}_{t+1} = \begin{cases} c, & if\ D_t^{c,\cdot} = 1 \\ \bar{c}_t, & if\ D_t^{c',\cdot} = 0\ for\ all\ c' \in \{1,2,\dots,C\} \end{cases} \tag{9}$$

where $D_t^{c,\cdot}$ is the indicator of consumer's choice, with $D_t^{c,\cdot} = 1$ denoting the consumer purchasing brand c $(c = 1,2,3,4,5,6,7)$ as the base product and any memory card (including no purchase) as an add-on product. $\bar{c}_t$ is the beginning camera inventory at time t.

On the other hand, since memory cards do not become obsolete, the inventory process of memory cards is simply the accumulation of purchased cards over time up to period t.

$$INVM_t^m = \begin{cases} INVM_{t-1}^m + D_t^{\cdot,m}, & if \sum_{k=1}^{M} INVM_{t-1}^k < \bar{B} \\ INVM_{t-1}^m, & otherwise \end{cases} \tag{10}$$

Where $INVM_{t-1}^m$ is the ending inventory, total number of memory card m at time $t-1$ and $D_t^{\cdot,m}$ is new purchase during period t. If no purchase is made at time t, $D_t^{\cdot,m} = 0$. This process is in contrast to that of fast-moving packaged goods, for which inventory is the cumulative purchases minus consumption throughout time. Similar to Hartmann & Nair (2010) and Derdenger and Kumar (2013), we restrict the number of memory cards in inventory to be less than $\bar{B}$[20] to keep the state space for the dynamic problem bounded. The transition matrix of inventory process from period t to period t+1 for the 18 choice - options is shown in Table A1.2 of Appendix A1.1.

[20] In our empirical application, we set $\bar{B} = 4$ and assume that a memory card of too old age (four years) will be obsolete.

Price and Quality Process

We assume that consumers have rational expectations about the stochastic processes governing the evolution of price and quality, which follow a first-order vector autoregressive (VAR) process. We also take into account competitive reaction, i.e. the price and quality expectation of one brand/standard depends on not only the lag price and quality of itself, but also that of all other competitors in the market. Furthermore, we capture cross-category effect where the price/quality of a product in one category (e.g. cameras) depends on the lagged price/quality of all products in the other category (memory cards, including both compatible and incompatible ones).

$$
\begin{aligned}
E(\ln \boldsymbol{P_t} \,|\Omega_{t-1}) &= \boldsymbol{\Lambda}_\mathbf{p} \ln \boldsymbol{P_{t-1}} + \omega Holiday_t + \boldsymbol{\eta_{pt}} \\
E(\ln \boldsymbol{Q_t} \,|\Omega_{t-1}) &= \boldsymbol{\Lambda}_\mathbf{q} \ln \boldsymbol{Q_{t-1}} + \boldsymbol{\eta_{qt}}
\end{aligned}
\tag{11}
$$

Letters in bold denote vectors of all choice alternatives. More specifically, $\boldsymbol{P_t}$ is a column vector that includes all prices of cameras and memory cards, i.e. $\boldsymbol{P_t} = [PC_t^1\ PC_t^2\ PC_t^3\ PC_t^4\ PC_t^5\ PC_t^6\ PC_t^7\ PM_t^1\ PM_t^2\ PM_t^3]^T$ (T denotes transpose) and $\boldsymbol{Q_t}$ is a column vector that includes all qualities of cameras and memory cards, $\boldsymbol{Q_t} = [QC_t^1\ QC_t^2\ QC_t^3\ QC_t^4\ QC_t^5\ QC_t^6\ QC_t^7\ QM_t^1\ QM_t^2\ QM_t^3]^T$. $\underbrace{\boldsymbol{\Lambda}_\mathbf{p}}_{(C+M)\times(C+M)}$ and $\underbrace{\boldsymbol{\Lambda}_\mathbf{q}}_{(C+M)\times(C+M)}$ are matrices that capture the influence of competitors' price/quality. We include $Holiday_t$, a dummy that indicates the fourth and first quarter of the year, in the price process because we observe significant discount during the holiday season (Figure 1A). $E(.\,|\Omega_{t-1})$ is the conditional expectation given a set of state variables Ω_{t-1}. $\boldsymbol{\eta_{pt}}$ is a column vector of random price shocks at time t and $\boldsymbol{\eta_{qt}}$ is a column vector of random quality shocks. We assume random shocks in prices/qualities follow a multivariate normal distribution:

$$
\boldsymbol{\eta_{pt}} \sim N\left(0, \Sigma_{\eta p}\right), \boldsymbol{\eta_{qt}} \sim N(0, \Sigma_{\eta q}) \tag{12}
$$

Allowing random shocks to be correlated can further capture the co-movement of prices (qualities) of the competing brands. In this fashion, we utilize all past variables (price/quality) to characterize market dynamic interaction in a reduced form representation. The price (quality) process parameters are estimated using the price (quality) data prior to the estimation of the model. They are then treated as known in the model estimation when we solve the consumer's dynamic optimization problem.

1.4.4 Dynamic Optimization Problem and Inter-temporal Tradeoffs

Given the base products and add-ons are durable in nature, we follow the standard literature and assume the objective of consumer i is to maximize the expected present value of utility over the (finite) planning horizon $t = 1,2, \ldots, T$

$$\max_{D_{it}^{c,m}}\{E[\sum_{\tau=t}^{T}\sum_{c=0}^{C}\sum_{m=0}^{M}\delta_i^{\tau}D_{it}^{c,m}U_{it}^{c,m}|\Omega_{it}]\} \tag{13}$$

where δ is the discount factor. $D_{it}^{c,m}$ is the choice indicator with $D_{it}^{c,m}=1$ indicating alternative (c,m) is chosen by the decision maker i at time t and $D_{it}^{c,m}=0$ indicates otherwise. Choice options are mutually exclusive, so that $\sum_{c=0}^{C}\sum_{m=0}^{M}D_{it}^{c,m}=1$. The state space for the dynamic optimization problem at time t for consumer i is Ω_{it} which consists of the set of inventory of camera and memory cards, their prices, qualities and the vector of unobserved taste shocks, so

$$\Omega_{it}=\{\bar{\boldsymbol{c}}_{it},\boldsymbol{INVM}_{it},\boldsymbol{PC}_{it},\boldsymbol{PM}_{it},\boldsymbol{QC}_{it},\boldsymbol{QM}_{it},\boldsymbol{\varepsilon}_{it}\} \tag{14}$$

with letters in bold denoting vectors of all choice alternatives.

Our model inherently allows for three important inter-temporal tradeoffs. First, within each product category a consumer faces the trade-off of purchase timing due to fast declining price and improving quality. This buy-now-or-later tradeoff is well documented in marketing literature (Song and Chintagunta 2003, Gordon 2009 and Gowrisankaran and Rysman 2012). Second, because a consumer makes purchase decision of both base products and add-ons at the same time, she might sacrifice the price loss of base products to achieve an optimal strategy of purchasing the whole bundle. For example, when purchasing cameras, a consumer has two alternatives, brand A with high quality-adjusted price and brand B with low quality-adjusted price. However, she anticipates the future price of memory cards compatible with brand A will be much lower than those of brand B compatible memory cards. In this case, she sacrifices a high price in the camera category but gains more utility in the memory card category so that the financial cost of the portfolio is minimized. A similar logic applies if future high quality of memory cards compensates for current low quality of camera. We refer to this trade-off as cross-category dynamic price/quality effect (Hartmann and Nair 2010, Derdenger and Kumar 2013). Third, a compatibility constraint between a camera and memory cards creates a tradeoff of switching standards (Farrell and Klemperer 2005). For example, think about a consumer who owns a Sony's Memory Stick. When deciding which camera to purchase in a replacement occasion if the consumer switches to a camera which uses a different standard of memory card, the consumer forgoes the continuous future consumption utilities provided by the Memory Stick. Moreover, she has to incur more financial cost to purchase additional memory cards to enhance the utility of the new camera. These losses can only be offset by higher total future utilities from the new brand of camera by offering higher quality at a lower price than Sony. In summary, our model incorporates trade-offs regarding own-product inter-temporal price and quality effect, cross-category price and quality effect and a cross-category dynamic inventory effect. To our knowledge, this is the first paper to study these three effects simultaneously.

1.4.5 Heterogeneity, Initial Value, and Estimation

We adopt a latent class approach (Kamakura and Russell 1989) to incorporate unobserved heterogeneity for quality preference, the add-on-to-base effect, and consumer price sensitivity.

With our data originating near the inception of the digital camera industry, we set the initial state variables for camera and memory card inventories to be zeroes for nearly all consumers. We support this assumption with several facts. The penetration rate of digital camera in the US in 1998 was a mere 0.35%.[21] Moreover, before 1998, only 9 models[22] of camera were launched, among which four models did not allow for external memory. The other five cameras were extremely expensive (average price of $1220) and had expensive compatible memory cards (SmartMedia Card: $259 for 30MB[23] or CompactFlash Card: $140 for 24MB[24]). Nevertheless, there could be rare exceptions for the very early adopters of camera and we accommodate this off chance in our estimation procedure. In our dataset, we observe roughly 1.09% (9/828) of the total sample occasions where a consumer buys a memory card before a camera. To rationalize this data pattern, we assume the consumer had adopted a compatible camera before the observation period (exact purchase time is randomly assigned to a quarter between 1994 and 1998 and for a memory card standard that is compatible with multiple camera brands, we randomly assign a brand.[25]

The maximization of (13) is accomplished by choosing the optimal sequence of control variables $\{D_{it}^{c,m}\}$ for $c \in \{0,1,\dots,C\}, m \in \{0,1,2,\dots,M\}$ and $\tau \in \{1,2,\dots,T\}$. Define the maximum expected value of discounted lifetime utility as

$$V_{it}(\Omega_{it}) = \max_{\{D_{it}^{c,m}\}} \{D_{it}^{c,m} U_{it}^{c,m} + \delta E[\sum_{\tau=t+1}^{T} \sum_{c'=0}^{C} \sum_{m'=0}^{M} \max_{D_{i\tau}^{c',m'}} \delta_i^{\tau} D_{i\tau}^{c',m'} U_{i\tau}^{c',m'} | \Omega_{it}, D_{it}^{c,m}]\} \tag{15}$$

The value function V depends on the state at t. Given t takes values from an interval of finite length, the value function can be written as

$$V_{it}(\Omega_{it}) = \max_{c,m}(V_{it}^{c,m}(\Omega_{it})). \tag{16}$$

Based on the Bellman equation (Bellman 1957),

---

[21] Worldwide Digital Camera Market Review and Forecast, 1997-2003 (IDC #B99S2172)

[22] Within the seven brands we are considering. The 9 models are Olympus D-200L, Olympus D-300L, Olympus D-500L, Olympus D-600L, Fujifilm DS-300, Canon PowerShot 600, Canon PowerShot 350, Nikon Coolpix 100, Nikon Coolpix 300.

[23] http://www.epi-centre.com/reports/9802seye.html

[24] http://zonezero.com/magazine/dcorner/texto8.html

[25] Note our model has already taken care of the state-dependence effect. So any missing value of initial purchases of cameras won't bias our estimated add-on-to-base effect. We admit that if unfortunately there's missing value of memory card purchases before the sample started, we might overestimate the add-on-to-base effect. But given so much evidence, we don't think the problem is severe enough to overturn our results.

$$V_{it}^{c,m}(\Omega_{it}) = \bar{U}_{it}^{c,m} + \varepsilon_{it}^{c,m} + \delta E \max_{D_{it+1}^{c',m'}} [D_{it+1}^{c',m'} V_{it+1}^{c',m'}(\Omega_{it+1}) | \Omega_{it}, D_{it}^{c,m} = 1] \quad (17)$$

at time T, the choice-specific value function is simply $V_{iT}^{c,m}(\Omega_{iT}) = \bar{U}_{iT}^{c,m} + \varepsilon_{iT}^{c,m}$. We assume the error term associated with deterministic components of utility above is $\varepsilon_{\mathrm{it}}^{c,m}$ and is independent and identically distributed with the Type I Extreme Value. The choice probability for consumer i to choose alternative (c, m) at time t has a closed-form solution:

$$P_{it}^{c,m} = \frac{\exp(\bar{V}_{it}^{c,m})}{\sum_{c=0}^{C}\sum_{m=0}^{M}\exp(\bar{V}_{it}^{c',m'})} \quad (18)$$

where $\bar{V}_{it}^{c,m}$ is the deterministic part of the choice-specific value function, i.e. $\bar{V}_{it}^{c,m} = V_{it}^{c,m} - \varepsilon_{it}^{c,m}$. The corresponding log-likelihood function to be maximized is

$$\mathrm{LL} = \sum_{i=1}^{I}\sum_{t=1}^{T}[\sum_{c}\sum_{m} D_{it}^{c,m}\log(P_{it}^{c,m})] \quad (19)$$

To estimate the dynamic model, we follow the convention and fix the discount factor δ at 0.95, for all consumers. To handle the problem of a large state space, we adopt the interpolation method developed by Keane and Wolpin (1994) and calculate the value functions at a subset of the state space, and then use these values to estimate the coefficients of an interpolation regression to correct for such a problem. More specifically, we draw 100 state-space points and adopt a linear interpolation function of the state variables. Next, we use the interpolation regression function to provide values for the expected maxima at any other state points for which values are needed in the backward recursion solution process. We also assume the planning horizon is 35 quarters (≈8.75 years, 1.75 times longer than our observation period).

1.5 Results and Discussion

1.5.1 Model Comparison

In order to evaluate the importance of incorporating the dynamic add-on-to-base effect, we compare the data fitting performance of our proposed model with several alternative benchmark models. The first assumes a zero discount factor, no add-on-to-base effect, and homogeneous consumers. This is a myopic model in which homogenous consumers are assumed to make independent purchase decisions of base and add-on products to maximize current utility—consumers do not consider the inter-temporal dependence between these two products. The second model adds to the first by incorporating forward-looking consumers. Even though customers are allowed to take into account future trends of prices and quality, their purchases of base products and add-ons are assumed to be independent since this model does not recognize compatibility. The third benchmark adds the add-on-to-base effect but assumes it is a constant, similar to Sriram, Chintagunta and Agarwal's (2009) estimated model. It is important to note this model implicitly assumes that the add-on and base products are not required to be purchased simultaneously like that of Sriram et. al. for consumers to recover the additional benefit from memory. The fourth benchmark is the aforementioned model without heterogeneous consumers. The last model adds heterogeneous consumers and is our proposed model.

Table 4. Model Comparison

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Proposed Dynamic Model**		
					Two Seg.	Three Seg.	Four Seg.
-LL	8071.83	8207.75	8104.55	7718.28	6390.99	6209.38	6352.78
AIC	16182.63	16450.31	16243.61	15465.50	12858.68	12528.23	12859.04
BIC	16495.70	16722.62	16544.34	15781.75	13500.59	13507.49	14166.62

We estimate our proposed model with one to four segments. Comparisons of the BIC measures suggest the two-segment model is the most preferred whereas the AIC measures identify the three-segment model. For ease of interpretation, we pick the two-segment model as our preferred model and report its model performance in the following discussion. Table 4 presents the log-likelihood, AIC and BIC of the five alternative models. All of our dynamic models (Models 2-5) outperform the myopic model (Model 1). This implies there is an inherent dynamic process associated with the data generating process. Similarly, models recognizing the add-on-to-base effect (Models 3-5) outperform the ones that treat the purchase decisions of base and add-on products independently (Models 1 and 2). AIC and BIC further improve when we replace the add-on-to-base effect in Model 3 with cumulative inventory term of memory cards in Model 4. Such a result shows that a model taking into account all previously purchased memory cards better approximate the dynamic decision process than a model with a simple constant effect. Our proposed model is superior because it captures the dynamic impact of add-ons on the purchases of the base product: when making brand/standard choices of base products, a consumer takes into account the quantity (and quality) of add-ons for each standard to evaluate the stream of future consumption utilities net of future re-investment costs.

Table 5. Estimation Results

	Proposed Dynamic Model					
	One Segment		Two Segments			
			Seg.1 (91.7%)		Seg.2 (8.3%)	
Parameters	Est.	SE	Est.	SE	Est.	SE
Intercept: Sony (α^1)	-0.219	(0.091)	-0.334	(0.044)	3.069	(0.061)
Intercept: Oly (α^2)	-0.637	(0.145)	-0.357	(0.043)	0.588	(0.052)
Intercept: Fuji (α^3)	-1.044	(0.042)	-0.829	(0.080)	-0.294	(0.027)
Intercept: Kodak (α^4)	-0.528	(0.108)	-0.466	(0.068)	0.853	(0.014)
Intercept: Canon (α^5)	-0.645	(0.104)	-0.532	(0.065)	0.722	(0.026)
Intercept: HP (α^6)	-1.825	(0.094)	-1.843	(0.101)	0.281	(0.069)
Intercept: Nikon (α^7)	-1.623	(0.080)	-1.751	(0.915)	1.303	(0.071)
Intercept: Std1 (θ^1)	-2.116	(0.057)	-2.226	(0.241)	-0.156	(0.049)
Intercept: Std2 (θ^2)	-0.452	(0.141)	-0.414	(0.379)	0.444	(0.076)
Intercept: Std3 (θ^3)	5.188	(0.085)	-0.273	(0.253)	-0.035	(0.055)
Cquality (ϕ)	0.601	(0.083)	0. 462	(0.084)	1.182	(0.027)
Mquality (ψ)	0.127	(0.031)	0.189	(0.037)	0.739	(0.003)
A-to-B: Std1 (γ^1)	0.389	(0.051)	0.319	(0.061)	0.176	(0.011)
A-to-B: Std2 (γ^2)	0.152	(0.080)	0.154	(0.085)	0.069	(0.004)
A-to-B: Std3 (γ^3)	0.303	(0.075)	0.200	(0.040)	0.108	(0.007)
Price (λ)	-2.138	(0.007)	-2.198	(0.001)	-0.590	(0.001)
State Dep (κ)	0.044	(0.022)	0.032	(0.009)	0.051	(0.030)

In Table 5, we report the estimated coefficients for the proposed model. All the parameter estimates are statistically significant at the 0.05 level. The intercept terms represent consumer intrinsic preference for the seven brands of camera and three standards of memory cards. Comparison of these intercepts reflects the relative attractiveness of different brands within each category, after controlling for other factors included in the utility function. For example, consumers in segment 1 prefer Sony and Olympus, followed by Kodak, Canon, Fuji, Nikon, and HP in sequential order for cameras and Standard 3, Standard 2, and Standard 1 for memory cards. However, the preference order is Sony, Nikon, Kodak, Canon, Olympus, HP, and Fuji for cameras and Standard 2, Standard 3, and Standard 1 for memory cards for segment 2 consumers.

The coefficients of quality for camera and memory cards are positive for both segments, implying consumers care about the quality of the products. Not surprisingly, the coefficients of state dependence term are positive for both segments, which suggest consumers are more likely to purchase the same brand of camera as the one they have in hand. As expected, the price coefficient is estimated to be negative, showing consumers are price sensitive to the base and add-on products.

Comparing the estimates across the two segments, we find segment 1 consumers are defined by higher price sensitivity (-2.198 vs. -0.590) and low quality sensitivity (0.462 vs.

1.182 for camera and 0.189 vs. 0.739 for memory card). Consumers in segment 2 are characterized as being sensitive to quality but less sensitive to price. For the remainder of the discussion, we refer to the first segment as the price-sensitive segment and the second segment as the quality-sensitive segment. The price-sensitive segment constitutes the majority of the population (90.7%).[26] Interestingly, price-sensitive consumers are found to have higher add-on-to-base effects. This is not surprising because price-sensitive consumers are relatively more concerned about the future financial burden of purchasing memory cards. Thus, they value the memory cards in inventory more.

1.5.2 Add-on-to-Base Effect and Cross-category Dynamic Price Effect

Below we discuss in details the three inter-temporal trade-offs consumers face when making a purchase of a base and/or add-on product. We first highlight how future prices of add-ons moderate the dynamic add-on-to-base effect. Next, we discuss how consumer inventory levels of a given memory card standard lock consumers in to a specific camera standard due to the incompatibility of memory across camera standards. We finish with a discussion of dynamic price effects with the presentation of consumer price elasticities. We focus on the cross-category price effects as within prices effects are less germane to our analysis. In summary, we discuss how two new inter-temporal effects (i) a cross-category pricing effect and (ii) the cross-category inventory effect impact consumer purchase behavior in addition to the standard dynamic price effect.

---

[26] One may wonder that some segment could be primarily related with consumers who never replaced a camera. Given the two segments $k \in \{1,2\}$, we can calculate the probability of membership for consumer i belonging to segment i based on her purchase history $H_i = \{D_{it}^{c,m}\}_{t=1}^{T}$ (where $D_{it}^{c,m}$ is the choice indicator for any $c \in \{0,1,2,\dots,C\}$ and $m \in \{0,1,2,\dots,M\}$). Following Bayes' rule

$$P(i \in k|H_i) = \frac{L(H_i|k)f_k}{\sum_{k'} L(H_i|k')f_{k'}}$$

Where $L(H_i|k)$ is the likelihood of consumer i's purchase history H_i given she belongs to segment k and f_k is the relative size of segment k.

Replacement	Probability of Belonging to Segment 1	
	Mean	Std. Dev.
Yes	91.7%	0.0479
No	90.1%	0.0685

According to the table above, for consumers with replacement purchases, the probability of belonging to segment 1 is 91.7% (the probability of belonging to segment 2 is 1-91.7%=8.3%). For consumers without replacement purchases, the probability of belonging to segment 1 is slight less (90.1%). A two-sample t-test reveals a p-value of 0.0758. So we reject the hypothesis that there's significant difference of the segment membership between consumers who have camera replacement purchases and those who don't.

1.5.2.1 Dynamic Add-on-to-Base Effect and Interaction with Future Prices of Add-ons

Figure 6. Purchase Probability of Camera is Driven by the Expected Future Price and Current Inventory of Memory Card

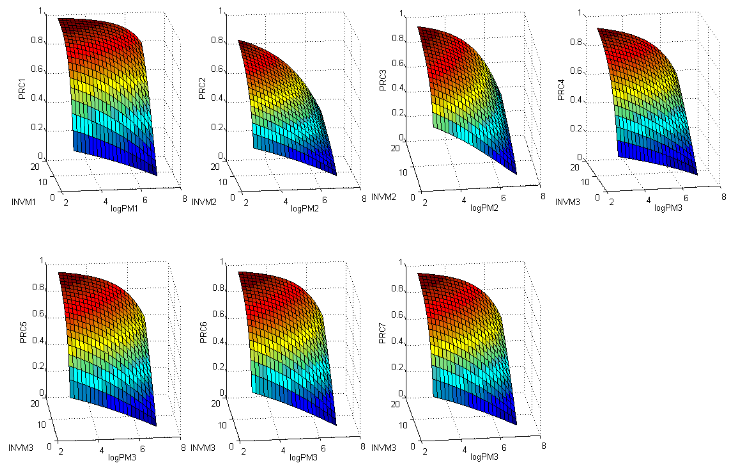


Figure 6 characterizes a consumer's decision rule describing how forward-looking consumers make a dynamic choice of base products based on current inventory and the expected future price sequence of compatible memory cards. The purchase probability of a compatible camera increases with inventory of compatible memory cards. This is because when planning her future purchase sequence, a consumer with higher inventory of memory cards, and thus extra storage, not only enjoys a long-term consumption utility stream, but also avoids a stream of future spending on new memory cards. This is the dynamic add-on-to-base effect captured by our model. Interestingly, the dynamic add-on-to-base effect is most prominent for Standard 1 (Sony's) and Standard 3 cameras in the sense that the purchase probability increases faster for the same amount of accumulation in memory card inventory. This is because when compared with those of Standard 2, Sony's memory cards offer a higher consumption utility stream while Standard 3 memory cards offer lower financial commitment. This implies that switching to an incompatible camera means not only incurring a purchase price, but also a loss of long-term consumption utility as well as a future of purchasing additional memory cards of another standard.

Figure 6 also presents how a current purchase decision of a camera is driven by the future price trend of compatible memory cards. As expected, for all brands, when the expected future price of a memory card decreases, the purchase probability of the compatible camera increases because the financial commitment related to the planned purchase sequence for owning a composite of camera and memory card(s) is lower compared with other pairs.

It is interesting to discuss how the future price expectations interact with the aggregate dynamic add-on-to-base effect. Now although the add-on-to-base effect does not explicitly account for the price of the memory cards, it does indirectly through a consumer's accumulation. The model determines and we illustrate in Figure 6 that the add-on-to-base effect becomes more prominent when consumers expect future prices of compatible memory cards to be lower. This is because when expecting lower future prices it increases consumers' likelihood of purchasing new memory cards, leading to a larger add-on-to-base effect in the future and thus making consumers even more likely to purchase compatible cameras. To summarize, lower future prices of memory cards can enhance the dynamic add-on-to-base effect for compatible cameras.

1.5.2.2 Quantify Purchase "Lock-In" due to Compatibility

Our dynamic model allows us to quantify the cost of switching associated with the purchase of a camera which is incompatible with a consumer's current inventory of memory cards. We define the cost of switching to be the minimum lump-sum payment needed for a manufacturer to induce a consumer to switch to its brand of camera. With the consumer being forward-looking, this cost of switching measures the difference between the total discounted values of two streams of utilities associated with the purchasing of two different cameras. More specifically, it is the difference between the continuation value of purchasing a compatible camera and the continuation value of switching to an incompatible brand divided by the price sensitivity coefficient. We divide by the price coefficient in order to convert the measure into dollars. Given this definition, the cost of switching is time and state-dependent. Thus, we arbitrarily select period ten to calculate the monetary equivalent of switching under the scenario of a representative consumer who has only one compatible memory card in inventory during this period.

Table 6. Cost of switching

Average	Sony	Olympus/Fuji	Kodak/Canon/HP/Nikon
Sony	$0.000	$23.595	$21.590
Olympus/Fuji	$8.873	$0.000	$8.029
Kodak/Canon/HP/Nikon	$15.495	$17.629	$0.000
Segment1	Sony	Oly/Fuji	Kodak/Canon/HP/Nikon
Sony	$0.000	$24.629	$23.018
Olympus /Fuji	$9.312	$0.000	$8.684
Kodak/Canon/HP/Nikon	$16.443	$18.707	$0.000
Segment2	Sony	Oly/Fuji	Kodak/Canon/HP/Nikon
Sony	$0.000	$10.143	$10.272
Olympus /Fuji	$4.703	$0.000	$4.511
Kodak/Canon/HP/Nikon	$7.222	$8.121	$0.000

We report the cost of switching for the two segments as well as for the three brand groups in Table 6. On average, Olympus or Fujifilm need to offer a $23.595 discount and

Kodak/Canon/HP/Nikon need to offer $21.590 to induce consumers to switch from Sony. However, Sony only has to offer $8.873 to steal a consumer from Olympus/Fujifilm and $15.495 to induce brand switching from Kodak/Canon/HP/Nikon. When comparing each consumer segment's cost of switching, the cost is highest among consumers in segment 1 (price-sensitive consumers) than consumers in segment 2 (quality-sensitive consumers). These price-sensitive consumers also have larger add-on-to-base effects and thus have the most utility to lose by eliminating their current memory card inventory when switching standards.

From the above comparison, we see Sony (the first row) has the highest cost of switching. For consumers who hold the same amount of memory cards in inventory, it is more costly to attract consumers from Sony to other brands than vice versa. Thus, Sony enjoys the highest rate of "lock-in" or loyalty partly because of its incompatibility with rival products. This can also be explained by the higher dynamic add-on-to-base effect moderated by consumer price expectations: conditional on having the same amount of memory cards on hand, Sony owners enjoy higher total discounted future utility from purchasing a compatible camera than purchasing a non-compatible camera (the coefficient of add-on-to-base effect is highest for Sony). However, this is partially mitigated by the higher expected future financial commitment in purchasing new memory cards because we have shown that higher future prices lower the total dynamic add-on-to-base effect. Consequently, the introduction of the Memory Stick assists Sony in building a strong brand loyalty because consumers are tied to the standard by a high cost of switching. When product replacement becomes more frequent as product quality improves over time, such a lock-in effect creates continuous sales for Sony.

The comparison of costs of switching also indicates that it takes nearly double the amount of discount to incentivize consumers to switch from Standard 3 cameras than from Standard 2 cameras. This is not only because Standard 3 cameras have a higher add-on-to-base effect, but also because the future prices of Standard 3 memory cards are lower than those of Standard 2. This enhances the dynamic add-on-to-base effect and competitiveness of Standard 3 cameras.

The total cost of switching comprises three effects, 1) Price/Quality difference between brands, 2) State dependence and 3) Add-on-to-base effect. Decomposition of the total cost of switching is required to know the relative contribution of each effect.[27]

Table 7[28]. Cost of switching Decomposition

	Cost of switching	Price/Quality	State dependence	Add-on-to-base effect

[27] Please find the detailed procedure in Appendix 5.

[28] Note that switching cost derived from state dependence is the same across brands. This is because in the utility function, the state dependence parameter SD_i is not brand specific as there's not enough replacement purchases to identify different SD_i's for different brands.

Olympus	$23.129	$-3.620	$10.358	$16.391
Fujifilm	$21.587	$-5.162	$10.358	$16.391
Kodak	$20.648	$-4.754	$10.358	$15.044
Canon	$23.553	$-1.849	$10.358	$15.044
HP	$18.964	$-6.438	$10.358	$15.044
Nikon	$20.399	$-5.003	$10.358	$15.044

Table 7 above shows the result of decomposition. We find that given Sony has a relatively low quality (adjusted by price) compared to Fujifilm[29], it is required to compensate consumers with 5.162 dollars to stay with Sony when consumers lack loyalty (state dependence) to the camera brand or have no add-on-to-base effect. State dependence accounts for a relative 38.7% (10.358/(10.358+16.391)) of the cost of switching while add-on-to-base effect accounts for the other 61.3%(1-38.7%).

For all brands, the add-on-to-base effect is the main source of the cost of switching. Interestingly we find that the cost of switching from add-on-to-base effect is higher for cameras compatible with the standard 3 memory card. This reinforces our estimate of a higher add-on-to-base effect coefficient for standard 3 than that for standard 2 (Table 5).

1.5.2.3 Inter-temporal Price Tradeoff—Price Elasticity

Unlike those in the existing literature, our model is built at the brand and standard choice level, allowing us to examine how price affects brand or standard switching decisions. In addition, our model takes into account the inter-temporal dependence of base and add-on products. In Table 8, we report the percentage changes in sales when the price increases by 10% for both camera brands and memory card standards. There are many notable results; however, we focus on the most interesting ones related to cross-category elasticities.

Table 8. Price Elasticities

	SonC	**OlyC**	**FujC**	**KodC**	**CanC**	**HPC**	**NikC**	**M1**	**M2**	**M3**
SonC	-1.047	0.353	0.196	0.578	0.215	0.181	0.311	-1.252	0.049	0.606
OlyC	0.443	-2.774	0.005	0.474	0.281	0.018	0.274	0.242	-1.645	0.664
FujC	0.419	0.489	-14.062	0.807	-0.067	0.411	0.570	0.183	-7.058	1.143
KodC	0.702	0.191	0.091	-1.501	0.175	0.188	0.068	0.234	0.213	-1.143
CanC	0.306	0.270	0.123	0.665	-3.560	0.244	0.238	0.679	0.369	-0.718
HPC	0.229	0.498	0.316	0.209	0.094	-4.408	-0.016	0.394	0.529	-1.015
NikC	0.138	0.455	0.043	0.353	0.121	0.134	-3.568	0.379	0.151	-1.786
OutC	0.015	0.093	0.039	0.062	-0.195	-0.050	0.001	-0.073	0.139	-0.007
M1	-1.267	0.272	0.304	0.508	0.310	0.277	0.068	-3.256	0.318	0.621

29 For example, Sony's famous Mavica models have generally higher prices and lower resolutions than competing models (http://www.imaging-resource.com/PRODS/CD1K/CD1KFLP.HTM)

M2	0.596	-2.822	-0.581	0.245	0.200	0.128	0.253	0.163	-4.534	0.776
M3	0.237	0.368	0.146	-0.697	-0.085	-0.198	-0.467	0.451	0.402	-2.428
OutM	-0.129	-0.143	-0.045	0.014	0.003	-0.067	0.037	0.064	0.014	0.135

First, it is interesting to note that own-category price effect dominates cross-category price effect for all brands with the exception of Sony. For instance, the purchase probability of the Sony camera decreases by 12.52% when the price of Sony memory increases by 10% but only by 10.47% when the price of the Sony camera increase by 10%. In other words, the change of purchase probability for the Sony camera decreases more when the price of the compatible Standard 1 memory card decreases than when its own price decreases. This is because the high price charged by Sony for its memory card prevents consumers from purchasing more memory cards; eroding the dynamic add-on-to-base effect to a point that consumers become highly sensitized to the price of memory cards.

Furthermore, when examining the cross-category elasticities listed in the last three columns, we find that when the price of a Standard 1 or 2 memory cards increase, most sales transfer to Standard 3 cameras. For example, when Sony increases the price of its memory card, the sales of Standard 3 cameras (Canon, HP and Kodak) increase more than those of Olympus and Fuji. Similarly, when the price of a Standard 2 memory card increases by 10% the sales of Standard 3 cameras also increase more than Sony. Consequently, higher memory card prices drive consumers to a more open standard in which more cameras can share the same memory card. It is also important to note that competition among camera brands is most fierce within standards.

1.6 Counterfactual Simulations

In order to address the impact of several important research questions pertaining to pricing and compatibility, we employ the above estimated model primitives in four counterfactual simulations. The first analyzes whether a firm can improve its market position by adopting a pricing strategy that better leverages its add-on-to-base effect. In the second simulation, we attempt to understand how the market changes when all compatibility constraints are eliminated. Consequently, what role does incompatibility play on market share? Thirdly, how does the market change when an inferior standard becomes compatible with a superior standard via an adaptor? Lastly, is incompatibility or a closed system beneficial for all firms? Specifically, how does brand equity moderate the effects of incompatibility on market share? It is important to highlight the fact the below simulations only recover partial equilibrium effects. We do not fully account for changes in product quality or rival firms responding to changes in compatibility across standards. The results are therefore partial equilibrium effects. A full equilibrium model wherein prices, quality and compatibility are endogenously determined is beyond the scope of this paper. In particular, such a general approach would require modeling how each firm's decision would impact its rivals' decisions--a nontrivial exercise when including structural demand side estimates.

1.6.1 Alternative Dynamic Pricing Strategies

As we mention in the above data description section, Olympus and Fujifilm employed a pricing strategy for cameras that first set a high price in years one and two (2000 and 2001) and then was lowered. We determine that the consequence of such a pricing scheme is that consumers delay purchase, leading to lower sales and hence fewer inventory of memory cards during the first two years. Moreover, this pricing strategy did not help these brands harvest their corresponding add-on-to-base effect. An opposite strategy would be to introduce its camera with a low price to attract consumers to purchase both its camera and memory cards and then to exploit the high add-on-to-base effect from consumers' high inventory of memory cards by increasing price in the future.

To determine the impact of this alternative pricing strategy, we allow the prices of Olympus and Fujifilm to fall by 10% each quarter during the first two years and then increase by 1% each quarter from 2002 to 2004. Comparing the second and third columns of**Error! Reference source not found.**Table 9 illustrates that under the new pricing scheme, the market share of Olympus and Fujifilm cameras would increase by 5.15 percentage points and 0.91 percentage points, respectively. Correspondingly, the overall market share of Standard 2 memory cards would rise by 2.33 percentage points.

We determine that Olympus's initial low price triggers consumers to adopt the camera early and enjoy the associated stream of utility from the camera and memory cards in future periods. With Standard 2 memory cards in hand, consumers are also more willing to buy Olympus cameras in later periods. In summary, this new pricing policy is more consistent with penetration pricing and product line pricing of complementary products, where lower initial camera prices boost camera and hence memory card sales. This increase in memory card sales generates higher consumer dynamic add-on-to-base effect and thus locks in consumers to purchase compatible cameras in future periods.

1.6.2 Compatibility

To investigate the implication of compatibility, we carry out a simulation wherein we estimate average choice probabilities of cameras and memory cards of different standards under the assumption that all cameras and memory card standards are compatible. For instance, a previously purchased Sony Memory Stick can be used on any newly purchased cameras from Olympus, Fujifilm, Kodak, Canon, HP, and Nikon in addition to Sony. Thus, all memory cards in inventory will exert the add-on-to-base effect to the purchased camera, though in various magnitudes determined by the coefficient of add-on-to-base-effect. To approximate this scenario, we set the standard-specific add-on-to-base effect to be the sum of inventory of all memory cards as if no compatibility constraints exist across standards.

Table 9. Policy Simulations

	Market Share of Cameras						
	Benchmark	Change Pricing		No Compatibility		Adapter	
Son	30.47%	29.00%	-4.84%	21.29%	-30.15%	25.54%	-16.18%
Oly	15.60%	20.75%	33.01%	15.85%	1.58%	14.78%	-5.28%
Fuji	5.68%	7.59%	33.50%	4.80%	-15.49%	4.79%	-15.65%
Kod	21.60%	21.50%	-0.48%	25.13%	16.33%	25.59%	18.47%
Can	11.14%	9.18%	-17.56%	11.94%	7.16%	12.76%	14.55%
HP	8.25%	7.33%	-11.19%	9.82%	19.01%	8.65%	4.87%
Nik	7.25%	4.66%	-35.79%	11.18%	54.15%	7.88%	8.68%
	Market Share of Memory Cards						
	Benchmark	**Change Pricing**		**No Compatibility**		**Adapter**	
Std1	31.00%	29.85%	-3.72%	25.25%	-18.56%	27.88%	-10.09%
Std2	20.93%	23.26%	11.13%	18.70%	-10.68%	20.03%	-4.32%
Std3	48.06%	46.89%	-2.45%	56.05%	16.63%	52.09%	8.39%

In column 5 and 6 of Table 9, we compare the purchase probabilities with those generated by the counterfactual simulation; from this we can understand the extent to which compatibility changes purchase probabilities of base products. The results suggest if Sony Memory Stick was compatible with the products of all its competitors, its camera market share would have dropped by 9.16 percentage points (from 30.47 percentage points to 21.29 percentage points) and by roughly 5.75 percentage points (from 31.00% to 25.25%) for memory. This occurs because consumers are no longer locked-in by the Memory Stick—consumers who own a Sony can purchase cheaper memory cards from its rivals. Without the compatibility constraint, consumers are free to choose whatever brand of new camera they like for their next purchase, which undermines Sony's brand equity, or brand synergy effect.

On the other hand, the market shares for Kodak, Canon, HP, and Nikon jump by 9.83 percentage points (from 48.24 percentage points to 58.07 percentage points), and the share for Standard 3 memory cards increases by 7.99 percentage points (from 48.06 percentage points to 56.05 percentage points). However, removing the incompatibility across standards has a marginal impact on the market shares of camera and memory cards of Standard 2 because of its relatively small add-on-to-base effect.

1.6.3 Partial Compatibility

The above simulation shows that Sony's proprietary standard of memory card (Standard 1) exerts strong pressure on the market share of Standard 3 memory cards. One defending strategy for Standard 3 might be to create an adapter that allows its compatible cameras to read Standard 1 cards. By this means, Kodak, Canon, HP, and Nikon can break down the lock-in effect of Sony's memory card, thus making their cameras more attractive. In order to determine the effectiveness of this strategy, we allow all cameras that are compatible with

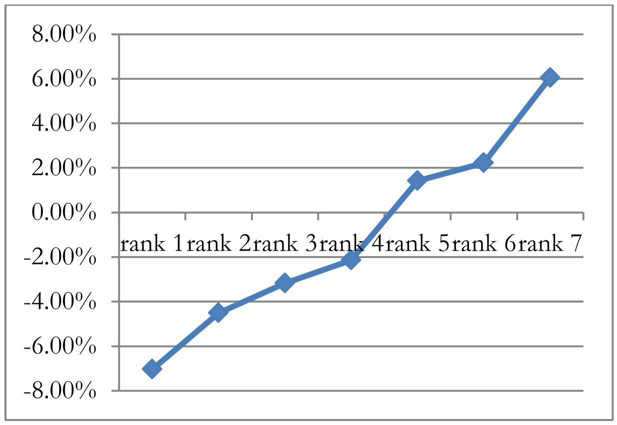
Standard 3 memory cards, i.e. Kodak, Canon, HP and Nikon, to be compatible with Standard 1 memory cards. Therefore, we increase the size of the choice set from 18 to 22 by adding four new choice alternatives, $\{i, j\} = \{(4,1), (5,1), (6,1), (7,1)\}$, because under this situation Kodak (c=4), Canon (c=5), HP (c=6), and Nikon (c=7) can use the Sony Memory Stick (m=1). Moreover, the add-on-to-base effect term is modified accordingly because Sony, Kodak, Canon, HP, and Nikon are now all compatible with a Standard 1 memory card. So in the occasion of purchasing any of these five cameras, a Standard 1 memory card in inventory will contribute to utility through the add-on-to-base effect term.

We show that all Standard 3 cameras can steal market share from Sony cameras by offering partial compatibility. For example, Kodak can increase its market share by 3.99 percentage points and Canon can increase sales by 1.62 percentage points. This is because partial compatibility enables consumers to switch more easily to standard 3 cameras due to the smaller cost of switching or lock-in effect. Similarly, Sony memory cards can be used with a third group of cameras, thus avoiding Sony's add-on-to-base effect. Consequently, the market share of Olympus and Fujifilm is smaller because of the added choice alternatives leading to more fierce competition in the market.

1.6.4 Incompatibility and Brand Equity

Recall in section 1.5.2 that Sony has the largest brand preference, or the strongest brand equity in the camera market. Such strong brand equity lays the foundation for its success. But what if this was not the case? What if its brand equity were not as strong? Would the aid of the add-on-to-base effect stemming from incompatibility be marginalized and thus have less influence on the market for base products? We find it necessary to examine how brand equity moderates the effects of incompatibility in order to answer these questions. We run a series of policy simulations where Sony's brand-specific intercept is set to that of the brand that ranks 2nd to 7th in the market. We compare the market share of Sony before and after eliminating incompatibility between memory cards and cameras (as done in section 1.6.1). Figure 7 depicts how the effect of incompatibility varies with Sony's brand equity rank. As we can see, when Sony had the strongest brand equity, creating compatibility with other standards had a significant impact on its market share — a decrease of 8.7 percentage points. This effect of compatibility diminishes as Sony's brand equity advantage vanishes (from rank 1 to rank 4). Strikingly, Sony's market share increases if it creates an open memory card format when Sony's brand equity falls below the industry average (rank 5 to rank 7). In other words, a market follower should not set up a compatibility constraint to bind itself (Katz and Shapiro (1985)). This policy simulation can rationalize a well-known case of Betamax. In 1975, Sony introduced the Betamax video standard and a year later JVC launched the competing standard VHS. For around a decade the two standards battled for dominance, with VHS eventually emerging as the winner. Why did Sony lose the video tape standard war but win the memory card standard war later? One possible reason is that Sony did not have as much relative brand equity in the VCR market as it did in digital cameras. Technically, (due to its solution to the recording head drum miniaturization,) Sony made its Beta camcorders only to record while VHS camcorders could review footage in the camcorder and copy to another VCR for editing. With this limitation, Sony's Betamax failed even though it was the market pioneer and tried to take advantage of the lock-in effect of video tape format.

Figure 7. Sony's Market Share Loss of Eliminating Incompatibility at Different Brand Equity Ranks



1.7 Conclusions and Future Research

High-technology durable products often comprise base products and add-ons. When making purchase decisions, forward-looking consumers take into account price, quality, and compatibility and make joint inter-temporal decisions. We develop a framework in which forward-looking consumers make joint choice decisions regarding the base and add-on products when multiple incompatible standards exist. We model consumers' repeated choices at brand and standard level given compatibility constraint. Compatibility makes the purchase behavior of two categories dynamically interdependent because when choosing which base product to buy, a consumer has to take into account the effect of forgoing future consumption utilities and incurring future financial costs for the add-ons if she switches standards. This novel model feature enables us to calibrate cross-brand, cross-standard, and cross-category price elasticity and compare the relative magnitude of each. Once given these elasticities, we further examined consumers' switching propensity in brand and standard, as well as interdependence across categories. Our results enrich the current literature by further probing competition at the standard and category level.

We found that when making a purchase decision for the base product, consumers take into account future prices of the add-on product because the financial commitment is related to the planned purchase sequence of both categories. Moreover, consumers are locked-in to the base product brand by the dynamic add-on-to-base effect which becomes stronger with greater inventory levels of add-ons. Furthermore, the dynamic add-on-to-base effect can be enhanced by lower future prices of add-ons. These interesting consumer behaviors have important firm strategy implications. We found that among three standards, Sony's Memory Stick enjoys the highest add-on-to-base effect, which further leads to highest cost of switching and greatest lock-in effect. Following this, we demonstrated that Sony gained profits from developing its proprietary standard of memory card (the Memory Stick). We also found such a strategy might not be as profitable for a manufacturer with lower brand equity. As to pricing strategies, we showed that if Standard 2 drops its initial price of memory cards, consumers will be triggered to adopt the camera early and the market size of Olympus will be expanded.

Insights from this stream of research will offer managers more comprehensive product strategies. For example, managers can employ pricing and promotion strategies for add-ons to improve base product market performance by taking advantage of the cross-category price effect and cross-category dynamic inventory effect. A cheaper price of add-ons in the early period of new product introduction may encourage adoption and lock consumers in. On the other hand, market leaders may consider designing exclusive add-ons, which can lead to greater market share of the base product. Followers though should elect to either be compatible with the leading brand or create a union with other players in the market to diminish the market power of the leading brand. Furthermore, pricing or promotion strategies of add-ons should be targeted heavily at price-sensitive consumers than quality-sensitive consumers.

Our research is subject to limitations that open areas for future research. First, with lacking product attributes in our data, we can't estimate intrinsic preference for various models of cameras and memory cards in a more refined fashion. Future works can further examine whether the documented add-on-to-base effect is more prominent for a high-end product or low-end product. Second, in high-technology product market with frequent innovations, consumer brand preferences might evolve over time. Researchers in the future might want to model ever-changing consumer intrinsic brand preference to better capture the demand dynamic. Third, the current paper assumes price and quality are exogenously given. A very interesting topic to explore is how firms design the full product line by deciding price trajectories for both base products and add-ons taking consumers' dynamic decision-making processes into consideration. A full equilibrium model is needed to solve this problem from both sides of supply and demand. Fourth, Gabaix and Laibson (2006) reveal very interesting phenomena regarding base product and add-ons where firms shroud information about add-ons to consumers. Only sophisticated consumers take advantage of the firm that shrouds information by avoiding add-on purchases; the unsophisticated fall into the trap of high add-on prices. Our paper supports the decision-making process of sophisticated consumers with evidence of their consideration of base products and add-ons at the same time. Future research can modify our model to allow only part of the consumers to be forward-looking with the rest short-sighted. Fourth, we keep other firm strategies, for example product design, pricing, cost structure, exogenous. But in reality, making add-on products compatible with base products involves engineering design, which will affect other firm decisions as well.

2 Chapter 2
A Structured Analysis of Unstructured Big Data Leveraging Cloud Computing

Xiao Liu, Param Vir Singh, Kannan Srinivasan

Abstract

Accurate forecasting of sales/consumption is particularly important for marketing because such information can be used to fine-tune marketing budget allocations and overall marketing strategies. In recent years, online social platforms have produced an unparalleled amount of data on consumer behavior. However, two challenges have limited the use of such data to obtain meaningful business insights in marketing. First, the data are typically in an unstructured format such as text, images, audio, or video. Second, the sheer volume of data makes standard analysis procedures computationally unworkable. In this study, we combine methods from cloud computing, machine learning and text mining to illustrate how content from social media such as Twitter can be effectively used for forecasting purposes. We conduct our analysis on a staggering volume of nearly two billion Tweets and 400 billion Wikipedia pages. Our main findings highlight that, in contrast to basic surface-level measures such as volume of Tweets or sentiment in Tweets, the information content of Tweets and their timeliness improve forecasting accuracy significantly. Our method endogenously summarizes the information contained in Tweets. The advantage of our method is that the classification of the Tweets is based on what is in the Tweets rather than preconceived topics that may not be relevant. We also find that in contrast to Twitter, other online search data (e.g., Google Trends, Wikipedia views, IMDB reviews or Huffington Post news) are very weak predictors of TV show demand because users Tweet about TV shows before, during and after a show, while Google searches typically lag the show.

2.1 Introduction

The average American watches 5.1 hours of television per day, which is more than the 4.6 self-reported daily hours spent on all work-related, educational and housework activities.[30] This statistic explains why TV is the largest ad spending medium in the US, and it remains relatively stable despite the spectacular growth of online advertising.

As technology has advanced, the scheduling of TV advertisements has become more efficient. Accurate forecasts of television ratings are critical for a number of reasons. First, depending on the forecasts, networks often adjust the number of new shows for each serial. Second, the pricing of advertising for shows can be made more dynamic and near real time. Current online advertising on content-based sites is realized in as little as 0.35 seconds based on dynamic real-time bidding models, and the spillover to traditional television advertising is beginning to occur. Thus, better forecasts of the size of a viewing audience will enhance ability to conduct price auctions, and advertisers can decide whether to participate and how much to bid. Third, depending on projected ratings, firms can fine-tune a number of (endogenous) actions such as advertisements for shows, paid blogs and Tweets to affect the ratings. Therefore, advertisers and broadcasting companies are both eager to accurately predict the ratings of TV shows.

A recent article in the New York Times (March 8, 2015) underscores the industry significance of the issue studied in our paper, "So far, however, Twitter and Nielsen have avoided the most important question posed by marketers and the TV industry: Exactly how much does chatter on Twitter lift the viewership of a particular show? Although Nielsen published data on the Twitter activity surrounding a show's broadcast as a complement to its more familiar TV ratings, it has said little about the relationship between the two." The issue we examine in this paper is of paramount importance to industry. It further appears that the companies do not have seemed to have a definite answer.

Currently, consumers use various online platforms to enhance their TV watching experience: they look for show-related information on search engines such as Google and share their watching experience with friends on social networks such as Twitter or Facebook. These footprints on online platforms can be very helpful for advertisers to forecast demand. However, the content on these online platforms produces two challenges: the data produced are in an unstructured form (for example, text, video, audio, or images), and the sheer volume of data makes standard data analysis procedures computationally unworkable or inefficient. The existing literature has attempted to incorporate user-generated content into their analyses by incorporating easy-to-calculate measures such as the volume or valence of relevant user-generated content on online platforms. Recent studies have attempted to dig deeper into user-generated content (e.g., Gopinath et al. 2014, Pauwels et al. 2013). These studies either depend primarily on manual coders to classify the user-generated content, thus limiting the scale of the application, or they follow a supervised learning approach for

[30] Nielsen (2011), BLS American Time Use Survey (2011).

classifying the user-generated content. However, all of these studies are limited to the extent that the user-generated content is typically classified into preconceived labels (such as sentiment, recommendation-oriented or emotion). We believe that the information contained in user-generated content cannot be captured by such simple measures. Rather, we show that the information contained in the textual content can provide much richer insight into user behavior and decision making, and an unsupervised learning approach can provide significant improvement in forecasting performance. We combine methods from machine learning and text mining to dig deeper into textual content, and we endogenously identify distinct streams of information contained in user-generated content. We utilize tools and methods from cloud computing such as Hadoop MapReduce to handle millions of text documents. In doing so, we confront a problem that is germane to text mining. We find that the number of distinct streams of information identified is typically greater by an order of magnitude than the number of observations on the variable of interest. To address this concern, we conduct a massive dimension reduction by employing the Apache Mahout machine learning library to summarize the enormous information in a few principal components. We find that these principle components have excellent predictive power in demand forecasting. None of the tasks described above is trivial when the volume of data is "giganormous". The memory space and computing capacity of a single workstation cannot handle data at our scale. Instead, we use Amazon Web Services to perform the cloud computing tasks and only pay with a minimal budget.

We use unstructured data of consumer behaviors on online platforms, including Twitter and Google search, to predict consumers' offline TV viewing behavior for thirty TV series as well as primetime National Football League (NFL) games. We argue that consumers reveal their interests in TV programs through online platforms before actually watching TV. For example, a Twitter user's post, "I am going to watch Breaking Bad tonight," is a direct indication of her future TV watching intent for a specific show. Or, if a user searches on Google for an NFL game before it starts, it is quite likely that he is going to watch it on TV. Therefore, by collecting publicly available Twitter Tweets and Google Trends data at negligible cost, marketers and advertisers can leverage consumer-generated data to accurately forecast future demand rather than purely relying on the historical information from the Nielsen Rating data.

To achieve this goal, we use a large dataset derived from five sources of social media: 1) Twitter: 1.8 *billion* Tweets for five years from 2008 to 2013; 2) Google Trends[1]: 113.3 *million* Google searches[2] (when combined with the Google AdWords keyword volume service, we are able to obtain the real search volume); 3) Wikipedia views: 433.6 *billion* Wikipedia page views; 4) IMDB reviews: 4.3 thousand reviews; and 5) Huffington Post news: 5.5 *million* articles. We find that the predictive power of surface-level measures of user-generated

[1] Google Trends is a public web facility of Google Inc. that is based on Google Search and that shows how often a particular search term is entered relative to total search volume across various regions of the world and in various languages.
[2] Google Trends data is structured in a numerical format. The other four sources of data come in text format.

content such as volume of Google searches (or Wikipedia views) or the volume and valence of Tweets (or IMDB reviews and Huffington Post news) is not as strong as the historical data to forecast TV ratings. However, refined information in Tweets exhibits stronger power to predict TV ratings than simply learning from the past.

We conduct a rigorous, structured econometrics analysis of the processed unstructured data. Our results show that Tweets and Google search volume have a positive impact on TV ratings. The impact of the valance of Tweets on ratings is not statistically significant. Twitter Tweets and Google searches are substitutes rather than complements. Carefully summarized Tweet content that indicates future action has the highest predictive power.

Our paper provides two key contributions. First, from a managerial standpoint, we show that easily accessible public information from online platforms can be used to predict TV ratings. However, surface-level information such as volume and valance is not more useful than historical data; only sophisticated content analysis can achieve high prediction accuracy. The method we propose has a distinct advantage in that it does not require one to classify content into pre-conceived topics that may not be relevant. Instead, it endogenously summarizes the information in Tweets into topics. That is, it classifies the relevant Twitter content into distinct streams of information that consumers are talking about. Second, we introduce state-of-the-art big data processing techniques through a cloud-based distributed computing framework called Hadoop MapReduce that we demonstrate with Amazon Web Service tools. We hope marketing researchers can put these methods into practice to conduct more structured research on large-scale unstructured data.

2.2 Literature Review

Our paper draws on three streams of literature: (1) using Twitter data for predictions, (2) the effect of online user-generated content on product sales and (3) text mining.

2.2.1 Social Media Predictions

Research has suggested that Twitter feeds are early indicators of various economic and social phenomena, such as book sales (Gruhl et. al. 2005), movie box office sales (Mishne and Lance 2006), opinion polls (O'Connor et. al. 2010), elections (Tumasjan et. al. 2010), the spread of contagious diseases (Paul and Dredze 2011), stock market performance (Bollen, Mao and Zeng 2011) and NFL game outcomes (Sinha, Dyer, Gimpel and Smith 2013). The predictive power of Twitter is derived from the information embedded in consumers' conversations. We follow this literature to predict the ratings of popular TV shows in the US market. Beyond Twitter, research has also investigated other social media data for predictions, such as Google searches (Preis et al. 2010) and Wikipedia views (Mestyán, M., Yasseri, T., & Kertész, J. 2013). These studies typically capture information in Tweets by volume, valence or emotion. Furthermore, most papers in this area use machine learning methods with the objective of merely minimizing the prediction error. We instead use an econometrics model that corrects for the Nickell bias (1981) to perform a more structured analysis (in the sense of providing economic explanations) of the unstructured data from multiple social media sources.

2.2.2 The Effect of Online UGC on Sales

Our paper is also closely related to the literature on the effect of user-generated content on demand/sales. As listed in Table 10, numerous studies have examined this topic in both marketing and computer science for either explanatory or prediction purposes. In these papers, researchers have used various forms of user-generated content, including online reviews, online ratings, and blogs, to investigate their impact on the demand for the focal products. We instead use two more popular online platforms, Twitter and Google, to collect UGC. The advantage of our approach is that these two platforms have a much wider user base, and therefore the predicted demand from using information from these two platforms is more likely to represent the emerging big data information sources.

Additionally, as Table 10 demonstrates, metrics such as the volume, valence and variance of UGC have been examined. However, the rich content information in text data has been left underexploited. In fact, all but three of the prior studies (Onishi and Manchanda 2012, Gopinath, Thomas and Krishnamurthi (2014) and Pauwels, Stacey and Lackmann (MSI, 2013)) tried to perform text mining beyond basic sentiment analysis. Our paper extends this stream of literature, with two major distinctions. First, we incorporate cloud-based, large-scale text mining to extract useful information from a vast amount of data whose size is larger than that in the previous literature by a magnitude of 1000. Second, we exploit unsupervised learning techniques to let the data speak for itself rather than imposing any label on the features (for example, Pauwels, Stacey and Lackmann (MSI, 2013) selected conversations related to "went there/purchased," and Gopinath, Thomas and Krishnamurthi (2014) classified tweets as action/emotion related). Instead we mine the data with an unsupervised learning, i.e. we adopt dimensionality reduction method (specifically, principal component analysis). By studying the loading of specific content, we can interpret the key principal components. Thus, our approach is consistent with the traditional approach in marketing where the dimensionality reduction is first undertaken (such as factor analysis and principal component analysis) and then the factors are interpreted.

Table 10 Overview of Literature on UGC

Author	Year	Product	UGC	Measure	Effect	Outcome measures	Text Data	Data size	Text mining tools
Godes and Mayzlin	2004	TV shows	Online review	Volume	~	Household rating	Yes	21,604	Independent raters
				Valence	~				
				Variance	+				
Chevalier and Mayzlin	2006	Books	Online rating	Volume	+	Sales rank	No		
				Valence	+				
Liu	2006	Movies	Online review	Volume	+	Box office revenue	Yes	12,136	Independent raters
				Valence	~				
Mishne and Glance	2006	Movies	Weblog	Volume	+	Sales	Yes	Unknown	Keyword detection
				Valence (Sentiment)	+				
Liu et al.	2007	Movies	Weblog	Sentiment		Box office revenue	Yes	45,046	Machine learning
Dhar and Chang	2009	Music	Online review, blog, SNS intensity	Volume	+	Sales rank	No		
				Rating	+				
				Social Network intensity	~				
Sadikov et al.	2009	Movies	Blog	Volume	+	Box office revenue	Yes	Unknown	Machine learning
				Sentiment	~				
Chintagunta et al.	2010	Movies	Online rating	Valence	+	Opening day revenue	No		
Chen et al.	2011	Digital cameras	Online rating	Valence	+	Sales rank	No		
Karniouchina	2011	Movies	Yahoo! movie site	Internet searches/Review count	+	Box office revenue	No		
Chakravarty, Liu, and Mazumdar	2011	Movies	Online review	Volume		Box office revenue	Yes		Independent raters
Moe and Trusov	2011	Multiple products	Online rating	Valence	+	Ratings and Sales	No		
Onishi and Manchanda	2012	Movies, cell phone subscriptions	Blog	Volume	+	Sales volume	Yes	200,000	Wordcount+ selected words
				Valence	+				
				(text mining)	+				
Stephen and Galak	2012	Loans	Press and blog	Volume	+	Sales	No	2012	Loans
Dewan and Ramaprasad	2012	Music	Blog	Volume	+	Sampling	No		
Tirunillai and Tellis	2012	Multiple	Product reviews	Volume	+	Stock prices	Yes	347,628	Machine learning
				Valence	+				
Gopinath, Chintagunta and Venkataraman	2013	Movies	Blogs	Volume	+	Box office revenue	Yes	unknown	Independent raters
				Valence	+				
Gopinath, Thomas and Krishnamurthi	2014	Cellular phones	Forum	Volume	~	Sales	Yes	unknown	Independent coders
				Valence	+				
				Content	+				
Pauwels, Stacey and Lackmann	2013	Clothing retailer	Blog, forum, Facebook, and Twitter	Volume	+	Store/web traffic	Yes	428,450	Machine learning
				Valence	+				
				Content	+				
This paper	2015	TV shows	Twitter, Google, Wikipedia, IMDB review, news	Volume	~	TV Ratings	Yes	1,096,057	Cloud-based machine learning
				Sentiment	~				
				Content	+				

2.2.3 Mining Unstructured Text Data

As noted by Archak et al. 2011, textual information embedded in user-generated content has largely been ignored in business research due to a lack of practical tools to analyze such unstructured data. Netzer, Feldman, Goldenberg and Fresko (2012) further noted that the overwhelmingly large volume of data has made analysis extremely difficult, if not impossible. Recently, with the aid of automated machine learning techniques, research has emerged to take advantage of the content of text data rather than its simple numeric volume to provide richer insights, including studies by Das and Chen 2007, Ghose et al. 2007, Eliashberg et al.

2007, Decker and Trusov 2010, Ghose and Ipeirotis 2010, Lee and Bradlow 2011, Ghose et al. 2012, Lee, Hosanagar and Nair 2013, to name a few. More applications utilizing text mining can be found in areas other than marketing, such as computer science (see Pang and Lee 2008 for a review).

Our paper goes beyond text mining by combining it with cloud computing techniques to analyze big text data quickly and cost efficiently. The amount of Twitter feed data that we process is much larger than the scale of any of the previous papers. Please refer to the data description (Section 2.3) and methodology sections (Section 2.4) for details.

2.3 Data description

Our primary goal is to use social media data to predict TV ratings. Below, we explain how we collect the data for TV ratings from five sources of social media, including Twitter Tweets, Google Trends, Wikipedia views, IMDB reviews and Huffington Post news (TGWIH).

2.3.1 TV series

We study a collection of thirty US TV series during the 2008 to 2012 TV seasons. Table 11 shows the number of episodes and each show's ranking in terms of total viewership over the five TV seasons from 2008 to 2012[31]. Among these TV shows, some are very popular, including The Big Bang Theory, Breaking Bad, Grey's Anatomy, NCIS and Two and A Half Men. We choose these shows because 1) advertisers are eager to know their ratings because these shows are quite costly on a CPM basis (for example, The Big Bang Theory commanded a staggering $326,260 per 30-second spot on CBS[32] in 2013, ranking only behind Sunday Night Football); and 2) their popularity may generate significant buzz on online platforms such as Twitter and Google searches. Other less well-known shows are also included, such as Allen Gregory, Charlie's Angels, Hellcats, Harry's Law, and Gary Unmarried. These shows did not last more than two seasons. We choose these shows because their ratings vary quite dramatically, so they are difficult to predict. We also examine some other shows that are neither too popular nor too unpopular to demonstrate the generalizability of our findings. We focus on shows whose titles are unique enough to not be confused with other common words that might appear in Tweets (for example, if we search for another popular show called Community in Tweets, many non-related Tweets with the generic word "community" may appear).

Table 11: TV Series, Number of Episodes and Rank 2008-2012

			2008		2009		2010		2011		2012	
	Show	Channel	Ep's	Rank	Ep's	Ran k	Ep's	Rank	Ep's	Rank	Ep's	Rank

[31] Among the shows, some (e.g., Breaking Bad) are cable shows (AMC is a cable network that generates revenue from user subscription fees), while the others (e.g., The Big Bang Theory) are network broadcast shows (CBS is a broadcasting network whose revenue mainly comes from advertisements). Because cable shows generally have fewer viewers, the ranks of cable shows and network broadcast shows are not directly comparable.

[32] http://www.adweek.com/news/television/big-bang-theory-gets-highest-ad-rates-outside-nfl-153087.

1	2 Broke Girls	CBS							24	32	24	32
2	30 Rock	NBC	22	69	22	86	23	106	22	130	13	99
3	90210	CW	24	172	22	137	22	133	24	145	22	147
4	Allen Gregory	Fox							7			
5	Blue Bloods	CBS					22	19	22	22	23	14
6	Body of Proof	ABC					9	13	20	44	13	34
7	Breaking Bad	AMC	7	86	13	51	13	39	13	22	16	1
8	Charlie's Angels	ABC							8			
9	Cougar Town	ABC			24	57	22	67	15	107	15	
10	Criminal Minds	CBS	26	11	23	16	24	10	24	15	24	20
11	Desperate Housewives	ABC	24	9	23	20	23	26	23	37		
12	Gary Unmarried	CBS	20	74	17	72						
13	Glee	FOX			22	33	22	43	22	56	22	50
14	Gossip Girl	CW	25	168	22	135	22	139	24	188	10	140
15	Grey's Anatomy	ABC	24	2	24	12	22	9	24	12	24	10
16	Harry’s Law	NBC							12	28	22	52
17	Hellcats	CW					22					
18	How I Met Your Mother	CBS	24	49	24	42	24	48	24	45	24	42
19	Lie to me	FOX	13	29	22	57	13	78				
20	Mike & Molly	CBS					24	35	23	31	23	37
21	NCIS	CBS	25	5	24	4	24	5	24	3	24	1
22	Nikita	CW					22	135	23	182	22	145
23	Parks and Recreation	NBC	6	96	24	108	16	116	22	134	22	111
24	Private Practice	ABC			22	10	23	37	22	48	22	49
25	Rules of Engagement	CBS	13	23	13	50	24	49	15	42	13	52
26	Shark Tank	ABC			14	102	9	113	15	98	26	63
27	Smallville	CW	22	152	21	129	22	131				
28	The Big Bang Theory	CBS	23	44	23	12	24	15	24	8	24	3
29	The Vampire Diaries	CW			22	118	22	193	22	166	23	133
30	Two and A Half Men	CBS	24	10	22	11	16	17	24	11	23	11

Data source: http://tvbythenumbers.zap2it.com

2.3.2 NFL

Football is the most popular sport in the US, and football games are among the most watched TV programs. We focus on professional football, the National Football League (NFL), and only their regular season[33] primetime games (8:30 pm): Sunday Night Football on NBC, Thursday Night Football on NFL Network and Monday Night Football on ESPN. The data are collected for three regular seasons from 2010-2012, for a total of 135 games. As shown in Table 12, there are more Sunday games than Monday or Thursday games.

Table 12 NFL Primetime Games 2010-2012

	2010	2011	2012
Sunday Night Football	18	18	19
Thursday Night Football	8	8	13
Monday Night Football	17	17	17

Data source: http://www.nfl.com/schedules

2.3.3 A.C. Nielsen Ratings

We crawl the A.C. Nielsen Ratings data for the 18-49 age range from the website http://tvbythenumbers.zap2it.com for each episode of the thirty TV series[34] and each

[33] The preseason and postseason games’ advertising slots are normally not on the scatter market.
[34] Unfortunately, this website did not collect any ratings data for Breaking Bad for the 2008 and 2009 seasons. Therefore, we only use the 2010-2012 ratings data for Breaking Bad.

primetime NFL game. By definition, rating means the percentage of all television-equipped households that were tuned in to that program at any given moment.

Figure 8a Nielsen Ratings – Five Popular TV Series

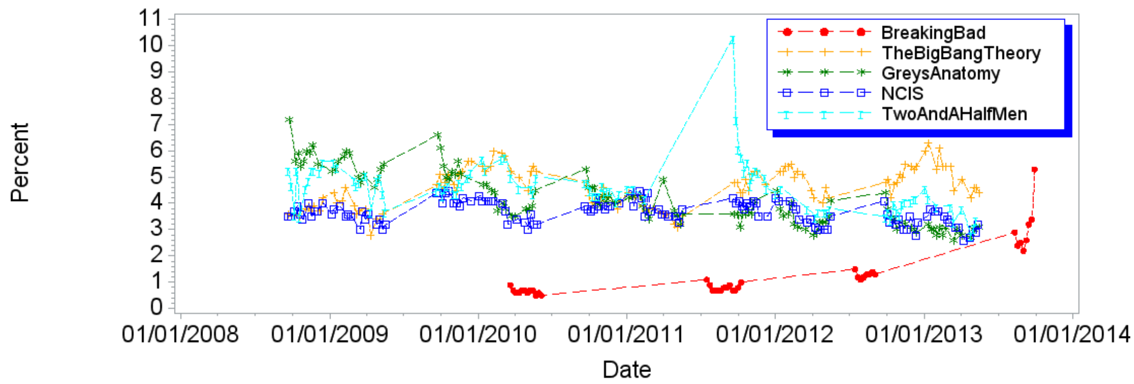


Figure 8b Nielsen Ratings – Five Unpopular TV Series

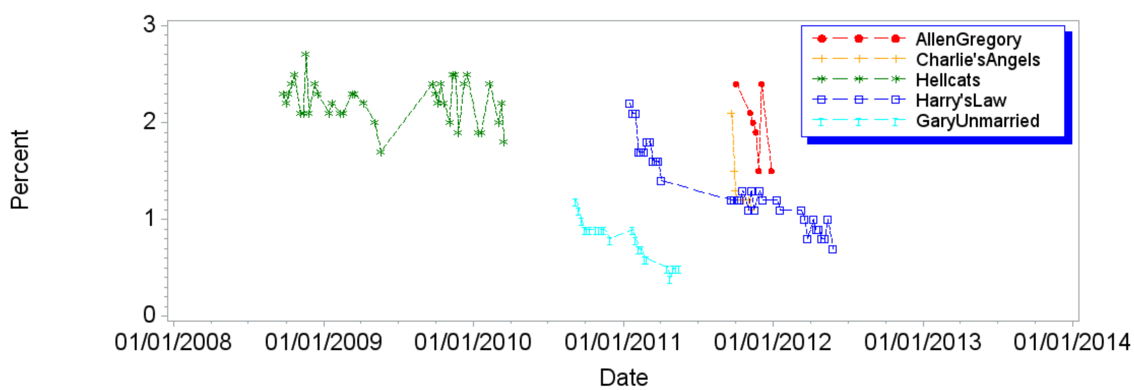


Figure 9 Nielsen Ratings – NFL

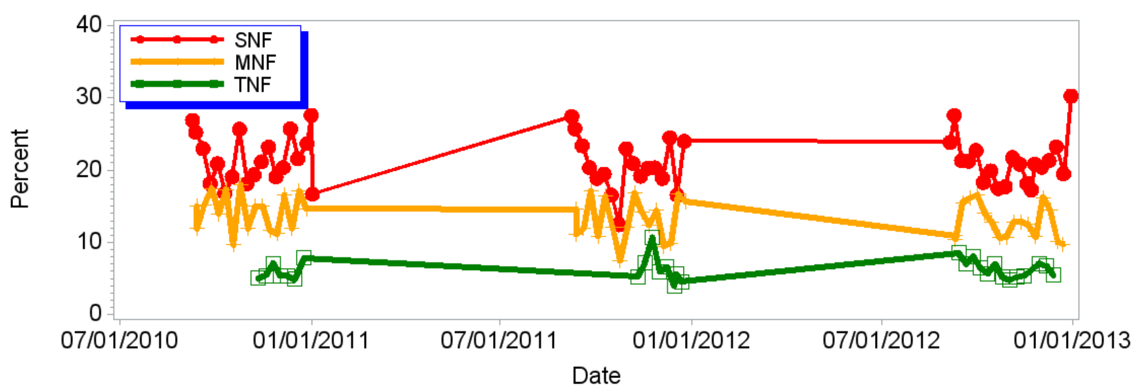


Figure 8 depicts the ratings of the five most popular TV series (Breaking Bad, The Big Bang Theory, Grey's Anatomy, NCIS and Two and A Half Men, in Figure 8a) and five unpopular TV series (Allen Gregory, Charlie's Angels, Gary Unmarried, Harry's Law and Hellcats, in

Figure 8b) for the five seasons from 2008 to 2012. It is interesting to note that each show has a unique trend throughout the five years. Breaking Bad experienced an upward trend over time and reached a quite dramatic spike for the final episode of the last season. The Big Bang Theory gradually improved its ratings each year. Grey's Anatomy's ratings decreased. NCIS remained stable, and Two and A Half Men lost popularity soon after the main character (Charlie Sheen) was replaced in 2011. All of the unpopular shows have a steep downward sloping ratings trend.

For NFL games (in Figure 9), Sunday Night Football is the most watched, while Thursday Night Football is the least watched. Despite the trend across years, when we zoom into each season, the ratings time series are relatively stationary.

Based on Figure 8 and Figure 9, we know that each TV series and each game day of the week (Sunday, Monday or Thursday) has a distinct time-series pattern and can be treated as one member of the panel data.

Next, we describe the data derived from five social media sites: Twitter, Google, Wikipedia, IMDB and Huffington Post (TGWIH).

2.3.4 Twitter

Founded in 2006, Twitter is a real-time information service where people around the world can post ideas, comments, news, photos and videos in 140 characters or less. In 2012, there were more than 500 million registered users who posted 340 million Tweets per day on Twitter. It ranks as the No. 2[35] most visited social networking website according to Global Alexa Page ranking[36].

2.3.4.1 Data gathering

The data we use are collected from Twitter (www.twitter.com) using the "garden hose" (10%) stream[37] on a daily basis from September 1, 2008 to October 27, 2013.[38] Table 13 below shows the size of the dataset per month as measured by number of Tweets (in millions) and the text file storage size.

Table 13 Number of Tweets and File Size by Month 2008-2013

	2008		2009		2010		2011		2012		2013	
	Tweets(Mill)	Size (GB)	Tweets(Mill)	Size (GB)	Tweets(Mill)	Size (GB)	Tweets(Mill)	Size (GB)	Tweets (Mill)	Size (GB)	Tweets (Mill)	Size (GB)
1			22.17	3.67	121.02	32.56	292.99	114.34	762.28	316.53	1282.95	577.33

---

35 Unfortunately, we could not access Facebook data. Moreover, although Facebook has a wider user base, most people choose to make their Facebook status updates and content viewable only to their friends, which leaves only a small percentage of public Facebook updates surrounding TV shows available for researchers. This restriction significantly constrains the number of Facebook posts that we can analyze. Moreover, Twitter is known as "the place that hosts a real-time, public conversation about TV at scale." (http://www.mediaweek.co.uk/article/1288398/twitter-buys-secondsync-mesagraph). Therefore, we do not use Facebook data in this paper.

36 http://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Online_Communities/Social_Networking.

37 http://blog.gnip.com/tag/gardenhose/.

38 We thank Brendan O'Connor and Professor Noah Smith from CMU for providing us with the data.

2			20.65	3.34	126.70	35.58	300.93	117.83	768.13	319.75	1170.41	524.69
3			20.89	3.30	163.07	47.48	369.83	147.29	860.95	361.32	1286.31	600.46
4			23.19	3.39	181.11	54.03	386.54	151.37	893.34	380.43	1267.80	579.93
5			9.21	1.76	210.18	63.37	429.47	167.58	949.08	407.89	1342.71	615.73
6			14.26	3.41	244.58	74.56	450.20	177.58	1004.92	434.91	1321.88	607.61
7			28.17	6.85	168.61	52.39	508.13	207.34	1095.50	478.38	1409.66	650.34
8			50.96	12.61	97.81	31.74	537.77	224.20	1120.58	493.33	1418.14	654.46
9	9.45	1.49	65.55	16.34	198.62	69.23	524.76	219.77	1050.63	464.29	1287.45	592.95
10	16.17	2.55	73.56	18.49	216.35	80.60	570.28	241.26	1111.40	506.26	1299.44	601.49
11	21.06	3.40	82.91	20.99	225.63	88.90	612.91	256.41	1131.26	519.72		
12	20.65	3.38	93.37	24.27	263.78	101.99	685.82	284.17	1220.59	555.67		

From this huge amount of Tweets, we first select the relevant Tweets that discuss the thirty TV series and NFL primetime games.

2.3.4.2 Selecting relevant Tweets

We use four types of identifiers to search for relevant Tweets for the five TV series: 1) name of the show (e.g., Breaking Bad)[39], 2) official Twitter account of the show (e.g., @TwoHalfMen_CBS), 3) a list of hashtags associated with the show (e.g., #AskGreys) and 4) character names on the show (e.g., Sheldon Cooper). As for the NFL, we use similar identifiers, including 1) a list of hashtags of the 32 teams[40] (e.g., #gosteelers) and 2) hashtags of the game (e.g., #SNF). We use Hadoop MapReduce to efficiently select all of the relevant Tweets. Please see the technical details in Section 2.4.

Tables 14 and 15 show some summary statistics of the Tweets for the five TV series and the NFL.

2.3.4.3 TV Series

Table 14: Tweet Frequency, Pre-During-Post

Show	Frequency				Frequency/Episode				Frequency /Hour			
	Pre[41]	During	Post	Total	Pre	During	Post	Total	Pre	During	Post	Total
2 Broke Girls	7258	3071	30861	41189	100.8	42.6	428.6	572.1	4.2	85.3	3.0	3.4
30 Rock	114258	31058	423803	569119	1077.9	293.0	3998.1	5369.0	44.9	586.0	27.9	32.0
90210	136434	38605	699171	874210	1196.8	338.6	6133.1	7668.5	49.9	338.6	42.9	45.6
Allen Gregory	796	1279	6106	8181	113.8	182.7	872.2	1168.7	4.7	365.4	6.1	7.0
Blue Bloods	21344	15445	132811	169600	239.8	173.5	1492.3	1905.6	10.0	173.5	10.4	11.3
Body of Proof	6741	4826	21066	32633	160.5	114.9	501.6	777.0	6.7	114.9	3.5	4.6
Breaking Bad	525234	687585	1199180	2412016	12505.5	16371.3	28552.0	57428.8	521.1	16371.3	199.7	341.8
Charlie's Angels	2365	1014	11994	15372	337.9	144.8	1713.4	2196.1	14.1	144.8	12.0	13.1
Cougar Town	31068	7411	99986	138465	349.1	83.3	1123.4	1555.8	14.5	166.5	7.8	9.3
Criminal Minds	97349	56446	809187	962982	804.5	466.5	6687.5	7958.5	33.5	466.5	46.8	47.4
Desperate Housewives	58109	29285	239505	326899	575.3	289.9	2371.3	3236.6	24.0	289.9	16.6	19.3
Gary Unmarried	532	67	1068	1667	14.4	1.8	28.9	45.0	0.6	3.6	0.2	0.3
Glee	356269	109596	1708721	2174587	3298.8	1014.8	15821.5	20135.1	137.4	1014.8	110.6	119.9
Gossip Girl	220750	70024	982357	1273131	2006.8	636.6	8930.5	11573.9	83.6	636.6	62.5	68.9
Grey's Anatomy	482976	870526	1067363	2420866	4093.0	7377.3	9045.4	20515.8	170.5	7377.3	63.3	122.1
Harry's Law	4493	2841	11994	19328	132.2	83.6	352.8	568.5	5.5	83.6	2.5	3.4

[39] And some variations of it, such as Breaking_Bad and BreakingBad.
[40] The hashtags include the names of the teams.
[41] "Pre" includes all Tweets 24 hours before the show starts. "During" includes Tweets only during the show time. "Post" includes Tweets between the end of one episode and the start of the next.

Hellcats	10994	4985	40707	56686	499.7	226.6	1850.3	2576.6	20.8	226.6	12.9	15.3
How I Met Your Mother	101727	11224	580399	693350	847.7	93.5	4836.7	5777.9	35.3	187.1	33.7	34.4
Lie to me	2994	680	14934	18608	62.4	14.2	311.1	387.7	2.6	14.2	2.2	2.3
Mike & Molly	1589	795	4070	6454	17.3	8.6	44.2	70.1	0.7	17.3	0.3	0.4
NCIS	223767	311939	452242	987933	1849.3	2577.9	3737.5	8164.8	77.1	2577.9	26.1	48.6
Nikita	49174	8495	588896	646564	673.6	116.4	8067.1	8857.0	28.1	116.4	56.4	52.7
Parks and Recreation	30637	10666	97953	139256	278.5	97.0	890.5	1266.0	11.6	193.9	6.2	7.5
Private Practice	41628	27003	133206	201836	408.1	264.7	1305.9	1978.8	17.0	264.7	9.1	11.8
Rules of Engagement	4457	1235	55764	61456	47.9	13.3	599.6	660.8	2.0	26.6	4.2	3.9
Shark Tank	36527	14428	114773	165728	392.8	155.1	1234.1	1782.0	16.4	155.1	8.6	10.6
Smallville	23383	7551	108993	139927	365.4	118.0	1703.0	2186.4	15.2	118.0	11.9	13.0
The Big Bang Theory	556636	327452	1607036	2491124	4717.3	2774.9	13618.9	21111.2	196.6	5550.0	94.9	125.7
The Vampire Diaries	148009	42410	562743	753162	1333.4	382.1	5069.8	6785.2	55.6	382.1	35.5	40.4
Two and A Half Men	156894	99753	471712	728359	1439.4	915.2	4327.6	6682.2	59.9	1830.3	30.2	39.8
Total	3454392	2797695	12278601	18530688	39939.9	35372.7	135648.9	210961.7	1664.1	39878.8	948	1255.8

The number of Tweets on Twitter still varies greatly. If we compare the number of Tweets per episode, Breaking Bad created the most buzz, with more than 57,000 tweets, while the least popular show, Gary Unmarried, had only 45.

For all thirty TV series, the number of Tweets peaks during the show time, and consumers Tweet more frequently before the show than after the show. This result reflects the fact that Twitter is a social platform that creates real-time engagement for viewers. It comes as no surprise that Nielsen is teaming up with Twitter to establish social TV ratings[42].

2.3.4.4 NFL

Table 15 Tweet Frequency – NFL

Time	Frequency	Freq/Game	Freq/Hour
Pre Game	1,520,044	3015.96	125.67
During Game	2,532,638	5045.10	1261.27
Post Game	5,402,517	10783.47	77.02
Total	9,455,200	18760.32	111.67

Similarly, the Tweets are more intensive during the NFL games than before or after the games.

Table 16 Tweet Summary Statistics – 32 NFL Teams

Variable	Mean	Median	Std Dev	Maximum	Minimum	N
New_York_Jets	487.15	299	1062.92	21350	1	1397
Dallas_Cowboys	412.18	123	1325.90	20595	0	1397
New_England_Patriots	335.86	105	1449.39	41909	0	1397
Pittsburgh_Steelers	323.05	88	1521.75	40416	0	1397
New_Orleans_Saints	304.54	123	1632.64	51592	0	1397
Baltimore_Ravens	299.67	70	1621.66	47427	0	1397
Philadelphia_Eagles	282.29	105	762.12	8558	0	1397
Green_Bay_Packers	277.72	70	1344.21	39485	0	1397
Chicago_Bears	273.73	123	779.28	18011	0	1397
San_Francisco_49ers	236.90	70	1181.30	30312	0	1397
New_York_Giants	189.97	70	687.34	19101	0	1397
Washington_Redskins	171.14	70	454.71	5412	0	1397
Indianapolis_Colts	163.26	35	894.73	25655	0	1397
Detroit_Lions	161.95	70	440.27	6467	0	1397

[42] http://www.nielsen.com/us/en/press-room/2013/nielsen-launches-nielsen-twitter-tv-ratings.html.

Oakland_Raiders	151.34	70	308.94	3251	0	1397
Denver_Broncos	143.24	53	463.74	9120	0	1397
Minnesota_Vikings	131.97	53	547.91	16043	0	1397
Atlanta_Falcons	128.68	35	509.82	8628	0	1397
Arizona_Cardinals	127.62	53	364.66	9014	0	1397
Cleveland_Browns	125.95	70	211.57	2407	0	1397
Houston_Texans	123.80	35	435.19	6414	0	1397
Buffalo_Bills	116.78	53	228.05	3497	0	1397
Kansas_City_Chiefs	101.52	35	250.94	3901	0	1397
Miami_Dolphins	93.66	53	179.54	2179	0	1397
Seattle_Seahawks	87.84	35	344.30	8224	0	1397
Carolina_Panthers	81.94	53	143.11	1423	0	1397
Cincinnati_Bengals	81.85	35	240.20	4393	0	1397
St__Louis_Rams	71.84	35	115.69	1494	0	1397
Tampa_Bay_Buccaneers	70.77	35	104.68	1125	0	1397
San_Diego_Chargers	68.10	18	215.54	5711	0	1397
Tennessee_Titans	66.09	35	136.66	1476	0	1397
Jacksonville_Jaguars	63.89	35	106.47	1037	0	1397

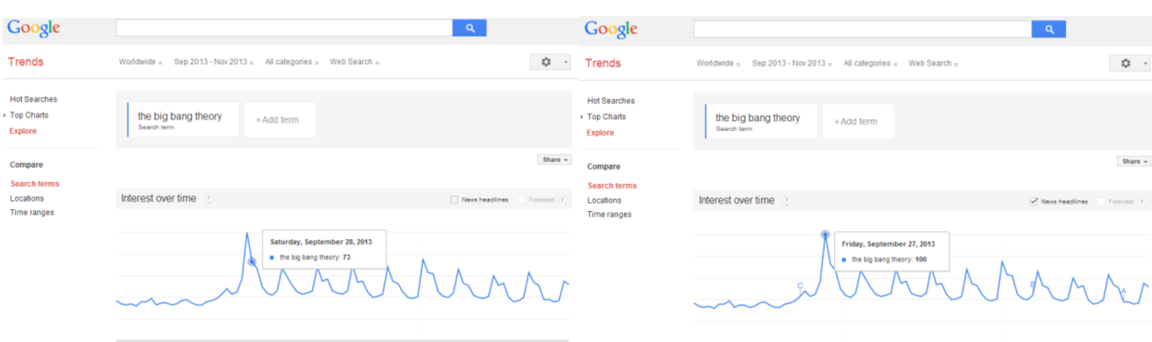
For the 32 NFL teams (Table 16), the most Tweeted team on Twitter was the New York Jets, which had five times more Tweets on average than the least Tweeted team, the Jacksonville Jaguars. The difference in Tweet frequency is largely due to the size of the fan base.

2.3.5 Google Trends

We also collect data from Google Trends (http://www.Google.com/trends/). Google Trends provides the total search volume for a particular search item. The results can be further customized to a certain time range, location and language. For the TV series data, we use the name of the show (e.g., Two and a Half Men) and character names on the show (e.g., Walden Schmidt) as the keywords. For the NFL data, we use the name of the football team (e.g., Pittsburgh Steelers) as the keyword.

In Figure 10, we present the results for the search item "the big bang theory" for the three-month range September 2013 to November 2013. When a search for a term on Google Trends is performed, a graph similar to Figure 10 is shown. The numbers on the graph reflect how many searches have been conducted relative to the total number of searches conducted on Google over time. They do not represent absolute search volume numbers because the data are normalized and presented on a scale from 0-100. Each point on the graph is divided by the highest point, or 100. For example, on September 28, 2013, the number 73 is relative to the highest point, 100, on September 27.

Figure 10 Google Trend Plot



In addition, we use the Google AdWords Keyword planner volume search service[43] to obtain the absolute search volume, or more specifically, a 12-month average of the number of searches for the exact keyword based on the selected location and Search Network targeting settings. Combining these data with the Google Trend relative numbers (multiplying the relative number by a ratio to transform it to an absolute number), we can obtain the absolute search volume for each day.

We record the Google Trends data on a daily basis (by restricting each search query to three months) for each of the thirty TV series and 32 NFL teams. When matching the Google search data to a particular NFL game, we sum the numbers for the two teams that participated in the game.

2.3.6 Wikipedia

Wikipedia is a free access, free content Internet encyclopedia. As of February 2014, it had 18 billion page views and nearly 500 million unique visitors each month[44]. Many of the Wikipedia editors are committed followers of the TV industry who gather information and edit related articles earlier than the show's release date. Consumers may view the edited information before the show or before NFL games; therefore, the Wikipedia edits or views might serve as good predictors of TV ratings.

We extract the edits and page view statistics from the Wikimedia Downloads site (http://dumps.wikimedia.org/other/pagecounts-raw). Table 17 shows the number of webpages and the size of the data files.

Table 17: Wikipedia Sizes

	2008		2009		2010		2011		2012		2013	
	Pages (Bil)	Size (GB)	Pages (Bil)	Size (GB)	Pages (Bil)	Size (GB)	Pages (Bil)	Size (GB)	Pages (Bil)	Size (GB)	Pages (Bil)	Size (GB)
1	2.5	19.9	4.9	39.0	5.4	42.9	6.2	49.2	7.7	61.5	8.7	69.4
2	2.4	18.8	4.4	35.0	5.3	42.2	6.1	48.3	7.2	57.2	7.7	61.5
3	2.5	19.9	5.1	40.8	5.7	45.7	3.1	24.6	7.5	59.6	8.6	68.4
4	2.6	20.9	4.8	38.7	5.3	42.4	6.2	49.1	7.1	56.8	8.1	64.4
5	3.6	28.9	5.1	40.9	5.4	42.8	6.6	52.3	7.5	59.6	8.6	68.8
6	4.3	34.4	5.0	39.7	4.3	33.9	6.4	51.5	7.3	57.9	8.4	66.9
7	4.3	34.0	5.0	39.9	4.4	35.5	6.6	53.0	7.5	60.0	8.1	64.8
8	4.3	34.1	5.1	40.5	5.1	40.6	6.8	54.0	7.5	60.2	8.2	65.8
9	4.3	34.6	3.5	27.7	5.8	46.6	6.1	48.6	7.8	61.9	7.9	63.0
10	4.5	35.7	5.5	43.7	5.9	47.3	7.3	58.4	8.1	64.6	8.5	67.7
11	4.4	35.0	5.5	43.5	5.7	45.8	7.1	56.7	8.0	64.2	8.0	63.5
12	4.7	37.1	5.4	42.9	6.0	47.6	6.6	52.4	8.2	65.6	8.3	66.2

Instead of only focusing on the Wikipedia pages that are designated to the TV programs (for example, http://en.wikipedia.org/wiki/How_I_Met_Your_Mother_(season_8)), we searched through all of the page names that contain the show names (for example, "A Change of Heart (How I Met Your Mother) or How I Met Your Mother (season 6)) or the keyword NFL (or National Football League). After selecting the relevant pages, we also find

[43] https://support.google.com/adwords/answer/2999770?hl=en.
[44] Cohen, Noam (9 February 2014). "Wikipedia vs. the Small Screen". *New York Times*.

the corresponding page edit history using http://tools.wmflabs.org/xtools. Table 18 summarizes the counts of views and edits for all of the shows.

Table 18: Counts for Wikipedia Views, Edits, IMDB Reviews, and Huffington Post Articles

Show	Wikipedia Views	Wikipedia Edits	IMDB Reviews	Huffington Post News
2 Broke Girls	7,072,696	6,282	161	105,000
30 Rock	13,777,360	1,126	80	657,000
90210	17,363,009	4,270	84	33,000
Allen Gregory	912,694	654	48	33,800
Blue Bloods	3,801,506	822	77	44,300
Body of Proof	2,912,555	828	40	187,000
Breaking Bad	41,449,326	4,437	688	229,000
Charlie's Angels	1,880,872	780	35	41,700
Cougar Town	6,586,439	1,622	60	15,100
Criminal Minds	21,460,772	4,752	141	45,400
Desperate Housewives	21,315,771	11,123	155	261,000
Gary Unmarried	978,092	742	20	363
Glee	57,126,905	7,232	158	411,000
Gossip Girl	18,298,602	7,072	183	380,000
Grey's Anatomy	31,073,450	9,463	251	489,000
Harry's Law	1,419,857	654	63	1,360
Hellcats	3,669,048	1,291	20	656
How I Met Your Mother	62,167,139	8,190	321	135,000
Lie to me	8,458,678	1,631	108	236,000
Mike & Molly	2,477,749	771	39	332,000
NCIS	24,884,220	6,381	153	13,300
Nikita	7,936,616	1,655	77	12,600
Parks and Recreation	9,458,196	1,882	90	263,000
Private Practice	6,213,890	1,893	35	84,700
Rules of Engagement	5,472,316	1,151	37	47,000
Shark Tank	1,025,637	1,023	10	10,600
Smallville	14,857,326	8,737	358	974
The Big Bang Theory	56,785,777	8,864	314	229,000
The Vampire Diaries	26,545,974	4,919	244	23,100
Two and A Half Men	30,355,421	7,674	268	499,000
NFL	10754293	10,064		762,000

2.3.7 IMDB Reviews

Consumers also post reviews on discussion forums such as IMDB – the Internet Movie Database. We choose IMDB because it has the highest web traffic ranking (according to Alexa) among all TV-show-related sites. As of January 18, 2015, IMDB had 58 million registered users. The site enables registered users to submit new material and request edits to existing entries[45]. The fourth column in Table 18 describes the number of reviews for each TV series. In contrast to Tweets or Wikipedia views, there are a very limited number of IMDB reviews. The show with the most reviews, Breaking Bad, only has 688 posts over more than 6 years.

2.3.8 The Huffington Post News

Consumers might also be driven by news articles to watch TV series. Therefore, we collect data from The Huffington Post, which is a site that offers news, blogs, and original content and covers entertainment, politics, business etc. The site has extensive traffic and ranks 26th on Alexa as of January 29, 2015. The last column in Table 18 lists the number of news

[45] It also features message boards that stimulate regular debates among authenticated users. However, we cannot access historical discussions before September 2014.

articles related to each TV series and NFL game. Interestingly, some shows that do not have much social buzz (measured by Tweets), such as 30 Rock, but are quite popular on news sites.

2.3.9 Comparing Ratings, Twitter Tweets, Google Trends, Wikipedia Views, IMDB Reviews and Huffington Post News

Figure 11 Comparing Ratings, Tweets and Google Searches – Breaking Bad 2011

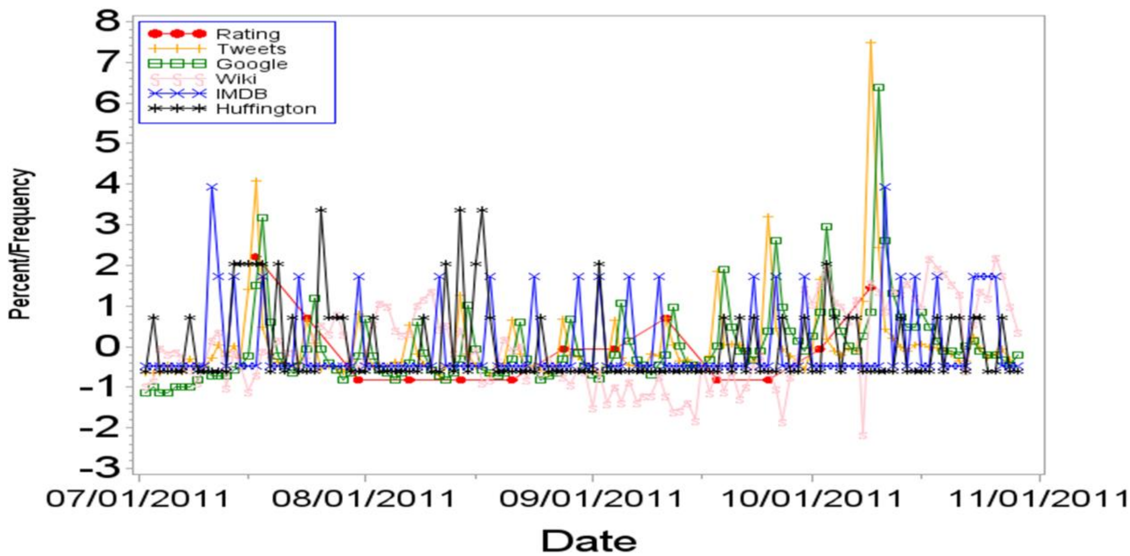
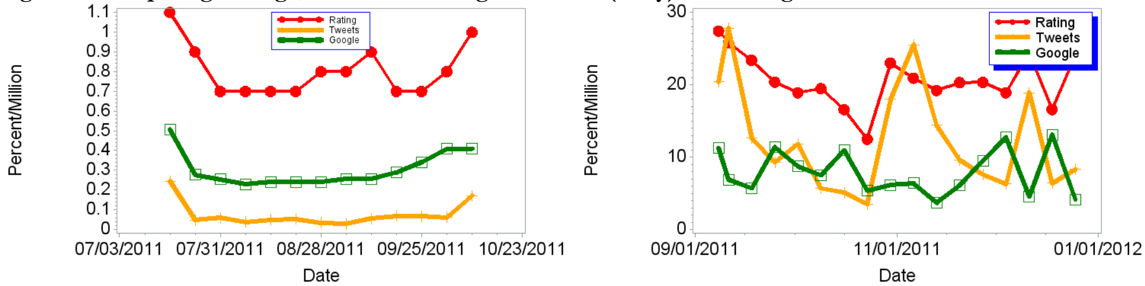


Figure 11 illustrates the relationship between Nielsen Ratings, Tweets, Google Trends, Wikipedia views, IMDB reviews, and Huffington Post news using the example of the fourth season of Breaking Bad in 2011. During that season, there was a new episode every week. Tweets spiked around every show day,[46] and a trail appeared the day after the show. Google searches peaked after each show day. Wikipedia views and Huffington Post news often gradually increased in the next couple of days after a show day. However, IMDB reviews' timing is quite irregular and has some misalignment with ratings. Over the whole season, TV ratings first moved down, then gradually increased and reached a crest on the finale.

[46] Viewings of popular shows such as Breaking Bad could be time deferred, which might reduce the information relevance of Tweets to predict on-air TV ratings. However, we find that Tweets peak during the show time, suggesting that the majority of consumers still watch the shows on air.

Figure 12 Comparing Ratings, Tweets and Google Searches (1 day) – Breaking Bad 2011[47] and SNF 2011



Because Tweets and Google searches cluster around the show days, in Figure 11 we zoom in to determine whether the pattern of Tweets and Google searches one day (24 hours) before each episode shows a similar pattern to the ratings. In Figure 12, we can see that the general patterns for the ratings, Tweets and Google searches are similar, but Tweets and Google searches fail to capture some local variations in the ratings. Regarding the NFL (right panel of Figure 12), the trend in Tweets and Google searches shows an even larger variation from ratings.

2.4 Structured Analysis

2.4.1 Dynamic Panel Data Linear Model

Our model is a dynamic panel data linear model (Bond 2002) that naturally extends Arellano and Bond 1991 (the AB approach). The AB approach is a single equation model with autoregressive dynamics and explanatory variables that are not strictly exogenous. It uses a Generalized Method of Moment (GMM) estimator to take all of the potential orthogonality conditions into account. It solves the Nickell bias problem (Nickell 1981) of a fixed effects model using dynamic panel data. Our model fits the assumptions of the AB approach in that we assume that the TV ratings in the current period can be explained by the ratings in previous periods and the information contained in TGWIH prior to the TV program. In Equation (1), $i \in \{1,2, \dots I\}$ denotes the TV series or the day of week of the NFL primetime game. For example, for TV series, 1 = "2 Broke Girls," with the orders shown in Table 11. For the NFL, {1 = Sunday Night Football, 2 = Thursday Night Football and 3 = Monday Night Football}. The subscript t denotes time. For TV series, one period of time is one episode of the show, while for the NFL, one period is one week in the regular season. $Rating_{it-j}$ is the j_{th}lag of the current period's TV show rating. $\mathcal{F}(Tweet_{it})$ denotes information from Tweets about show i at time t. Because Tweets are unstructured text data, we try several functional forms for $\mathcal{F}$ that are explained later. $Google_{it}$ measures the number of Google searches. $Wiki$ measures the number of Wikipedia views. $\mathcal{F}(IMDB_{it})$and $\mathcal{F}(Huffington_{it})$ contain the content information from IMDB reviews and Huffington Post News. $Controls_{it}$ are several control variables, including a premier indicator, a finale indicator and the age of the show (which episode). u_i is the show-specific, time-invariant fixed

---

[47] We adjusted Tweets and Googles search by a proportion to make the three series more comparable with a similar scale on one plot.

effect that is unobserved to the researcher, and ϵ_{it} is the unobserved shock. Note that we use TGWIH shortly (to be defined later) before the TV program starts so that they are exogenous to the current errors ($u_i + \epsilon_{it}$), but not strictly exogenous to the past errors.

$$\begin{aligned} Rating_{it} = \alpha + \sum_{j=1}^{J} \beta_j Rating_{it-j} + \gamma\mathcal{F}(Tweet_{it}) + \theta Google_{it} + \delta Wiki_{it} + \eta\mathcal{F}(IMDB_{it}) \\ + \phi\mathcal{F}(Huffington_{it}) + \lambda Controls_{it} + u_i + \epsilon_{it} \end{aligned} \tag{1}$$

According to Arellano and Bond (1991), we start with first-differencing the model to eliminate all of the show-specific effects u_i. Then, we use lags (up to period t-2) of both dependent variables and explanatory variables as instruments to perform a GMM estimation. These lags are valid instruments because they are uncorrelated with the difference in the error term but are correlated with the first differences of the endogenous variables. We are careful to avoid using too many lags as instruments, which Roodman (2009) emphasized was a problem[48].

2.4.2 Information in Tweets, IMDB Reviews and Huffington Post News

2.4.2.1 Information measures of Content: Count, Sentiment and n-grams PCA

We use three measures to extract content information from the unstructured text data, including Tweets, IMDB reviews, and Huffington Post news. Below, we use Tweets as an illustration.

One easy measure of information in Tweets is how many times a TV program is discussed by Twitter users. When more users mention a TV program in their Tweets, these users are very likely to watch the program, and the social network that they are connected to is likely to be influenced to watch the program as well. We call this model the "Tweet volume model".

A second measure of information is sentiment, whereby Tweets are classified by polarity, i.e., positive, neutral and negative. Positive (negative) Tweets express favorable (unfavorable) feelings toward a show, while neutral Tweets do not express opinions but state facts. Hypothetically, positive Tweets are likely to generate positive feedback for a show, thus increasing TV ratings, whereas negative Tweets signal that consumers might stop watching the show because they are dissatisfied, thus lowering future ratings. We construct two variables, $t\_pos = \frac{\#\, of\, positive\, tweets}{total\, \#\, of\, tweets}$ and $\_neg = \frac{\#\, of\, negative\, tweets}{total\, \#\, of\, tweets}$, and we test their effect on ratings in the model called "Tweet sentiment model". We constructed a sentiment analysis classifier using the LingPipe[49] linguistic analysis package, which provides a set of open-source java libraries for natural language processing tasks. We used the

[48] We finally chose lags up to 5 periods ahead as instruments to prevent weakening the Hansen test (Anderson and Sorenson 1996; Bowsher 2002)

[49] http://alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html.

DynamicLMClassifier, which is a language model classifier that accepts training events of categorized character sequences. Training is based on a multivariate estimator for the category distribution and dynamic language models for the per-category character sequence estimators. To obtain labeled training data for the classifier, we hired two independent coders who have expertise in TV shows and NFL games to manually label 4% of the Tweets for each show/NFL team. We further adjusted the classified Tweets using the list of positive and negative opinion words provided by Hu and Liu 2004.

The third measure digs deeper into the variety of content in Tweets. We discover that some Tweets related to a program might only express users' opinion about the program rather than indicating an action to watch the upcoming program. For instance, consider the following two sample Tweets regarding the show "Breaking Bad":

"I learnt how to cook meth like the guy in breaking bad"
"Pumped for the season finale of Breaking Bad tonight. Only 4 hours and 37 minutes to go."

The first Tweet talks about a featured behavior of the actor in the show. From this Tweet, we can infer that the Twitter user has watched Breaking Bad and is interested in its story. Nevertheless, the second Tweet directly states the future watching behavior of the user. If, before the show starts, there are many Tweets similar to the second one, the show's rating is very likely to be high. In contrast, variations of Tweets similar to the first Tweet may have far less predictive power.

Based on this rationale, we construct a third measure of information to make inferences from the full content of Tweets. More specifically, we use the counts of each n-gram in the Tweet. An n-gram is a continuous sequence of n words in the text. For example, "big data" is a 2-gram, and 'big' is a 1-gram. The sample Tweet 'I love Pittsburgh Steelers' contains four 1-grams, three 2-grams, two 3-grams and one 4-gram. Because phrases provide more interpretable information than a single word, we choose to count n-grams rather than only counting words. We label this model the "Tweet content model".

2.4.2.2 Information Timeliness

Another decision to make concerns the length of time to collect TGWIH before one show starts. The shows are generally broadcast on a weekly basis. Echoing what we find in Section 2.3, Tweets and Google searches also follow a weekly trend whereby more instances take place around (shortly before and after) the show. Intuitively, on the first or second day after the previous show, consumers are more likely to focus on the old show, while one day or two before the new show, consumers are more likely to express buzz about the new show. Following this intuition, we use TGWIH 24 or 48 hours before the new show starts as the input variable. It is interesting to compare the performance of the 24-hour measure against the 48-hour measure to evaluate the value of information over time[50].

[50] We also conducted analysis for the one week window. Results are available upon request.

2.4.3 Challenges in Processing Enormous Unstructured Data – Cloud Computing Techniques

As shown in Section 2.3, our analysis involves an enormous amount of unstructured data. For example, we have approximately 1.8 billion Tweets, 433 billion Wikipedia pages, and 5.5 million Huffington Post news articles. Our data-cleaning process includes three major processes: 1) selecting relevant Tweets/Wikipedia pages, 2) n-gram count, and 3) the stochastic singular value decomposition (SSVD). The first two tasks can be performed in a streaming fashion (no out-of-memory problem) but are extremely time consuming on a single machine given the volume of our data. The last task cannot even be performed on a single machine because the size of the matrix does not fit into memory.

For example, the content information in Tweets is enormous. Even when using the 24-hour measure, we selected 6,894,624 Tweets related to the thirty TV series and 2,004,987 Tweets related to the NFL. These Tweets generate 28,044,202 and 9,028,774 n-grams that have appeared at least 5 times (interestingly, NFL Tweets are much more concentrated on high frequency phrases). Moreover, in our regression model for the TV shows, we hope to incorporate all of the content information. One way to do this is to use the frequency of all n-grams as features. This approach provides us a huge feature space. Therefore, we must rely on dimension reduction techniques such as Principle Component Analysis (PCA) to make the task more tractable. However, performing PCA on a 2339*28,044,202[51] matrix cannot be done on a single machine because the matrix is too large to be stored in memory.

Our solution to this challenge is to use the stochastic singular value decomposition (SSVD) method developed by Helko (2012). The key idea behind the SSVD is that when a data matrix is too large to be stored in memory, randomized sampling (the stochastic element of SSVD) allows the algorithms to work in a streaming environment to rapidly construct a low-rank approximation of the matrix. It parallelizes and distributes both the randomized sampling stage and the factorization stage using Hadoop MapReduce.

Next, we explain exactly how we solve these challenges using cloud computing services.

Because the computing task cannot be handled by a single machine, programs have been developed to exploit the capacities of massively distributed computational resources. MapReduce is a good example[52]. MapReduce is a programming model for processing large datasets using a parallel, distributed algorithm on a cluster. It is very powerful because it abstracts the complexities of parallel programming down to two operations: a map and a reduce. Both the map and the reduce steps can be parallelized on a cluster of computers. Between the map and the reduce, the job involves shuffling and sorting the keys such that all key-value pairs of the same key go to the same reduce for the next step. Thus, the

[51] The number of episodes of shows is 2,339. The number of n-grams generated from Tweets 24 hours before a show is 28,044,202. For the larger dataset of Tweets 48 hours before a show, the corresponding number of n-grams is 34,855,764.

[52] MapReduce developers tout MapReduce for its scalability, fault tolerance and elasticity. Google uses it to process 20Pb of data per day.

communication of data is only between map and reduce. The details, such as distributed computation, file storage, communication and data transfer, are left to the framework, such as Hadoop[53], to handle.

We implement MapReduce using Amazon Elastic MapReduce (EMR)[54]. Specifically, we used EMR for all three tasks: 1) selecting the relevant Tweets/Wikipedia pages, 2) performing the n-gram counts, and 3) conducting the SSVD. Table 19 summarizes how we design the map and reduce the jobs for each task.

Table 19 MapReduce for Three Tasks

No	Task	Map		Reduce
		Key	Value	
1	Select Relevant Tweets/Wiki	Keyword	1, Text	Summation
2	N-gram Frequency Count	N-gram	1	Summation
3.1	SSVD - Matrix Multiplication	Matrix row index	Matrix row vector	Null
3.2	SSVD - Orthogonalization	Submatrix	QR matrix	Summation

We can use the Tweet n-gram count task as an illustration. In the first procedure, "Map" filters the input data into key value pairs. When reading one Tweet as the input, if we find one n-gram in the Tweet, then we set the key as the n-gram and the value as 1. The second procedure, "Reduce", then summarizes the key-value pairs generated by the "Map" procedure. In other words, "Reduce" adds all of the values of the same key (n-gram) as the summary count of the n-gram.

When implementing the SSVD, we employed Mahout, an open-source machine learning library that utilizes Hadoop MapReduce to implement scalable distributed algorithms. It essentially breaks down the singular value decomposition of a huge matrix into two basic operations, matrix multiplication and orthogonalization. Because both operations can rely on MapReduce to be performed in distributed clusters, our computational challenge is resolved.

2.4.4 Alternative Machine Learning Models

In addition to the cloud-based PCA model and the dynamic panel linear model explained above, we also employ alternative content extraction models and machine learning models for prediction comparison. In terms of content extraction, we compare the current cloud-based PCA model with the Latent Dirichlet Allocation Topic Modeling approach (Blei et al. 2003). Moreover, we compare the dynamic panel linear model with machine learning models that are widely used in other papers related to Twitter predictions, including Auto Regression with Exogenous Independent Variables (Autoregression X), Multi-layered feedforward

[53] Hadoop is an open-source software framework that allows distributed processing of large datasets across clusters of computers using simple programming models. It contains 1) the Hadoop Common package, which provides file system and OS level abstraction; 2) Yarn, a MapReduce engine; and 3) the Hadoop Distributed File System. These mechanisms automatically break down jobs into distributed tasks, schedule jobs and tasks efficiently at participating cluster nodes, and tolerate data and task failures.
[54] EMR was created to enable researchers, businesses, and developers to process vast amounts of data easily and efficiently in a pay-as-you-go fashion. For more detailed information about implementation, please see Appendix A4.

neural networks and the Self-organizing Fuzzy Neural Network (SOFNN) model. For detailed information about these alternative models, please see Appendix A2.1. We discuss the prediction performance of these competing models in Section 2.5.

2.5 Results

2.5.1 Specification Tests and Fit

2.5.1.1 Stationarity Tests

We apply the Augmented Dickey-Fuller (ADF) unit root test on all of the variables (Rating, Twitter Tweets, Tweet sentiments, Tweet Principal Components, Google searches, Wikipedia views, IMDB reviews and Huffington Post news). In all cases, the null hypothesis that the series is non-stationary is rejected.

2.5.1.2 IV validity and Serial correlation

We use the Sargan Chi-squared statistic to test the validity of the instruments. Tables 23 to 25 report that all of the over-identifying test statistics are insignificant. Therefore, we cannot reject the joint validity of the instrument sets. In addition, we use the method developed by Arellano and Bond (1991) to check whether the errors are serially correlated (AB test score Z in Tables 23-25). There is no evidence of first-order serial correlation.[55]

2.5.1.3 Number of Lags

To decide the number of lags (J in Equation (1)) of the dependent variable in our model, we use the MMSC-BIC statistics developed by Andrews and Wu (2001). The comparisons show that only including the first lag yields the lowest MMSC-BIC. Therefore, in the following Tables (23-25), we only report the results with the first lag included.

2.5.2 Main Findings

2.5.2.1 People Tweet about what they are going to do

2.5.2.1.1 The n-grams

Table 20: Most Popular N-grams in Shows

Show	1st	Count	%	2nd	Count	%	3rd	Count	%	Total
2 Broke Girls	2 Broke Girls	21959	5.0%	OfficialKat	19199	4.3%	Watch	7227	1.6%	442665
30 Rock	30 Rock	568593	8.0%	watch	127239	1.8%	Love	41667	0.6%	7148180
90210	90210	893597	9.3%	watch	177318	1.8%	Doctor	61888	0.6%	9601842
Allen Gregory	Allen Gregory	8181	9.2%	JonahHill	2100	2.4%	Watch	1713	1.9%	89161
Blue Bloods	Blue Bloods	169526	8.5%	watch	21809	1.1%	Love	14294	0.7%	1995877
Body of Proof	Body of Proof	32658	8.8%	watch	7609	2.0%	DanaDelany	3958	1.1%	372839
Breaking Bad	Breaking Bad	113950	2.7%	aaronpaul	28519	0.7%	Tonight	18407	0.4%	4145246
Charlie's Angels	Charlie's Angels	15372	9.4%	watch	3379	2.1%	Cancel	1182	0.7%	162895
Cougar Town	Cougar Town	138465	8.8%	watch	28818	1.8%	Show	14161	0.9%	1575302
Criminal Minds	Criminal Minds	962982	9.4%	watch	215756	2.1%	Gublernaton	127008	1.2%	10284082
Desperate Housewives	Desperate Housewives	326899	9.6%	watch	91934	2.7%	Season	26254	0.8%	3404224
Gary Unmarried	Gary Unmarried	1664	9.1%	watch	389	2.1%	Funny	143	0.8%	18334
Glee	Glee	2174587	8.7%	watch	311489	1.2%	Love	208418	0.8%	25074162
Gossip Girl	Gossip Girl	1273131	8.4%	watch	257260	1.7%	Now	116795	0.8%	15177057
Grey's Anatomy	Grey's Anatomy	111321	2.6%	tonight	19369	0.4%	Watch	15302	0.4%	4307147

[55] There is also no second-order serial correlation. The results are not reported here but are available upon request.

Harry's Law	Harry's Law	19328	8.6%	watch	4130	1.8%	Kathy Bates	2379	1.1%	224088
Hellcats	Hellcats	56686	9.8%	watch	10595	1.8%	Ashley Tisdale	7811	1.4%	577529
How I Met Your Mother	How I Met Your Mother	693350	8.0%	watch	139585	1.6%	Neil Patrick Harris	117773	1.4%	8625859
Lie to me	Lie to me	18608	8.7%	Tim Roth	6525	3.1%	Watch	1844	0.9%	213067
Mike & Molly	Mike & Molly	6454	9.9%	watch	2095	3.2%	checked-in	891	1.4%	65037
NCIS	NCIS	54054	2.3%	watch	8014	0.3%	Tonight	7091	0.3%	2325168
Nikita	Nikita	646564	9.1%	go	91309	1.3%	Love	47038	0.7%	7075657
Parks and Recreation	Parks and Recreation	139256	8.5%	watch	23400	1.4%	Office	11899	0.7%	1631277
Private Practice	Private Practice	201836	8.6%	Grey's Anatomy	53815	2.3%	Watch	46560	2.0%	2354062
Rules of Engagement	Rules of Engagement	61456	7.7%	David Spade	27061	3.4%	Go	8554	1.1%	795710
Shark Tank	Shark Tank	165728	7.7%	watch	21795	1.0%	Dragon's Den	3905	0.2%	2141261
Smallville	Smallville	139927	8.5%	watch	26197	1.6%	Season	16833	1.0%	1651297
The Big Bang Theory	Big Bang *	121590	2.6%	watch	16103	0.3%	Tonight	10379	0.2%	4639872
The Vampire Diaries	The Vampire Diaries	753162	8.3%	tv	255581	2.8%	Watch	136691	1.5%	9025715
Two and A Half Men	Two and a half men	32337	2.3%	watch	5043	0.4%	Charlie Sheen	3901	0.3%	1390918
NFL	Cowboy	47251	0.5%	game	36550	0.4%	Tonight	28432	0.3%	8603269
Total										135138799

*N-grams: Big, Bang, and Big Bang have the same frequency. The same logic applies to the names of the other shows.

In Table 20, we list the n-grams with the 1st, 2nd and 3rd highest frequency for each of the five TV series and the NFL. Across all five TV series, the highest mentioned topic is the name of the program, such as Breaking Bad. Moreover, "watch" and "tonight" appear with very high frequency. In fact, we find many Tweets talking about the consumer's plan to watch the show. For example, "I can't wait to watch Breaking Bad tonight". The content in this type of Tweet is a useful predictor for the rating of the program. Not surprisingly, we also find that celebrities on the TV programs are another hot topic, such as Charlie Sheen in Two and A Half Men. Twitter users express their preference for a celebrity and also reTweet what the celebrity says. For example, many people Tweeted that "Two and A Half Men is never the same without Charlie Sheen" after Sheen was replaced on the show in 2011. Fans of The Big Bang Theory often Tweeted phrases such as "Bazinga! Love Sheldon", where "Sheldon" is a main character on the show, and "Bazinga" is a phrase that is frequently uttered by him.

2.5.2.1.2 Sentiment

Table 21 Sentiment Distribution

Show	Positive	Neutral	Negative	Total
2 Broke Girls	21.23%	78.16%	0.6%	41189
30 Rock	30.11%	69.85%	0.05%	569119
90210	4.85%	95.13%	0.02%	874210
Allen Gregory	31.56%	41.59%	26.84%	8181
Blue Bloods	18.93%	80.66%	0.40%	169600
Body of Proof	17.84%	81.38%	0.78%	32633
Breaking Bad	12.33%	87.11%	0.56%	2412016
Charlie's Angels	41.76%	46.31%	11.93%	15372
Cougar Town	12.34%	87.40%	0.26%	138465
Criminal Minds	8.35%	91.47%	0.18%	962982
Desperate Housewives	20.42%	79.55%	0.03%	326899
Gary Unmarried	26.17%	56.64%	17.18%	1667
Glee	23.77%	75.37%	0.86%	2174587
Gossip Girl	8.35%	91.47%	0.18%	1273131
Grey's Anatomy	33.17%	61.22%	5.60%	2420866
Harry's Law	10.26%	88.21%	1.54%	19328
Hellcats	31.46%	67.45%	1.08%	56686
How I Met Your Mother	14.55%	85.12%	0.33%	693350

Lie to me	18.17%	80.55%	1.29%	18608
Mike & Molly	18.66%	80.60%	0.75%	6454
NCIS	36.38%	63.56%	0.06%	987933
Nikita	5.92%	71.39%	5.92%	646564
Parks and Recreation	21.60%	78.29%	0.12%	139256
Private Practice	32.37%	67.09%	0.54%	201836
Rules of Engagement	4.29%	95.61%	0.10%	61456
Shark Tank	16.93%	78.74%	4.33%	165728
Smallville	6.21%	93.74%	0.04%	139927
The Big Bang Theory	31.91%	66.90%	1.19%	2491124
The Vampire Diaries	27.87%	72.08%	0.05%	753162
Two and A Half Men	11.84%	85.30%	2.86%	728359
NFL	36.31%	57.37%	6.32%	2040139

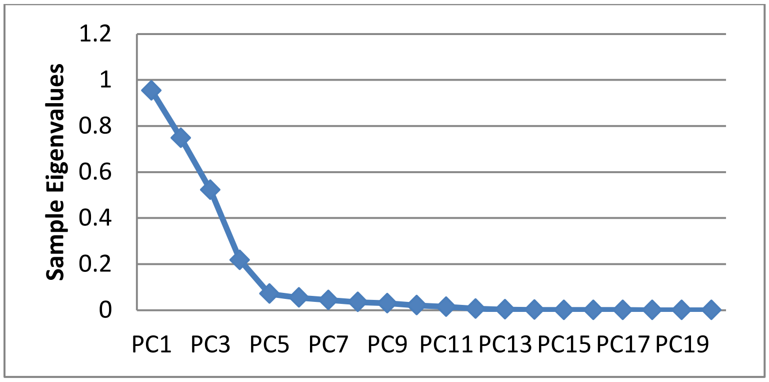
We find several interesting phenomena in the sentiment analysis of the Tweets. First, the majority of the Tweets are neutral and document consumers' mundane lives, such as their actions and plans. Second, consistent with the previous findings of Godes and Mayzlin 2004, consumers are much more positive than negative about the TV shows they watch. On average, there are 7.7 times more positive Tweets than negative Tweets. However, this ratio is relatively lower for NFL games compared with TV series.

One way to explain this result is self-selection bias. The consumers who Tweet about TV shows are those who enjoy the shows the most. Thus, they are more positive than the whole population. However, the theory of cognitive dissonance indicates that people tend to justify their own mistakes. In our context, even if a consumer dislikes a show that she watched, she might be unwilling to admit that she made a mistake in choosing the show by broadcasting it on Twitter. These properties of sentiment analysis of Tweets prevent this aspect from being predictive of TV ratings.

2.5.2.1.3 PCA

In the principal component analysis, we follow the standard approach and use a "scree plot" (Cattell 1966) to decide how many principal components to retain. If the largest few eigenvalues in the covariance matrix dominate in magnitude, then the scree plot will exhibit an "elbow". We apply this "elbow" rule and select four principal components out of the 28,044,202 n-gram features based on Figure 13.

Figure 13 Scree Plot



It is interesting to examine the matrix of eigenvectors (loadings) to see which n-grams contribute the most to the chosen principal components. Table 22 shows the three n-grams with the largest loadings on each of the first four principal components (PC).

Table 22 N-grams with Highest Loadings on First Four PCs

PC1	PC2	PC3	PC4
tonight	bed	season	Excited
can't wait	home	start	Finale
watch	tv	premiere	Love

Consistent with our findings from the n-gram count, words and phrases such as "tonight", "can't wait" and "watch" have the largest projection on the first PC. Location and device-related words such as "home" and "tv" contribute most to the second PC. The third PC captures consumers' attention to the "premiere" of the "season", and the fourth PC contains positive emotions such as "excited" and "love" as well as another hot topic, "finale". Overall, the first four PCs cover consumers' intention to watch the shows. Later in the analysis of the regression results, we confirm that this summary of information is indicative of users' upcoming consumption (watching the show).

2.5.2.2 Twitter Content is a lead indicator of TV Ratings

Table 23 Impact of Previous 24 Hours of Tweets and Google Searches on TV Ratings: TV Series[56]

24 hr	1	2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Rl	T	G	W	I	H	Rl + T	RI + G	T + Sen	Rl+T+Sen	Topic	Rl+Topic	RI+T+G+W+I+H+Topic	PC	Rl+PC	RI+T+G+W+I+H+PC
Rating_lag	0.459						0.560	0.404		0.441		0.460	0.389		0.401	0.391
p_value	<.001						<.001	<.001		<.001		<.001	<.001		<.001	<.001
Tweets		0.001					0.001		0.001	8.53E-04			0.001			0.001
p_value		<.001					0.013		0.027	0.161			0.023			0.025
Google			1.05E-06					4.86E-06					3.04E-06			1.52E-06
p_value			0.028					<.001					<.001			0.002
Wiki				1.87E-05									8.37E-06			2.81E-05
p_value				<.001									0.102			<0.001
IMDB					-0.182											
p_value					0.395											
Huffington						0.033										
p_value						0.247										
Tweet_Pos									0.001	0.002						
p_value									0.264	0.421						
Tweet_Neg									0.006	0.002						
p_value									0.461	0.84						
Tweet_PC1														0.357	0.533	0.439
p_value														<.001	<.001	<.001
Tweet_PC2														0.594	0.560	0.859
p_value														<.001	<.001	<.001
Tweet_PC3														0.934	0.429	0.668
p_value														<.001	<.001	<.001
Tweet_PC4														1.013	0.916	1.050
p_value														<.001	<.001	<.001
Tweet_T1											0.423	0.385	0.492			
p_value											<.001	<.001	<.001			
Tweet_T2											1.265	0.740	0.438			
p_value											<.001	<.001	<.001			
Tweet_T3											0.559	0.578	0.928			
p_value											<.001	<.001	<.001			
Tweet_T4											1.467	0.459	1.123			
p_value											<.001	<.001	<.001			
Tweet_T5											1.840	0.781	0.635			
p_value											<.001	<.001	<.001			
Premier	0.369	0.270	0.355	0.270	0.407	0.413	0.329	0.235	0.269	0.284	0.349	0.365	0.191	0.329	0.362	0.384
p_value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001	<.001	<.001
Finale	-0.009	-0.037	-0.023	-0.078	-0.138	-0.096	-0.004	-0.072	-0.037	-0.054	-0.022	-0.008	-0.115	-0.017	-0.019	-0.041
p_value	0.822	0.313	0.530	0.029	0.008	0.020	0.916	0.078	0.307	0.097	0.540	0.854	0.011	0.699	0.636	0.256
Age	-0.115	-0.137	-0.135	-0.211	-0.142	-0.140	-0.093	-0.104	-0.137	-0.137	-0.145	-0.113	-0.142	-0.101	-0.123	-0.116
p_value	0.016	<.001	<.001	<.001	<.001	<.001	0.001	0.006	<.001	0.001	<.001	0.017	0.001	0.002	<.001	<.001
R2**	0.755	0.065	0.049	0.119	0.012	0.008	0.759	0.776	0.066	0.777	0.814	0.830	0.832	0.823	0.857	0.860
Wald Chi2	244.810	150.960	113.200	203.880	98.460	72.320	274.010	274.450	152.640	279.120	283.960	300.790	521.300	404.150	422.930	442.180
p-value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
AR(1)	-2.944						-2.765	-3.141		-2.887		-2.923	-3.111		-2.994	-2.7482
p-value	0.003						0.006	0.002		0.004		0.004	0.002		0.003	0.006
AR(2)	0.347						1.938	0.379		0.652		0.311	-0.333		0.738	0.4832
p-value	0.729						0.053	0.705		0.514		0.756	0.739		0.316	0.6817
Sargan Chi2	24.961						20.717	20.862		23.292		17.764	7.264		25.640	21.87
p-value	1						1	0.9612		1		1	1		1	1

[56] We also conduct the sentiment analysis and n-gram PCA for IMDB reviews and Huffington Post news. The results are similar to columns 5 and 6 in Table 23.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
MMSC-BIC	-11735						-3475	-11631		-11525		-11773	-11543		-11968	-11964

The results show that the lagged rating itself explains 75.5% of the variation in the current rating (column 1 of Table 23). In contrast, the sheer volume of Tweets within 24 hours before the show only accounts for 6.5% of the variation (column 2). Similarly, Google searches, Wikipedia views, IMDB review count and Huffington Post news articles in the past 24 hours only explain 4.9% (column 3), 11.9% (column 4), 1.2% (column 5) and 0.8% (column 6) of the variation, respectively. Moreover, when we combine the lagged rating and Tweet counts (column 7) or Google searches (column 8), the R-squared does not increase much from the R-squared that merely includes the lagged ratings (column 1). This result implies that the pure "Tweet volume model" does not have much explanatory power. Sentiments such as the proportion of positive or negative Tweets are not much better than volume in predicting TV ratings. As can been observed from Models 9 and 10 in Table 23, in the presence of Tweet volume, the effects of positive or negative Tweets are not statistically significant. We also experiment with other functional forms of the sentiment variables, such as quadratic terms and exponential terms; the results are qualitatively similar to the linear form. Therefore, the "Tweet sentiment model" is also not good at predicting TV ratings.

However, the first four principal components from the Tweet n-gram features produce an R2 of 0.823 (column 14), which is comparable to that of the model with only the lagged rating included. A better model fit is found when we combine PCs with the lagged rating (column 15) and both lagged rating and Google searches (column 16). This result shows that the "Tweet content model" outperforms the "Tweet volume model" and the "Tweet sentiment model" in predicting TV series ratings.

Similarly, when we employ the Latent Dirichlet Allocation (Blei et. al 2003) technique to extract content information, we obtain an R2 of 0.814 (column 11), which is quite close to what we obtain in the n-gram PCA case. Further adding lagged rating and other social media information improves the R2 (columns 12 and 13). However, n-gram PCA performs better overall than LDA (we confirm this result in the next section on the forecasting task).

Another aspect to note is that we find marginal improvement in the R2 by adding Google searches. This result suggests that information derived from Google searches and Twitter posts may be mostly substitutes.

Table 24 Impact of Previous 48 Hours of Tweets and Google Searches on TV Ratings: TV Series

48 hr	1	2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Rl	T	G	W	I	H	Rl + T	RI + G	T + Sen	Rl+T+Sen	Topic	Rl+Topic	RI+T+G+W+Topic	PC	Rl+PC	RI+T+G+W+PC
Rating_lag	0.466						0.554	0.407		0.443		0.458	0.393		0.404	0.387
p_value	<.001						<.001	<.001		<.001		<.001	<.001		<.001	<.001
Tweets		-0.002					0.002		0.001	7.45E-04			0.001			0.001
p_value		<.001					0.014		0.026	0.160			0.022			0.025
Google			-2.70E-09					3.87E-06					2.12E-06			7.50E-07
p_value			0.027					<.001					<.001			0.002
Wiki				1.76E-05									7.38E-06			2.75E-05
p_value				<.001									0.102			<0.001
IMDB					-0.157											
p_value					0.313											
Huffington						0.03										
p_value						0.281										
Tweet_Pos									0.001	0.004						
p_value									0.254	0.419						
Tweet_Neg									0.005	0.002						

p_value									0.460	0.834						
Tweet_PC1														0.350	0.538	0.437
p_value														<.001	<.001	<.001
Tweet_PC2														0.583	0.557	0.841
p_value														<.001	<.001	<.001
Tweet_PC3														0.916	0.427	0.672
p_value														<.001	<.001	<.001
Tweet_PC4														0.981	0.907	1.058
p_value														<.001	<.001	<.001
Tweet_T1											0.412	0.390	0.380			
p_value											<.001	<.001	<.001			
Tweet_T2											1.349	0.772	0.640			
p_value											<.001	<.001	<.001			
Tweet_T3											0.637	0.516	1.070			
p_value											<.001	<.001	<.001			
Tweet_T4											1.496	0.543	0.827			
p_value											<.001	<.001	<.001			
Tweet_T5											1.931	0.856	0.717			
p_value											<.001	<.001	<.001			
Premier	0.371	0.303	0.364	0.269	0.348	0.385	0.331	0.234	0.273	0.288	0.315	0.365	0.186	0.319	0.371	0.356
p_value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001	<.001	<.001
Finale	-0.009	-0.036	-0.022	-0.087	-0.179	-0.083	-0.002	-0.063	-0.050	-0.049	-0.023	-0.005	-0.104	-0.009	-0.018	-0.038
p_value	0.714	0.379	0.578	0.022	0.003	0.021	0.986	0.076	0.303	0.110	0.553	0.767	0.012	0.666	0.589	0.234
Age	-0.114	-0.135	-0.043	-0.058	-0.178	-0.122	-0.146	-0.144	-0.071	-0.091	-0.152	-0.159	-0.186	-0.198	-0.115	-0.064
p_value	<.001	<.001	<.001	<.001	<.001	<.001	0.001739	<.001	<.001	<.001	<.001	0.016	0.001	0.006	<.001	<.001
R2**	0.758	0.068	0.051	0.120	0.023	0.015	0.760	0.778	0.067	0.779	0.817	0.832	0.835	0.831	0.859	0.863
Wald Chi2	244.908	151.089	113.323	203.874	100.336	73.281	274.199	274.517	152.805	279.262	283.930	301.055	521.723	404.245	423.217	442.597
p-value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
AR(1)	-3.040						-2.839	-3.341		-3.087		-3.123	-3.311		-3.194	-3.15055
p-value	0.001						0.002	0.000		0.001		0.001	0.000		0.001	0.001
AR(2)	0.048						1.538	0.279		0.452		0.209	-0.533		0.440	0.283
p-value	0.481						0.062	0.390		0.325		0.417	0.298		0.331	0.388
Sargan Chi2	24.764						20.477	20.651		23.103		17.518	7.231		25.371	21.418
p-value	1						1	1		1		1	1		1	1
MMSC-BIC	-11738						-3466	-11637		-11522		-11768	-11546		-11974	-11973

Table 24 considers Tweets and Google searches 24 hours before the new TV shows start. Will it help to involve more information by extending the data collection period to 48 hours? Comparing the R2s from Table 24 with those from Table 23, we find that adding more information into TGWIH increases the model fits, but the increments are very small. For example, if we compare column 14 in Table 24 with column 8 in Table 23, the R2 improves by less than 1%. This result implies that consumers' buzz on online platforms about TV shows is very transient. Information value decays with time very quickly. Most of the other findings from Table 23 are replicated in Table 24.

Table 25 Impact of Tweets and Google Searches on TV Ratings: NFL

Rating	1	2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Rl	T	G	W	H	D	Rl+T+D	RI+G+D	T+Sen	Rl+T+Sen	Topic	Rl+D+Topic	RI+T+G+W+H+D+Topic	PC	Rl+D+PC	RI+T+G+W+H+D+PC	RI+T+G+W+H+PC
Rating_lag	0.04						0.07	0.17		0.16		0.01	-0.07		0.00	0.10	0.07
p_value	0.55						0.30	0.00		0.01		0.67	0.89		0.79	0.13	0.41

Tweets		0.03					0.02		0.06	0.08			0.01			0.02	0.03
p_value		<.001					<.001		<.001	<.001			<.001			<.001	<.001
Google			0.00					0.00					0.00			0.00	0.00
p_value			1.00					0.73					0.77			0.82	0.83
Wiki				0.00									0.00			0.00	0.00
p_value				0.08									0.09			0.08	0.10
Huffington					0.07								0.11			0.05	0.06
p_value					0.47								0.39			0.54	0.47
Team Dummy						~	~	~				~	~		~	~	
Tweet_Pos									0.01	0.01							
p_value									0.39	0.26							
Tweet_Neg									0.01	0.00							
p_value									0.70	0.74							
Tweet_PC1														1.25	1.17	0.63	0.85
p_value														<.001	<.001	<.001	<.001
Tweet_PC2														0.48	1.05	1.45	1.56
p_value														<.001	<.001	<.001	<.001
Tweet_PC3														1.06	1.34	0.88	1.00
p_value														<.001	<.001	<.001	<.001
Tweet_PC4														1.04	1.15	1.27	1.29
p_value														<.001	<.001	<.001	<.001
Tweet_T1											0.70	0.72	1.90				
p_value											<.001	<.001	<.001				
Tweet_T2											1.20	1.39	0.75				
p_value											<.001	<.001	<.001				
Tweet_T3											1.59	1.28	1.12				
p_value											<.001	<.001	<.001				
Tweet_T4											1.26	0.84	1.18				
p_value											<.001	<.001	<.001				
Tweet_T5											0.42	1.05	0.67				
p_value											<.001	<.001	<.001				
Premier	1.67	-2.07	1.62	1.63	1.74	1.46	-1.60	1.65	1.62	1.59	1.32	1.00	1.17	1.47	1.59	1.96	1.43
p_value	0.20	0.06	0.17	0.16	0.12	0.23	0.00	0.04	0.17	0.18	0.27	0.40	0.31	0.19	0.17	0.07	0.18
Finale	1.18	0.90	1.24	1.68	1.72	2.17	1.81	1.94	2.01	1.00	1.49	2.16	1.99	1.66	1.83	1.57	1.82
p_value	0.38	0.26	0.21	0.09	0.23	0.07	0.23	0.14	0.21	0.39	0.25	0.20	0.22	0.24	0.23	0.24	0.23
Age	-0.31	-0.28	-0.25	-0.23	-0.32	0.02	-0.20	-0.12	-0.30	-0.20	-0.30	-0.24	-0.26	-0.22	-0.40	-0.28	-0.31
p_value	0.31	0.31	0.46	0.47	0.55	0.96	0.25	0.57	0.47	0.41	0.55	0.45	0.50	0.44	0.83	0.57	0.56
R2**	0.24	0.34	0.03	0.06	0.04	0.68	0.78	0.77	0.36	0.77	0.81	0.83	0.86	0.84	0.85	0.90	0.89
Wald Chi2	62.80	235.76	225.26	247.51	236.98	67.30	127.05	113.32	252.20	246.13	323.45	302.74	314.05	328.15	334.59	331.82	336.41
p-value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.002
AR(1)	-1.98						-2.24	-2.06		-1.97		-2.01	-1.98		-2.40	-2.05	-2.52
p-value	0.05						0.03	0.04		0.05		0.04	0.05		0.02	0.04	0.01
AR(2)	0.05						0.76	-0.32		0.44		0.65	0.84		0.63	0.88	0.81
p-value	0.96						0.45	0.75		0.66		0.52	0.40		0.53	0.38	0.42
Sargan Chi2	124.03						57.17	63.14		105.34		52.35	47.86		45.83	46.24	75.86
p-value	0.43						0.65	0.44		0.81		0.65	0.67		0.88	0.76	1.00
MMSC-BIC	-466.00						-201.64	-236.71		-470.18		-223.32	-208.46		-234.67	-214.92	-470.64

In the case of the NFL (Table 25), we find that the lagged rating is not a good predictor of the current rating (R2=0.24). The reason for this result is probably that the size of the fan base changes because the teams that play the game change each week. For example, if the last week's Sunday Night Football (SNF) game was between the Dallas Cowboys and Pittsburgh Steelers, the rating would be much higher than this week's SNF game between the Jacksonville Jaguars and Tennessee Titans.

Instead, when we use the number of Tweets related to the two teams within 24 hours before the game kicks off (Table 25, column 2) in the model, the R2 becomes 0.34, which is much higher than when we use the lagged rating as the only explanatory variable (column 1). Surprisingly, the number of Google searches, Wikipedia views, and Huffington Post news articles related to the two teams 24 hours before the game starts can only explain 3%, 6% and 4% of the variation in the rating, respectively, and the estimated coefficients are not significant.

To fix the problem of changing teams, we add team dummies (home and away separately) in the 6th specification of the model. The resulting R2 of 0.68 supports our conjecture that the size of the fan base is a very important determinant of ratings. In columns 7 and 8, we combine the lagged rating with team dummies and Tweets or Google searches. Together, they can explain approximately 78% of the variation in ratings.

We confirm that content analysis (using n-gram PCs or topics as features) of Tweets is a very powerful predictor of TV ratings. Most strikingly, the good model fits remain even after we remove the team dummies, as shown in columns 11, 14 and 17 of Table 25. Topics and PCs alone can explain 83% and 84% of the variations in ratings, respectively. If we combine all of the information, including lagged rating, Tweets, Google searches, Wikipedia views, Huffington Post news articles and PCs, almost 90% of the variation can be explained.

This result indicates that team-specific Tweets that capture consumers' intention to watch the shows are lead indicators of the actual future consumption.

2.5.3 Forecasting

After the model is calibrated, we want to test how well it can be used to forecast TV ratings. For this purpose, we use Nielsen ratings data from September to November 2013 as the test sample. These data include 411 episodes[57] of the thirty TV series and 39 NFL games. To test the model performance, we use two different measures: (1) mean absolute percentage error (MAPE) and (2) mean squared error (MSE).

[57] Shows that have already ended are not included, such as Breaking Bad and Allen Gregory.

Table 26 Prediction: TV Series

Model		MAPE	MSE	Model		MAPE	MSE
	1	0.1214	0.0816		1	0.1232	0.0828
	2	0.4652	0.3332		2	0.4641	0.3334
	3	0.4757	0.3423		3	0.4207	0.3370
	4	0.4392	0.3142		4	0.4048	0.3099
	5	0.5003	0.3748		5	0.4833	0.3670
	6	0.5279	0.3891		6	0.5059	0.3914
	7	0.1247	0.0839		7	0.1208	0.0831
	8	0.1099	0.0746		8	0.1135	0.0771
	9	0.4618	0.3325		9	0.4531	0.3136
	10	0.1111	0.0754		10	0.1060	0.0729
	11	0.0849	0.0574		11	0.0790	0.0543
	12	0.0893	0.0614		12	0.0802	0.0552
	13	0.0809	0.0547		13	0.0780	0.0536
	14	0.0810	0.0537		14	0.0779	0.0526
	15	0.0577	0.0443		15	0.0571	0.0435
24 hr	16	0.0709	0.0468	48 hr	16	0.0688	0.0454

Largely consistent with the results in Tables 23 and 24, Table 26 demonstrates that without content analysis (Models 2 to 6), the mere volume of TGWIH performs much worse in predicting ratings than using content analysis (Models 11 to 16).

Table 27 Prediction MAPE of Popular vs. Unpopular Shows

Show	M1 (Only lagged rating) MAPE	M14 (Twitter content) MAPE
Unpopular	0.2853	0.0563
Medium	0.1379	0.0578
Very popular	0.1035	0.0583
Total	0.1214	0.0577

Interestingly, we find that (Table 27) the content information from Tweets can still do a good job in forecasting ratings for the obscure titles provided that the predictive power of the lagged rating is greatly decreased. Specifically, for the five unpopular shows that did not last more than two seasons[58], including Allen Gregory, Charlie's Angels, Hellcats, Harry's Law, and Gary Unmarried, we found that the prediction MAPE for the model only including the lagged ratings is significantly larger than that for the other popular shows. However, the prediction MAPE for the model that includes Twitter content information is not much different from that of the popular shows.

Table 28 Prediction: NFL

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
R2	0.64	0.34	0.03	0.06	0.04	0.68	0.78	0.77	0.36	0.77	0.81	0.83	0.86	0.84	0.85	0.90	0.89
MAPE	0.18	0.34	0.48	0.47	0.49	0.16	0.11	0.12	0.32	0.11	0.09	0.09	0.07	0.07	0.07	0.06	0.05
MSE	0.12	0.24	0.35	0.34	0.35	0.11	0.07	0.08	0.23	0.08	0.06	0.06	0.05	0.05	0.05	0.04	0.04

Similarly, Models 11 through 17 using content-selected Tweets have the smallest prediction errors for the NFL sample (Table 28).

[58] We use the last season of these obscure titles for the prediction.

Table 29: Prediction Comparison with LDA

MAPE (Model\Feature)	Cloud (Hadoop) PCA	LDA
Dynamic Panel Linear Model	0.0577	0.0595
Auto regression X	0.0613	0.0622
Multi-layered feedforward neural networks	0.0664	0.0681
Self-organizing Fuzzy Neural Network (SOFNN) model	0.0640	0.0679

Our Dynamic Panel Linear Model with Cloud PCA outperforms the alternative models (including Auto Regression with Exogenous Independent Variables (Autoregression X), Multi-layered feedforward neural networks and the Self-organizing Fuzzy Neural Network (SOFNN) model) in terms of out-of-sample prediction accuracy.

Table 30: Computational Time Comparison

Time (Minutes)	Cloud (Hadoop) PCA	LDA
	6.2	36.8

Moreover, cloud-based PCA is much faster than LDA. As shown in Table 30, cloud-based PCA takes approximately 6 minutes for the TV series task, while LDA takes approximately 37 minutes. This result shows that leveraging cloud computing can make computation more efficient, which can help advertisers predict demand faster, potentially in real time.

2.6 Conclusions

Our paper shows that easily accessible online information such as Twitter Tweets and Google Trends can be useful for marketers to accurately predict consumer demand for TV shows. We demonstrate the power of employing machine learning, text mining, and cloud computing techniques to process large-scale unstructured data to conduct a structured econometric analysis. We conclude that information from online platforms, if carefully extracted and sifted, can provide timely representation of consumers' intentions. These results have important implications for forecasting purchase/consumption that should be of great interest to all firms.

Our paper has certain limitations. First, the real mechanism between online behavior and offline consumption is not revealed. Our study is not based on a well-grounded theory that explains the whole path of a consumer's consumption experience. Thus, caution should be exercised in interpreting the results. Second, although Twitter has a wide user base, it is relatively more appealing to the young and urban demographic group, which is different from the general US population. For example, our data show a discrepancy in the rankings of TV series based on the volume of Tweets and ratings. NCIS has the highest average rating but the lowest number of Tweets. This result is probably due to a mismatch between the Twitter user population and the fan base for the show. This limitation constrains Twitter's predictive power for consumption targeted at other demographics. Finally, we choose to predict the most popular TV shows, which have relatively stable ratings patterns. Predicting low-rated shows or newly debuted shows may pose significant additional challenges.

Our paper is only a first step in utilizing consumers' online activities to predict their offline consumption. Future research may consider gathering information from more consumer touch points to predict demand and for other non-information goods. More heterogeneous or location-specific analyses can also be performed to predict demand for certain demographics or in a certain local market. Methodologically, other text mining methods could be developed to extract the most useful information for predictions. Another promising venue to explore is to disrupt the current Nielsen Rating system and replace it with a real-time measure of audience size/composition based on Twitter or Facebook conversations. This approach could provide a viable solution to combat some criticisms of Nielsen's accuracy and bias. Researchers might also want to link TV advertising to social media to accurately measure consumers' responses to TV ads in real time. We hope that future research will address these issues.

3 Chapter 3

Overhaul Overdraft Fees: Creating Pricing and Product Design Strategies with Big Data[59]

Xiao Liu, Alan Montgomery, Kannan Srinivasan

Abstract

In 2012, consumers paid an enormous $32 billion overdraft fees. Consumer attrition and potential government regulations to shut down the overdraft service urge banks to come up with financial innovations to overhaul the overdraft fees. However, no empirical research has been done to explain consumers' overdraft incentives and evaluate alternative pricing and product strategies. In this paper, we build a dynamic structural model with consumer monitoring cost and dissatisfaction. We find that on one hand, consumers heavily discount the future and overdraw because of impulsive spending. On the other hand, a high monitoring cost makes it hard for consumers to track their finances therefore they overdraw because of rational inattention. In addition, consumers are dissatisfied by the overly high overdraft fee and close their accounts. We apply the model to a big dataset of more than 500,000 accounts for a span of 450 days. Our policy simulations show that alternative pricing strategies may increase the bank's revenue. Sending targeted and dynamic alerts to consumers can not only help consumers avoid overdraft fees but improve bank profits from higher interchange fees and less consumer attrition. To alleviate the computational burden of solving dynamic programming problems on a large scale, we combine parallel computing techniques with a Bayesian Markov Chain Monte Carlo algorithm. The Big Data allow us to detect the rare event of overdraft and reduce the sampling error with minimal computational costs.

[59] We acknowledge support from the Dipankar and Sharmila Chakravarti Fellowship. All errors are our own.

3.1 Introduction

An overdraft occurs when a consumer attempts to spend or withdraw funds from her checking accounts in an amount exceeding the account's available funds. In the US, banks allow consumers to overdraw their accounts (subject to some restrictions at banks' discretion) and charge an overdraft fee. Overdraft fees have become a major source of bank revenues since banks started to offer free checking accounts to attract consumers. In 2012, the total amount of overdraft fees in the US reached $32 billion, according to Moebs Services[60]. This is equivalent to an average of $178 for each checking account annually[61]. According to the Center for Responsible Lending, US households spent more on overdraft fees than on fresh vegetables, postage and books in 2010[62]. The unfairly high overdraft fee has provoked a storm of consumer outrage and therefore caused many consumers to close the account. The US government has taken actions to regulate these overdraft fees through the Consumer Financial Protection Agency[63] and may potentially shut down the overdraft service[64]. Without overhauling the current overdraft fee, banks encounter the problem of losing valuable customers and possibly totally losing the revenue source from overdrafts.

Financial institutions store massive amounts of information about consumers. The advantages of technology and Big Data enable banks to reverse the information asymmetry (Kamenica, Mullainathan, and Thaler 2011) as they may be able to generate better forecasts about a consumer's financial state than consumers themselves can. In this paper, we extract the valuable information embedded in the Big Data and harness it with structural economic theories to explain consumers' overdraft behavior. The large scale financial transaction panel data allows us to sort through consumers' financial decision making processes and discover rich consumer heterogeneity. As a consequence, we come up with individually customized strategies that can increase both consumer welfare and bank revenue.

In this paper, we aim to achieve two substantive goals. First, we leverage rich data about consumer spending and balance checking to understand the decision process for consumers to overdraw. We address the following research questions. Are consumers fully attentive in monitoring their checking account balances? How great is the monitoring cost? Why do attentive consumers also overdraw? Are consumers dissatisfied because the overdraft fee?

Second, we investigate pricing and new product design strategies that overhaul overdraft fees. Specifically, we tackle these questions. Is the current overdraft fee structure optimal? How will the bank revenue change under alternative pricing strategies? More importantly, what new revenue model can make the incentives of the bank and consumers better aligned? Can the bank benefit from helping consumers make more informed financial decisions, like sending alerts to consumers? If so, what's the optimal alert strategy? How can the bank

---

[60] http://www.moebs.com
[61] According to Evans, Litan, and Schmalensee 2011, there are 180 million checking accounts in the US.
[62] http://www.blackenterprise.com/money/managing-credit-3-ways-overdraft-fees-will-still-haunt-you/
[63] http://banking-law.lawyers.com/consumer-banking/consumers-and-congress-tackle-big-bank-fees.html
[64] http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf

leverage its rich data about consumer financial behaviors to reverse information asymmetry and create targeted strategies?

We estimate the dynamic structural model using data from a large commercial bank in the US. The sample size is over 500,000 accounts and the sample length is up to 450 days. We find that some consumers are inattentive in monitoring their finances because of a substantially high monitoring cost. In contrast, attentive consumers overdraw because they heavily discount future utilities and are subject to impulsive spending. Consumers are dissatisfied to leave the bank after being charged the unfairly high overdraft fees. In our counterfactual analysis, we show that a percentage fee or a quantity premium fee strategy can achieve higher bank revenue compared to the current flat per-transaction fee strategy. Enabled by Big Data, we also propose an optimal targeted alert strategy. The bank can benefit from sending alerts to let consumers spend their unused balances so that the bank can earn more interchange fees. Helping consumers make more informed decisions will also significantly reduce consumer attrition. The targeted dynamic alerts should be sent to consumers with higher monitoring costs and both when they are underspending and overspending.

Methodologically, our paper makes two key contributions. First, we build a dynamic structural model that incorporates inattention and dissatisfaction into the life-time consumption model. Although we apply it to the overdraft context, the model framework can be generalized to analyze other marketing problems regarding consumer dynamic budget allocation, like electricity and cellphone usage.

Second, we estimate the model on Big Data with the help of parallel computing techniques. Structural models have the merit of producing policy invariant parameters that allow us to conduct counterfactual analysis. However, the inherent computational burden prevents it from being widely adopted by industries. Moreover, the data size in a real setting is typically much larger than what's used for research purposes. Companies, in our case a large bank, need to have methods that are easily scalable to generate targeted solutions for each consumer. Our proposed algorithm takes advantage of state-of-the-art parallel computing techniques and estimation methods that alleviate computational burden and reduce the curse of dimensionality.

The rest of the paper is organized as follows. In section 3.2 we first review related literature. Then we show summary statistics in section 3.3 which motivate our model setup. Section 3.4 describes our structural model and we provide details of identification and estimation procedures in section 3.5. Then in sections 3.6 and 3.7 we show estimation results and counterfactual analysis. Section 3.8 concludes and summarizes our limitations.

3.2 Related Literature

A variety of economic and psychological models can explain overdrafts, including full-information pure rational models and limited attention, as summarized by Stango and

Zinman (2014). However, no empirical paper has applied these theories to real consumer spending data. Although Stango and Zinman (2014) had a similar dataset to ours, their focus was on testing whether taking related surveys can reduce overdrafts. We develop a dynamic structural model that incorporates theories of heavy discounting, inattention and dissatisfaction in a comprehensive framework. The model is flexible to address various overdraft scenarios, thus it can be used by policy makers and the bank to design targeted strategies to increase consumer welfare and bank revenue.

Our model inherits from the traditional lifetime consumption model (Modigliani & Brumberg 1954, Hall 1978) but adds two novel features, inattention and dissatisfaction. First of all, a large body of literature in psychology and economics has found that consumers pay limited attention to relevant information. In the review paper by Card, DellaVigna and Malmendier (2011), they summarize findings indicating that consumers pay limited attention to 1) shipping costs, 2) tax (Chetty et. al. 2009) and 3) ranking (Pope 2009). Gabaix and Laibson (2006) find that consumers don't pay enough attention to add-on pricing and Grubb (2014) shows consumers' inattention to their cell-phone minute balances. Many papers in the finance and accounting domain have documented that investors and financial analysts are inattentive to various financial information (e.g., Hirshleifer and Teoh 2003, Peng and Xiong 2006). We follow Stango and Zinman (2014) to define inattention as incomplete consideration of account balances (realized balance and available balance net of coming bills) that would inform choices. We further explain inattention with a structural parameter, monitoring cost (Reis 2006), which represents the time and effort to know the exact amount of money in the checking account. With this parameter estimated, we are able to quantify the economic value of sending alerts to consumers and provide guidance for the bank to set its pricing strategy. We also come up with policy simulations about alerts because we think a direct remedy for consumers' limited attention is to make information more salient (Card, DellaVigna and Malmendier 2011). Past literature also finds that reminders (Karlan et. al. 2010), mandatory disclosure (Fishman and Hagerty 2003), and penalties (Haselhuhn et al. 2012) all serve the purpose of increasing salience and thus mitigating the negative consequences of inattention.

Second, as documented in previous literature, unfairly high price may cause consumer dissatisfaction which is one of the main causes of customer switching behavior (Keaveney 1995, Bolton 1998). We notice that consumers are more likely to close the account after paying the overdraft fee and when the ratio of the overdraft fee over the overdraft transaction amount is high. This is because given the current banking industry practice, a consumer pays a flat per-transaction fee regardless of the transaction amount. Therefore, the implied interest rate for an overdraft originated by a small transaction amount is much higher than the socially accepted interest rate (Matzler, Wurtele and Renzl 2006), leading to price dissatisfaction.

We aim to estimate this infinite horizon dynamic structural model on a large scale of data and obtain heterogeneous best response for each consumer to prepare targeted marketing strategies. After searching among different estimation methods, including the nested fixed point algorithm (Rust 1987), the conditional choice probability estimation (Arcidiacono and Miller 2011) and the Bayesian estimation method developed in Imai, Jain and Ching (2009) (IJC), we finally choose the IJC method for the following reasons. First of all, the hierarchical Bayes framework fits our goal of obtaining heterogeneous parameters. Second, in order to apply our model to a large scale of data, we need to estimate the model with Bayesian MCMC so that we can implement a parallel computing technique. Third, IJC is the state-of-the art Bayesian estimation algorithm for infinite horizon dynamic programming models. It provides two additional benefits in tackling the computational challenges. One is that it alleviates the computational burden by only evaluating the value function once in each MC iteration. Essentially, the algorithm solves the value function and estimates the structural parameters simultaneously. So the computational burden of a dynamic problem is reduced by an order of magnitude similar to those computational costs of a static model. The other is that the method reduces the curse of dimensionality by allowing state space grid points to vary between estimation iterations. On the other hand, as our sample size is huge, traditional MCMC estimation may take a prohibitively, if not impossible, long time, since for N data points, most methods must perform O(N) operations to draw a sample. A natural way to reduce the computation time is to run the chain in parallel. Past methods of Parallel MCMC duplicate the data on multiple machines and cannot reduce the time of burn-in. We instead use a new technique developed by Neiswanger, Wang and Xing (2014) to solve this problem. The key idea of this algorithm is that we can distribute data into multiple machines and perform IJC estimation in parallel. Once we obtain the posterior Markov Chains from each machine, we can algorithmically combine these individual chains to get the posterior chain of the whole sample.

3.3 Background and Model Free Evidence

We obtained data from a major commercial bank in the US. During our sample period in 2012 and 2013, overdraft fees accounted for 47% of the revenue from deposit account service charges and 9.8% of the operating revenue.

The bank provides a comprehensive overdraft solution to consumers. (For general overdraft practices in the US, please refer to Stango and Zinman (2014) for a good review. Appendix A3.1 tabulates current fee settings in top US banks.) In the standard overdraft service, if the consumer overdraws her account, the bank might cover the transaction and charge $31[65] Overdraft Fee (OD) or decline the transaction and charge a $31 Non-Sufficient-Fund Fee (NSF). Whether the transaction is accepted or declined is at the bank's discretion. The OD/NSF fee is at a per-item level. If a consumer performs several transactions when the

[65] All dollar values in the paper have been rescaled by a number between .85 and 1.15 to help obfuscate the exact amounts without changing the substantive implications. The bank also sets the first time overdraft fee for each consumer at $22. All the rest overdraft fees are set at $31.

account is already overdrawn, each transaction item will incur a fee of 31 dollars. Within a day, a maximum of four per-item fees can be charged. If the account remains overdrawn for five or more consecutive calendar days, a Continuous Overdraft Fee of $6 will be assessed up to a maximum of $84. The bank also provides an Overdraft Protection Service where the checking account can link to another checking account, a credit card or a line of credit. In this case, when the focal account is overdrawn, funds can be transferred to cover the negative balance. The Overdraft Transfer Balance Fee is $9 for each transfer. As you can see, the fee structure for the bank is quite complicated. In the empirical analysis below, we don't distinguish between different types of overdraft fees and assume that money is fungible so that the consumer only cares about the total amount of overdraft fee rather than the underlying pricing structure.

The bank also provides balance checking services through branch, automated teller machine (ATM), call center and online/mobile banking. Consumers can inquire about their available balances and recent activities. There's also a notification service to consumers via email or text message, named "alerts". Consumers can set alerts when certain events take place, like overdrafts, insufficient funds, transfers, deposits, etc. Unfortunately, our dataset only includes the balance checking data but not the alert data. We'll discuss this limitation in section 3.8

In 2009, the Federal Reserve Board made an amendment to Regulation E (subsequently recodified by the Consumer Financial Protection Bureau (CFPB)) which requires account holders to provide affirmative consent (opt in) for overdraft coverage of ATM and non-recurring point of sale (POS) debit card transactions before banks can charge for paying such transactions[66]. This Regulation E aimed to protect consumers against the heavy overdraft fees. The change became effective for new accounts on July 1, 2010, and for existing accounts on August 15, 2010. Our sample contains both opt-in and opt-out accounts. However, we don't know which accounts have opted in unless we observe an ATM/POS initiated overdraft occasion. We also discuss this data limitation in section 3.8.

3.3.1 Summary Statistics

Our data can be divided into two categories, checking account transactions and balance inquiry activities. In our sample, there are between 500,000 and 1,000,000[67] accounts, among which 15.8% had at least one overdraft incidence during the sample period between June 2012 and Aug 2013. The proportion of accounts with overdraft is lower than the 27% (across all banks and credit unions) reported by the CFPB in 2012[68]. In total, all the counts performed more than 200 million transactions, including deposits, withdrawals, transfers, and payments etc. For each transaction, we know the account number, transaction date, transaction amount, and transaction description. The transaction description tells us the type of transaction (e.g., ATM withdrawal or debit card purchase) and location/associated

[66] http://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-43.html
[67] For the sake of privacy, we can't disclose the exact number.
[68] http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf.

institution of the transaction, like merchant name or branch location. The description helps us identify the cause of the overdraft, for instance whether it's due to an electricity bill or due to a grocery purchase.

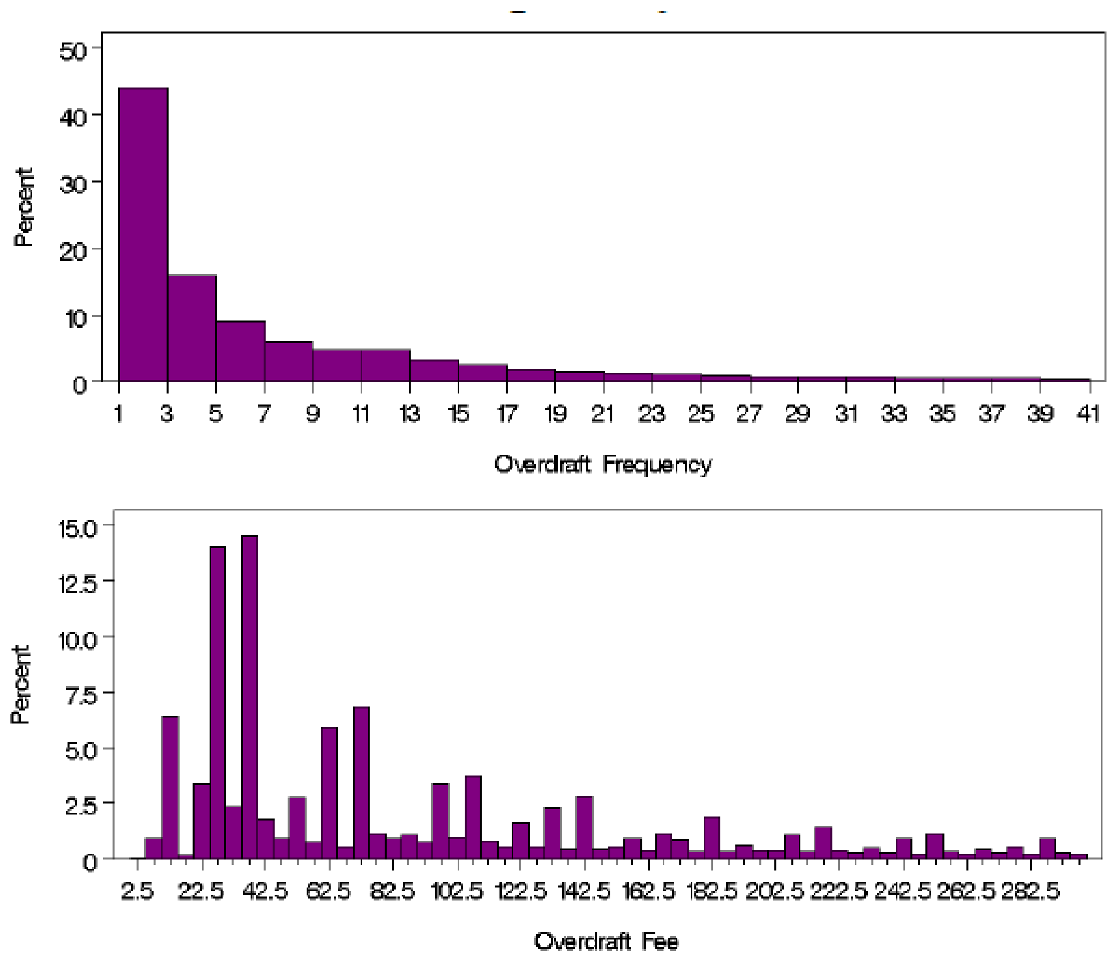
Table 31 Overdraft Frequency and Fee Distribution

	Mean	Std	Median	Min	99.85 Percentile
OD Frequency	9.84	18.74	3	1	>100
OD Fee	245.46	523.04	77	10	>2730

As shown in Table 3, consumers who paid overdraft fees, on average, overdrew nearly 10 times and paid $245 during the 15 month sample period. This is consistent with the finding from the CFPB that the average overdraft- and NSF-related fees paid by all accounts that had one or more overdraft transactions in 2011 were $225 [69]. There is significant heterogeneity in consumers' overdraft frequency and the distribution of overdraft frequency is quite skewed. The median overdraft frequency is three and more than 25% of consumers overdrew only once. In contrast, the top 0.15% of heavy overdrafters overdrew more than 100 times. A similar skewed pattern applies to the distribution of overdraft fees. While the median overdraft fee is $77, the top 0.15% of heaviest overdrafters paid more than $2,730 in fees.

[69] http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf

Figure 14 Overdraft Frequency and Fee Distribution



Now let's zoom in to take a look at the behavior of the majority overdrafters that have overdrawn less than 40 times. The first panel in Figure 14 depicts the distribution of overdraft frequency for those accounts. Notice that most consumers (> 50%) only overdrew less than three times. The second panel shows the distribution of the paid overdraft fee for accounts that have overdrawn less than $300. Consistent with the fee structure where the standard per-item overdraft fee is $22 or $31, we see spikes on these two numbers and their multiples.

Table 32 Types of Transactions That Cause Overdraft

Type	Frequency	Percentage	Amount
Debit Card Purchase	946,049	48.65%	29.50
ACH Transaction	267,854	13.77%	294.57
Check	227,128	11.68%	417.78
ATM Withdrawal	68,328	3.51%	89.77

What types of transactions cause overdraft? We find that nearly 50% of overdrafts are caused by debit card purchases with mean transaction amounts around $30. On the other hand, ACH (Automated Clearing House) and Check transactions account for 13.77% and 11.68% of overdraft occasions. These transactions are generally for larger amounts, $294.57 and $417.78, respectively. ATM withdrawals lead to another 3.51% of the overdraft transactions with an average amount of around $90.

3.3.2 Model Free Evidence

This section presents some patterns in the data that suggest the causes and effects of overdrafts. We show that heavy discounting and inattention may drive consumers' overdraft behaviors. And consumers are dissatisfied because of the overdraft fees. The model free evidence also highlights the variation in the data that will allow for the identification of the discount factor, monitoring cost and dissatisfaction sensitivity.

3.3.2.1 Heavy Discounting

First of all, we argue that a consumer may overdraw because she prefers current consumption much more than future consumption, i.e. she heavily discounts future consumption utility. At the point of sale, the consumer sharply discounts the future cost of the overdraft fee to satisfy immediate gratification[70]. If that's the case, then we should observe a steep downward sloping trend in the spending pattern within a pay period. That is, the consumer will spend a lot right after getting a pay check and then reduce spending during the course of the month. But because of overspending at the beginning, the consumer is going to run out of budget at the end of the pay period and has to overdraw.

We test this hypothesis with the following model specification. We assume that the spending for consumer i at time t $Spending_{it}$ can be modeled as

$$Spending_{it} = \beta * LapsedTimeAfterIncome_{it} + \mu_i + v_t + \epsilon_{it}$$

where $LapsedTimeAfterIncome_{it}$ is the number of days after the consumer received income (salary), μ_i is the individual fixed effect and v_t is the time (day) fixed effect. To control for the effect that consumers usually pay for their bills (utilities, phone bills, credit card bills, etc) after getting the pay check, we exclude checks and ACH transactions which are the common choices for bill payments from the daily spendings and only keep debit card purchases, ATM withdrawals and person-to-person transfers.

We run this OLS regression for heavy overdrafters (whose overdraft frequency is in the top 20 percentile among all overdrafters), light overdrafters (whose overdraft frequency is not in the top 20 percentile among all overdrafters) and non-overdrafters (who didn't overdraw during the 15 months sample period) separately. The results are reported in column (1) (2) and (3) of Table .

[70] We also considered hyperbolic discounting with two discount factors, a short term present bias parameter and a long term discount factor. With more than three periods of data within a pay period, hyperbolic discount factors can be identified (Fang and Silverman 2009). However, our estimation results show that the present bias parameter is not significantly different from 1. Therefore we only keep one discount factor in the current model. Estimation results with hyperbolic discount factors are available upon requests.

Table 33 Spending Decreases with Time in a Pay Cycle

	(1)	(2)	(3)
	Heavy Overdrafters	Light Overdrafters	Non-Overdrafters
Lapsed Time after Income (β)	-6.8374***	-0.00007815	-0.00002195
	(0:00006923)	(0:00006540)	(0:00002328)
Fixed Effect	Yes	Yes	Yes
Number of Observations	17,810,276	53,845,039	242,598,851
R2	0.207	0.275	0.280

Note: *p<0.01;**p<0.001;***p<0.0001

We find that the coefficient of $LapsedTimeAfterIncome_{it}$ is negative and significant for heavy overdrafters but not light overdrafters or non-overdrafters. This suggests that heavy overdrafters have a steep downward sloping spending pattern during a pay period while light overdrafters or non-overdrafters have a relatively stable spending stream. The heavy overdrafters are likely to overdraw because they heavily discount their future consumptions.

3.3.2.2 Inattention

Next we explain the overdraft incentives for the light overdrafters with inattention. The idea is that consumers might be inattentively monitoring their checking accounts so that they are uncertain about the exact balance amount. Sometimes the perceived balance can be higher than the true balance and this might cause an overdraft. We first present a representative example of consumer inattention. The example is based upon our data, but to protect the privacy of the consumer and the merchants, amounts have been changed. However, the example remains representative of the underlying data.

Figure 15 Overdraft due to Balance Perception Error

DATE	+/-	AMOUNT	BALANC	DESCRIPTION
8/17/2012	+	734.11	1705.34	SALARY
8/17/2012	-	535	1170.34	BILLPAY RENT
8/17/2012	-	96.85	1073.49	PURCHASE DEPARTMENT STORE
8/17/2012	-	87	986.49	PURCHASE ELECTRONICS
8/18/2012	-	56.99	929.5	PURCHASE CLOTHING
8/18/2012	-	15.23	914.27	PURCHASE RESTAURANT
8/18/2012	-	585.05	329.22	BILLPAY MORTGAGE
8/19/2012	-	106.3	222.92	PURCHASE HOME
8/19/2012	-	92.52	130.4	PURCHASE GROCERY
8/20/2012	-	38.59	91.81	PURCHASE RESTAURANT
8/20/2012	-	37.13	54.68	PURCHASE ONLINE SPORTS
8/20/2012	-	33.52	21.16	PURCHASE CLOTHING
8/21/2012	-	25	-3.84	**PURCHASE RESTAURANT**
8/21/2012	-	17.12	-20.96	**PURCHASE BEAUTY**
8/22/2012	-	6.31	-27.27	**PURCHASE GAME STORE**
8/22/2012	-	4.95	-32.22	**PURCHASE COFFEE**
8/23/2012	+	180	147.78	ATM DEPOSIT
8/23/2012	-	31	116.78	**OVERDRAFT FEE**
8/23/2012	-	31	85.78	**OVERDRAFT FEE**
8/23/2012	-	31	54.78	**OVERDRAFT FEE**
8/23/2012	-	31	23.78	**OVERDRAFT FEE**

As shown in Figure 15, the consumer first received her salary on August 17th. After a series of expenses she was left with $21.16 on August 20th. As she had never checked her balance, she continued spending and overdrew her account for several small purchases, including a $25 restaurant bill, a $17.12 beauty purchase, a $6.31 game and a $4.95 coffee purchase. These four transactions added up to $53.38 but caused her to pay four overdraft item fees, a total of $124. We speculate that this consumer was careless in monitoring her account and overestimated her balance.

Beyond this example, we find more evidence of inattention in the data. Intuitively, a direct support of inattention is that the less frequent a consumer checks her balance, the more overdraft fee she pays. To test this hypothesis, we estimate the following specification:

$$TotODPmt_{it} = \beta_0 + \beta_1 BCFreq\_it + \mu_i + v_t + \epsilon_{it}$$

where for consumer i at time t (month), $TotODPmt_{it}$ is the total overdraft payment, $BCFreq_{it}$ is the balance checking frequency.

We estimate this model on light overdrafters (whose overdraft frequency is not in the top 20 percentile) and heavy overdrafters (whose overdraft frequency is in the top 20 percentile) separately and report the result in the column (1) and (2) in Table 3.

Table 34 Frequent Balance Checking Reduces Overdrafts for Light Overdrafters

	(1)	(2)	(3)
	Light Overdrafters	Heavy Overdrafters	All Overdrafters
Balance Checking Frequency ($BCFreq, \beta_1$)	-0.5001***	-0.00001389	-0.6823***
	0.00000391	0.00000894	0.00000882
Overdraft Frequency ($ODFreq, \beta_2$)			16.029 4***
			0.00002819
$BCFreq * ODFreq$ (β_3)			27.8136***
			0.00000607
Number of Observations	1,794,835	593,676	2,388,511
R^2	0.1417	0.1563	0.6742

Note: Fixed effects at individual and day level; Robust standard errors, clustered at individual level.*p<0.01;**p<0.001;***p<0.0001

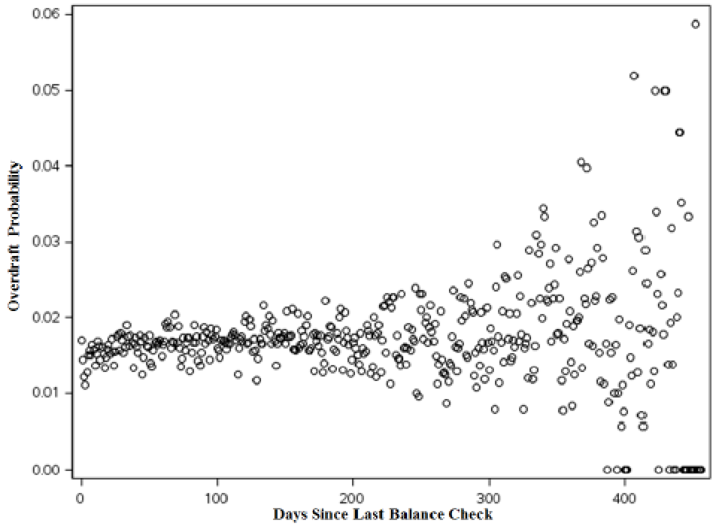
The result suggests that more balance checking decreases overdraft payment for light overdrafters but not for heavy overdrafters. We further test this effect by including overdraft frequency ($ODFreq_{it}$) and an interaction term of balance checking frequency and overdraft frequency $BCFreq_{it} * ODFreq_{it}$ in the equation below. The idea is that if the coefficient for this interaction term is positive while the coefficient for balance checking frequency ($BCFreq_{it}$) is negative, then it implies that checking balances more often only decreases the overdraft payment for consumers who overdraw infrequently but not for those who do it with high frequency.

$$TotODPmt_{it} = \beta_0 + \beta_1 BCFreq_{it} + \beta_2 ODFreq_{it} + \beta_3 BCFreq_{it} * ODFreq_{it} + \mu_i + v_t + \epsilon_{it}$$

The result in column (3) of Table 3 confirms our hypothesis.

Interestingly, we find that a consumer's balance perception error accumulates overtime in the sense that the longer a consumer hasn't check balances, the more likely that she is going to overdraw and pay higher amount of overdraft fees. Figure 16 below exhibits the overdraft probability across number of days since a consumer checked balance last time for light overdrafters (whose overdraft frequency is not in the top 20 percentile). It suggests that the overdraft probability increases moderately with the number of days since the last balance check.

Figure 16 Overdraft Likelihood Increases with Lapsed Time Since Last Balance Check



We confirm this relationship with the following two specifications. We assume that overdraft incidence $I(OD)_{it}$ (where $I(OD)_{it} = 1$ denotes overdraft and $I(OD)_{it} = 0$ denotes no overdraft) and overdraft fee payment amount $ODFee_{it}$ for consumer i at time t can be modeled as:

$$I(OD)_{it} = \Phi(\rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} + \rho_2 BeginBal_{it} + \mu_i + v_t)$$

$$ODFee_{it} = \rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} + \rho_2 BeginBal_{it} + \mu_i + v_t + \epsilon_{it}$$

where Φ is the cumulative distribution function for standard normal distribution. The term $DaysSinceLastBalanceCheck_{it}$ denotes the number of days consumer i hasn't checked her balance until time t and $BeginBal_{it}$ is the beginning balance at time t. We control for the beginning balance because it can be negatively correlated with the days since last balance check due to the fact that consumers tend to check when the balance is low and a lower balance usually leads to an overdraft.

Table 35 Reduced Form Evidence of Existance of Monitoring Cost

	I (OD)	ODFee
Days Since Last Balance Check (ρ_1)	0.0415***	0.0003***
	(0.00000027)	(0.00000001)
Beginning Balance (ρ_2)	-0.7265***	-0.0439***
	(0.00000066)	(0.00000038)
Individual Fixed Effect	Yes	Yes
Time Fixed Effect	Yes	Yes
Number of Observations	53,845,039	53,845,039
R2	0.5971	0.6448

Note: The estimation sample only includes overdrafters. Marginal effects for the Probit model; Fixed effects at individual and day level; robust standard errors, clustered at individual level.*p<0.01;**p<0.001;***p<0.0001.

Table 3 reports the estimation results which support our hypothesis that the longer a consumer hasn't checked balance, the more likely she overdraws and the higher overdraft fee she pays.

Since checking balances can effectively help prevent overdrafts, why don't consumers do it often enough to avoid overdraft fees? We argue that it's because monitoring the account is costly in terms of time, effort and mental resources. Therefore, a natural consequence is that if there's a means to save consumers' time, effort or mental resources, the consumer will indeed check balances more frequently. We find such support from the data about online banking ownership. Specifically, for consumer i we estimate the following specification:

$$CheckBalFreq_i = \beta_0 + \beta_1 OnlineBanking_i + \beta_2 LowIncome_i + \beta_3 Age_i + \epsilon_i$$

where $CheckBalFreq_i$ is the balance checking frequency, $OnlineBanking_i$ is online banking ownership (1 denotes the consumer has online banking while 0 denotes otherwise), $LowIncome_i$ is whether the consumer belongs to the low income group (1 denotes yes and 0 denotes no) and Age_i is age (in years).

Table 36 Reduced Form Evidence of Existence of Monitoring Cost

Dependent variable	Check Balance Frequency
Online Banking (β_1)	58.4245***
	0.5709
Low Income (β_2)	3.3812***
	0.4178
Age (β_3)	0.6474***
	0.0899
Number of Observations	602.481
R2	0.6448

*p<0.01;**p<0.001;***p<0.0001.

Table 3 shows that after controlling for income and age, consumers with online banking accounts check the balance more frequently than those without, which suggests that monitoring costs exist and when they are reduced, consumers monitor more frequently.

3.3.2.3 Dissatisfaction

Table 37 Account Closure Frequency for Overdrafters vs Non-Overdrafters

Total %	Closed
Heavy Overdrafters	23.36%
Light Overdrafters	10.56%
Non-Overdrafters	7.87%

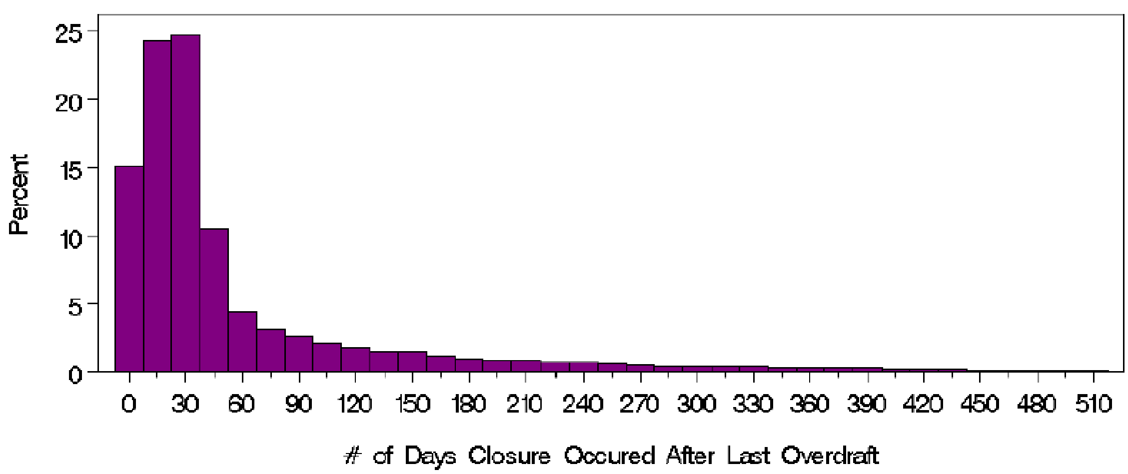
We also find that overdrafters are more likely to close their accounts (Table 37). Among non-overdrafters, 7.87% closed their accounts during the sample period. This ratio is much higher for overdrafters. Specifically, 23.36% of heavy overdrafters (whose overdraft frequency is in the top 20 percentile) closed their accounts, while 10.56% of light overdrafters (whose overdraft frequency is not in the top 20 percentile) closed their accounts.

Table 38 Closure Reasons

	Overdraft Forced Closure	Overdraft Voluntary Closure	No Overdraft Voluntary Closure
Heavy Overdrafters	86.34%	13.66%	-
Light Overdrafters	52.58%	47.42%	-
Non-Overdrafters	-	-	100.00%

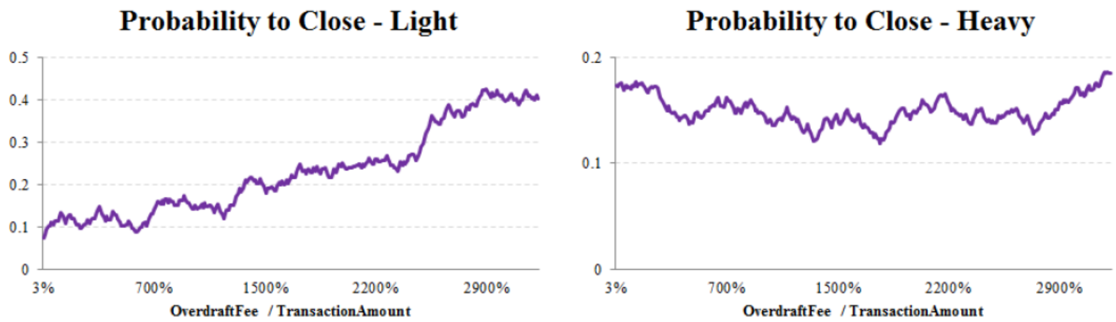
From the description field in the data, we can distinguish the cause of account closure: forced closure by the bank because the consumer is unable or unwilling to pay back the negative balances and the fee (charge-off) or voluntary closure by the consumer. Among heavy overdrafters, 13.66% closed voluntarily and the rest (86.34%) were forced to close by the bank (Table 3). In contrast, 47.42% of the light overdrafters closed their accounts voluntarily. We conjecture that the higher voluntary closures may be due to customer dissatisfaction with the bank, with evidence shown below.

Figure 17 Days to Closure After Last Overdraft



First, we find that overdrafters who closed voluntarily were very likely to close soon after the overdraft. In Figure 17 we plot the histogram of number of days it took the account to close after its last overdraft occasion. It shows that more than 60% of accounts closed within 30 days after the overdraft occasion.

Figure 18 Percentage of Accounts Closed Increases with Fee/Transaction Amount Ratio



Second, light overdrafters are also more likely to close their accounts when the ratio of overdraft fee over the transaction amount that caused the overdraft fee is higher. In other words, the more unfair the overdraft fee (higher ratio of overdraft fee over the transaction amount that caused the overdraft fee), the more likely it is that she will close the account. We show this pattern in the left panel of Figure 18. However, this effect doesn't seem to be present for heavy overdrafters (right panel of Figure 18).

The model free evidence indicate that consumer heavy discounting and inattention can help explain consumers' overdraft behaviors as consumers might be dissatisfied after being charged the overdraft fees. Below we'll build a structural model that incorporates consumer heavy discounting, inattention and dissatisfaction.

3.4 Model

We model a consumer's daily decision about non-preauthorized spending in her checking account. Alternatively we could describe this non-preauthorized spending as immediate or discretionary; not discretionary in the sense that economists traditionally use the term, but in the sense that immediate spending likely could have been delayed. To focus on rationalizing the consumer's overdraft behavior, we make the following assumptions. First, we abstract away from the complexity associated with our data and assume that the consumer's income and preauthorized spendings are exogenously given. We refer to preauthorized spending to mean those expenses for which the spending decision was made prior to payment. For example, a telephone bill or a mortgage due are usually arranged before the date that the actual payment occurs. We assume that decisions for preauthorized spending are hard to change on a daily basis after they are authorized and more likely to be related to consumption that has medium or long-run consequences. In contrast, non-preauthorized spending involves a consumer's frequent day-to-day decisions and the consumer can adjust the spending amount flexibly. We make this distinction because non-preauthorized spending is at the consumer's discretion and thus affects the overdraft outcome directly. To ease explanation, we use "coming bills" to represent preauthorized spending for the rest of the paper. Second, we allow the consumer to be inattentive to monitoring her account balance and coming bills. But she can decide whether to check her balance. When a consumer hasn't checked the balance, she comes up with an estimate of the available balance and forms an expectation about coming bills. If she makes a wrong estimate or expectation, she faces the risk of overdrawing her account. Last, as consumption is not observed in the data, we make a bold assumption that spending is equivalent to consumption in terms of generating utility. That is, the more a consumer spends, the more she consumes, the higher utility she obtains. In what follows, we use consumption and spending interchangeably.

We'll describe the model in the next four parts: (1) timing, (2) basic model (3) inattention and balance checking and (4) dissatisfaction and account closing.

3.4.1 Timing

The timing of the model is as follows (Figure 19). On each day:

1. The consumer receives income, if there is any.

2. Her bills arrive if there is any.

3. Balance checking stage (CB): She decides whether to check her balance. If she checks, she incurs a cost and knows today's beginning balance and the bill amount. If not, she recalls an estimate of the balance and bill amount.

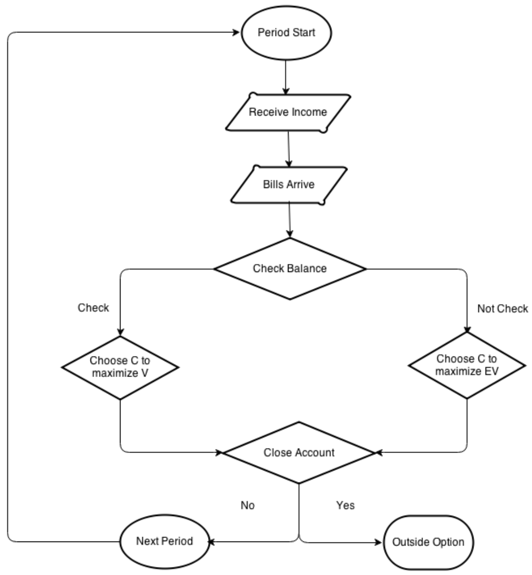
4. Spending stage (SP): She makes the discretionary spending decision (Choose C) to maximize total discounted utility V (or expected total discounted utility EV if she didn't check balance)for today and spends the money.

5. Overdraft fee is charged if the ending balance is below zero.

6. Account closing stage (AC): She decides whether to close the account (after paying the overdraft fee if there's any). If she closes the account, she receives an outside option. If she doesn't choose the account, she goes to 7.

7. Balance updates and the next day comes.

Figure 19 Model Timing



3.4.2 Basic Model

Following the lifetime consumption literature (Modigliani & Brumberg 1954, Hall 1978), we assume the consumer's per-period consumption utility at time t is a constant relative risk averse utility (Arrow 1963):

$$u_C(C_t) = \frac{C_t^{1-\theta}}{1-\theta_t} \tag{1}$$

where θ_t is the relative risk averse coefficient which represents the consumer's preference about consumption. The higher θ_t , the higher utility the consumer can derive from a marginal unit of consumption.

$$\theta_t = exp(\theta + \varepsilon_t)$$

$$\varepsilon_t \sim N(0, \varsigma^2)$$

As consumers' preference for consumption might change over time and the relative risk averse coefficient is always positive, we allow θ_t to follow a log-normal distribution. Essentially, θ_t is the exponential of the sum of a time-invariant mean θ and a random shock ε_t. The shocks capture unexpected needs for consumption and follow a normal distribution with mean 0 and variance ς^2 (Yao et. al. 2012).

Notice that the consumption plan C_t depends on the consumer's budget constraint, which further depends on her current balance B_t , income Y_t and future bills Ψ_t . For example, when the coming bill is for a small amount, the consumption can be higher than when the bill is for a large amount.

3.4.3 Inattention and Balance Checking

In practice, the consumer may not be fully attentive to her financial well-being. According to the theory of rational inattention (Sims 1998, 2003), individuals have many things to think about and limited time, they can devote only limited intellectual resources to these tasks of datagathering and analysis. Because monitoring her account balance takes time and effort, she may not check her balance frequently. As a consequence, instead of knowing the exact (available) balance B_t[71], she recalls a perceived balance $\widetilde{B_t}$. Following Mehta, Rajiv and Srinivasan (2003), we allow the perceived balance $\widetilde{B_t}$ to be the sum of the true balance B_t and a perception errorn $\eta_t \omega_t$. The first component of the perception error η_t is a random draw from the standard normal distribution[72] and the second component is the standard deviation of the perception error, ω_t . So $\widetilde{B_t}$ follows a normal distribution

$$\widetilde{B_t} \sim N(B_t + \eta_t \omega_t, \omega_t^2)$$

The variance of the perception error ω_t^2 measures the extent of uncertainty. Based on the evidence from section 3.3.2.2, we allow this extent of uncertainty to accumulate through time which implies that the longer the consumer goes without checking her balance, the more inaccurate her perceived balance is. That is,

$$\omega_t^2 = \rho \Gamma_t \tag{2}$$

where Γ_t denotes the lapsed time since the consumer last checked her balance, and ρ denotes the sensitivity to lapsed time as shown in the equation (2) above[73]. Notice that the expected utility is decreasing in the variance of the perception error ω_t^2 . This is true because the larger the variance of the perception error, the less accurate the consumer's estimate of her true balance, and the more likely she is going to mistakenly overdraw, which lowers her utility.

We further assume that the consumer is sophisticated inattentive[74] in the sense that she is aware of her own inattention (Grubb 2014). Sophisticated inattentive consumers are rational in that they choose to be inattentive due to the high cost of monitoring her balances from day-to-day. We also model the consumer's balance checking behavior. We denote the balance checking choice as $Q_t \in \{1,0\}$ where 1 means check and 0 otherwise. If a consumer checks her balance, she incurs a monitoring cost but knows exactly what her balance is. So

[71] Available balance means the initial balance plus income minus bills. For the ease of exposition, we omit the word "available" and only use "balance".

[72] The mean balance perception error $\bar{\eta}$ cannot be separately identified from the variance parameters ρ because the identification sources both come from consumers' overdraft fee payment. Specifically, the high overdraft payment for a consumer can be either explained by a positive balance perception error or large perception error variance caused by large ρ . So we fix $\bar{\eta}$ at zero, i.e. the perception error is assumed to be unbiased.

[73] We considered other specifications for the relationship between perception error variance and lapsed time since last balance check. Results remain qualitatively unchanged

[74] Consumers can also be naively inattentive, but we don't allow it here. See discussion in Grubb 2014.

the perception error is reduced to zero and she can make her optimal spending decision with all information. In mathematics form, her consumption utility function changes to

$$u_t = \frac{C_t^{1-\theta}}{1-\theta_t} - Q_t\xi + \chi_{tQ_t} \quad (3)$$

where ξ is her balance checking cost and χ_{Q_t} is the idiosyncratic shock that affects her balance checking cost. The shock χ_{tQ_t} can come from random events like a consumer checks balance because she's also performing other types of transactions (like online bill payments) or she is on vacation without access to any bank channels so it's hard for her to check balances. The equation implies that if the consumer checks her balance, then her utility decreases by a monetary equivalence of $|(1-\theta_t)\xi|^{\frac{1}{1-\theta_t}}$. We assume that χ_{tQ_t} are iid and follow a type I extreme value distribution.

If she doesn't check, she recalls her balance $\widetilde{B_t}$ with the perception error η_t. So her perceived balance is

$$\widetilde{B_t} \sim Q_t B_t + (1-Q_t) N(B_t + \eta_t\omega_t, \omega_t^2)$$

She forms an expected utility based on her knowledge about the distribution of her perception error. The optimal spending will maximize her "expected" utility after integrating out the balance perception error, which is

$$u_t = \int_{\widetilde{B_t}} \int_{\eta_t} u_t(C_t; \widetilde{B_t}) dF(\eta_t) dF(\widetilde{B_t})$$

3.4.4 Dissatisfaction and Account Closing

We assume that the consumer also has the option of closing the account (e.g., an "outside option"). If she chooses to close the account, she might switch to other competing banks or become unbanked. With support from section 3.3.2.3, we make an assumption that consumers are sensitive to the ratio of the overdraft fee to the overdraft transaction amount and we use Ξ_{it} to denote this ratio as a state variable. We assume that the higher the ratio, the more likely it is that the consumer will be dissatisfied to close the account because the forward-looking consumer anticipates that she's going to accumulate more dissatisfaction (as well as lost consumption utility due to overdrafts) in the future so that it's not beneficial for her to keep the account open any more. Furthermore, we assume that the consumer keeps updating her belief of the ratio and only remembers the highest ratio that has ever incurred. That is, if we use Δ_t to denote the per-period ratio then

$$\Delta_t = \frac{OD_t}{|B_t - C_t|}$$

and

$$E[\Xi_t + 1|\Xi_t] = max(\Xi_t, \Delta_t)$$

This assumption reflects a consumer's learning behavior over time in the sense that after experiencing many overdrafts, a consumer realizes how costly (or dissatisfied) it could be for her to keep the account open. When she learns that the ratio can be high enough so that it's not beneficial for her to keep the account open any more, she'll choose to close the account. Specifically, we add the dissatisfaction effect to the per-period utility function where

$$U_t = u_t - \Upsilon * \Delta_t * I[B_t - C_t < 0]$$

In the above equation, u_t is defined in equation (3) and Υ is the dissatisfaction sensitivity, i.e., the impact of charging an overdraft fee on a consumer's decision to close the account.

We assume that closing the account is a termination decision. Once a consumer chooses to close the account, her value function (or total discounted utility function) equals an outside option with a value normalized to 0 for identification purposes[75]. If the consumer keeps the account open, she'll receive continuation values from future per-period utility functions. More specifically, let W denote the choice to close the account, where $W = 1$ is closing the account and $W = 0$ is keeping the account open. Then the value function for the consumer becomes

$$V_t = \begin{cases} U_t + \varpi_{t0} + \beta\, E[V_{t+1}\,|S_t], & W_t = 0 \\ U_t + \varpi_{t1}, & W_t = 1 \end{cases}$$

where ϖ_{t0} and ϖ_{t1} are the idiosyncratic shocks that determine a consumer's account closing decision. Sources of the shocks may include (1) the consumer moved address; (2) competing bank entered the market, and so on. We assume these shocks follow a type I extreme value distribution.

3.4.5 State Variables and the Transition Process

We have explained the following state variables in the model: (beginning) balance B_t, income Y_t, coming bill ψ_t, lapsed time since last balance check Γ_t, overdraft fee OD_t, ratio of overdraft fee to the overdraft transaction amount Ξ_t, preference shock ε_t, balance checking cost shock χ_t and account closure utility shock ϖ_t. The other state variable to be introduced later, L_t, is involved in the transition process.

For (available) balance B_t, the transition process satisfies the consumer's budget constraint, which is

$$B_{t+1} = B_t - C_t - OD_t * I(B_t - C_t < 0) + Y_{t+1} - \psi_{t+1}$$

[75] Although the outside option is normalized to zero for all consumers, the implicit assumption is that we allow for heterogeneous utility of the outside option. The heterogeneity is reflected by the other structural parameters, including the dissatisfaction sensitivity.

where OD_t is the overdraft fee. As we model the consumer's spending decision at the daily level rather than transaction level, we aggregate all overdraft fees paid and assume the consumer knows the per-item fee structure stated in section 3.3. This assumption is realistic in our setting because we have already distinguished between inattentive and attentive consumers. The argument that a consumer might not be fully aware of the per-item fee is indirectly captured by the balance perception error in the sense that the uncertain overdraft fee is equivalent to the uncertain balance because they both tighten the consumer's budget constraint. As for the attentive consumer who overdraws because of heavy discounting, she should be fully aware of the potential cost of overdraft. So in both cases we argue that the assumption of a known total overdraft fee is reasonable.

The state variable OD_t is assumed to be iid over time and to follow a discrete distribution with support vector and probability vector $\{X, p\}$. The support vector contains multiples of the per-item overdraft fee.

Consistent with our data, we assume an income distribution as follows

$$Y_t = Y * I(DL_t = PC)$$

where Y is the stable periodic (monthly/weekly/biweekly) income, DL_t is the number of days left until the next payday and PC is the length of the pay cycle. The transition process of DL is deterministic $DL_{t+1} \ = DL_{t-1} + PC * I(DL_t = 1)$ where it decreases by one for each period ahead and goes back to the full length when one pay cycle ends.

The coming bills are assumed to be iid draws from a compound Poisson distribution with arrival rate ϕ and jump size distribution , $\Psi_t \sim CP(\phi, G)$. This distribution can capture the pattern of bills arriving randomly according to a Poisson process and bill sizes are sums of fixed components (each separate bill).

The time since last checking the balance also evolves deterministically based on the balance checking behavior. Formally, we have

$$\Gamma_{t+1} = 1 + \Gamma_t(1 - Q_t)$$

which means that if the consumer checks her balance in the current period, then the lapsed time goes back to 1 but if she doesn't check, the lapsed time accumulates by one more period.

The ratio of the overdraft fee to the overdraft transaction amount evolves by keeping the maximum amount over time.

$$E[\Xi_{t+1}|\Xi_t] = max(\Xi_t, \Delta_t)$$

The shocks ε_t , χ_t and ϖ_t are all assumed to be iid over time.

In summary, the whole state space for consumer is $S_t = \{\widetilde{B_t}, \Psi_t, Y_t, DL_t, OD_t, \Gamma_t, \Xi_t, \varepsilon_t, \chi_t, \varpi_t\}$. In our dataset, we observe $\widehat{S_t} = \{B_t, \psi_t, Y_t, DL_t, OD_t, \Gamma_t, \Xi_t\}$ and our unobservable state variables are $\widetilde{S_t} = \{\widetilde{B_t}, \eta_t, \varepsilon_t, \chi_t, \varpi_t\}$. $S_t = \widehat{S_t} \cup \widetilde{S_t} \cap \{ B_t, \psi_t\}$. Notice here that consumers also have unobserved states B_t and ψ_t due to inattention, which means that the consumer doesn't know the true balance (B_t) or the bill amount (ψ_t) if she doesn't check her balance but only the perceived balance ($\widetilde{B_t}$) and expected bill (Ψ_t).

3.4.6 The Dynamic Optimization Problem and Intertemporal Tradeoff

The consumer chooses an infinite sequence of decision rules $\{C_t, Q_t, W_t\}_{t=1}^{\infty}$ in order to maximize the expected total discounted utility:

$$\max_{\{C_t, Q_t, W_t\}_{t=1}^{\infty}} E_{\{S_t\}_{t=0}^{\infty}} \left\{ U_0(C_0, Q_0, W_0; S_0) + \sum_{t=1}^{\infty} \beta^t U_t(C_t, Q_t, W_t; S_t) \,|S_0 \right\}$$

where $U_t(C_t, Q_t, W_t; S_t) = \left[\int_{\widetilde{B_t}} \int_{\eta_t} \left(\frac{C_t^{1-\theta}}{1-\theta_t} - Q_t\xi + \chi_{tQ_t}\right) dF(\eta_t) dF\left(\widetilde{B_t}\right) - \Upsilon * \frac{OD_t}{|B_t - C_t|} * I[B_t - C_t < 0]\right] + \varpi_{t0}(1 - W_t) + \varpi_{t1} W_t.$

Let $V(S_t)$ denote the value function:

$$V(S_t) = \max_{\{C_\tau, Q_\tau, W_\tau\}_{\tau=t}^{\infty}} E_{\{S_\tau\}_{\tau=t+1}^{\infty}} \left\{ U_t(S_t) + \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} U_\tau(S_\tau) \,|S_t \right\} \tag{4}$$

according to Bellman (1957), this infinite period dynamic optimization problem can be solved through the Bellman Equation

$$V(S_t) = \max_{C,Q,W} E_{S_{t+1}} \{ U(C, Q, W; S_t) + \beta V(S_{t+1}) | S_t \} \tag{5}$$

In the infinite horizon dynamic programming problem, the policy function doesn't depend on time. So we can eliminate the time subscript. Then we have the following choice specific value function:

$$v\big(C,Q,W;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\big)$$

$$=\begin{cases} u_C(C)-\xi+\chi_1-\Upsilon*\dfrac{OD*I[B-C<0]}{|B-C|}+\varpi_0 \\ +\beta\, E_{S_{+1}}[V\big(\tilde{B}_{+1},\Psi_{+1},Y_{+1},DL_{+1},OD_{+1},\Gamma_{+1},\Xi_{+1},\varepsilon_{+1},\chi_{+1},\varpi_{+1}\big)], & if Q=1\ and\ W=0 \\ \displaystyle\int_{\tilde{B}}\int_{\eta}(u_C(C)+\chi_0)dF(\eta)dF\big(\tilde{B}\big)-\Upsilon*\dfrac{OD*I[B-C<0]}{|B-C|}+\varpi_0 & if\ Q=1\ and\ W=0 \\ \varpi_1, & if\ W=1 \end{cases}$$

where subscript+1 denotes the next time period. So the optimal policy is given by the following solution

$$\{C^*,Q^*,W^*\}=argmaxv\big(C,Q,W;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\big)$$

One thing that's worth noticing is that there's a distinction between this dynamic programming problem and traditional ones. Because of the perception error, the consumer observes $\tilde{B}_t=B_t+\eta_t\omega_t$ but doesn't know B_t or η_t. She only knows the distribution $(B_t+\eta_t\omega_t,\omega_t^2)$. The consumer makes a decision $C^*(\tilde{B}_t)$

based on the perceived balance $\tilde{B}_t$. But as researchers, we don't know the realized perception error η_t. We observe the true balance B_t and the consumer's spending $C^*(\tilde{B}_t)$. So we can only assume $C^*(\tilde{B}_t)$ maximizes the "expected ex-ante value function". Later we look for parameters such that the likelihood for $C^*(\tilde{B}_t)$ maximizes the expected ex-ante value function attains maximum. Following Rust (1987), we obtain the ex-ante value function which integrates out the cost shocks, preference shocks, account closing shocks and unobserved mean balance error.

$$EV(B,\psi,Y,DL,OD,\Gamma,\Xi)$$
$$=\int_{\varpi}\int_{\chi}\int_{\varepsilon}\int_{\eta}v\big(C^*,Q^*,W^*;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\big)\,d\eta d\,\varepsilon d\chi d\varpi$$

Consumers' intertemporal trade-offs are associated with the three dynamic decisions. First of all, given the budget constraint, a consumer will evaluate the utility of spending (or consuming) today versus tomorrow. The higher amount she spends today, the lower amount she can spend tomorrow. So spending is essentially a dynamic decision and the optimal choice for the consumer is to smooth out consumption over the time. Second, when deciding when to check balance, the consumer will compare the monitoring cost with the expected gain from avoiding the overdraft fee. She'll only check when the expected overdraft fee is higher than her monitoring cost. As the consumer's balance perception error might accumulate with time, the consumer's overdraft probability also increases with the lapse time since the last balance check. As a result, the consumer will wait until the overdraft probability reaches the certain threshold (when the expected overdraft fee equals the monitoring cost) to check the balance. Finally, the decision to close the account is an optimal stopping problem. The consumer will compare the total discounted utility of keeping the account with the utility from the outside option to decide when to close the account. When expecting too much overdraft fees as well as the accompanied dissatisfaction, the consumer will find it more attractive to take the outside option and close the account.

3.4.7 Heterogeneity

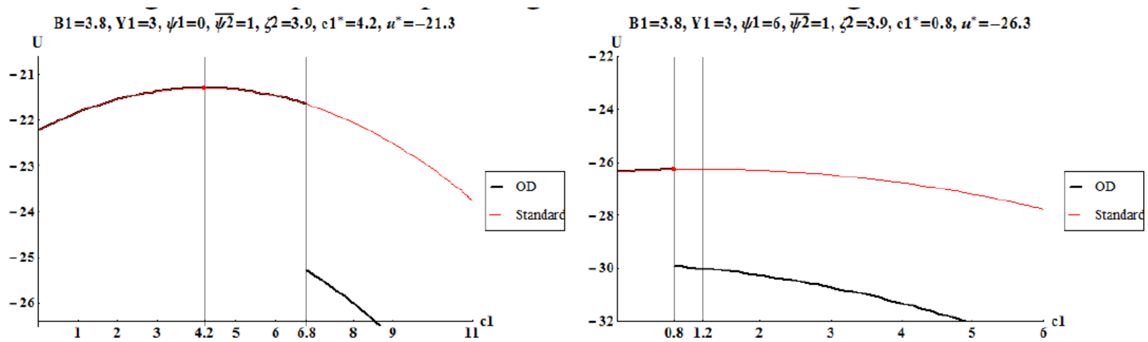
In our data, consumers exhibit different responses to their state conditions. For example, some consumers have never checked their balances and frequently overdraw while other consumers frequently check their balances and rarely overdraw. We hypothesize that it's due to their heterogeneous discount factors and monitoring costs. Therefore, our model needs to account for unobserved heterogeneity. We follow a hierarchical Bayesian framework (Rossi, McCulloch and Allenby 2005) and incorporate heterogeneity by assuming that all parameters: β_i (discount factor), ς_i (standard deviation of risk averse coefficient), ξ_i (monitoring cost), ρ_i (sensitivity of error variance to lapsed time since last checking balance) and Υ_i (dissatisfaction sensitivity) have a random coefficient specification. For each of these parameters, $\vartheta \in \{\beta_i, \varsigma_i, \lambda_i, \xi_i, \rho_i\}$, the prior distribution is defined as $\vartheta \sim N(\mu_\vartheta, \sigma_\vartheta^2)$. The hyper-prior distribution is assumed to be diffuse.

3.4.8 Numerical Example

Here we use a numerical example to show that inattention can explain the observed overdraft occasions in the data. More importantly, we display an interesting case in which an unbiased perception can make the consumer spend less than the desired level. In this example, there are two periods, $t \in \{1,2\}$. The consumer chooses the optimal consumption to maximize the expected total discounted utility. In order to obtain an analytical solution for the optimal spending, we assume a CARA utility $u_C(C_t) = \frac{1}{\theta} exp(-\theta\, C_t)$ and the coming bill following a normal distribution $\Psi_2 \sim N(\overline{\psi_2}, \zeta_2^2)$. The initial balance is B_1 and the consumer receives income Y_1 and Y_2. As period 2 is the termination period, the consumer will spend whatever is left from period 1, i.e., $C_2 = B_1 + Y_1 - \psi_1 - C_1 - OD * (B_1 + Y - \psi_1 - C_1 - \psi_2) + Y_2 - \psi_2$. So the only decision is how much to spend for period 1: C_1. Let $= 0.07$, $B_1 = 3.8$, $Y_1 = 3$, $Y_2 = 3$, $\overline{\psi}_2 = 1$, $\zeta_2 = 3.9$, $\beta = 0.99$, $OD = 3.58$ (The values seem small compared to spending in reality because we apply log to all monetary values).

3.4.8.1 Effect of Overdraft

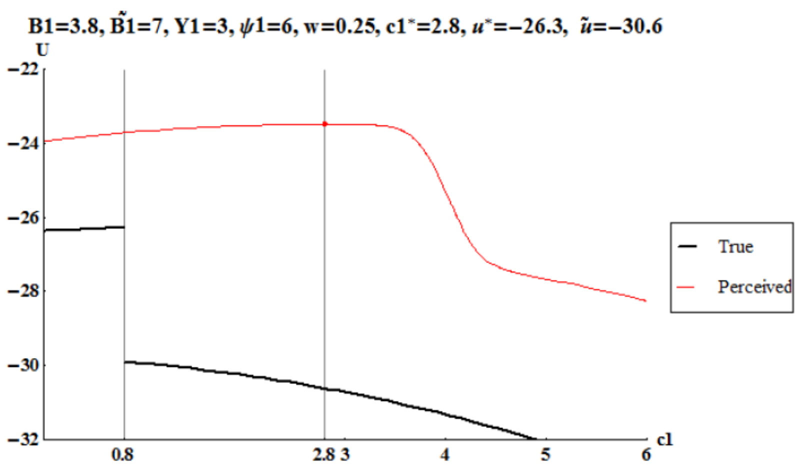
Figure 20 Optimal Spending with Neutral vs Negative Shock



In this example in Figure 20, when there's no bill to pay in the first period ($\psi_1 = 0$ in the left panel), the total budget for the consumer is 6.8 and she would like to spend 4.2 to attain the maximum utility. However, when she has to pay for a bill of 6 (right panel), she is left with only 0.8. Her optimal choice is to spend 0.8 and just clear the budget because the disutility of overdraft (utility function with overdraft is the black line labeled as OD) is too high. This example shows that since the overdraft fee is equivalent to an extremely high interest rate short-term loan, the consumer wouldn't want to overdraw her account.

3.4.8.2 Effect of Inattention--Overdraft

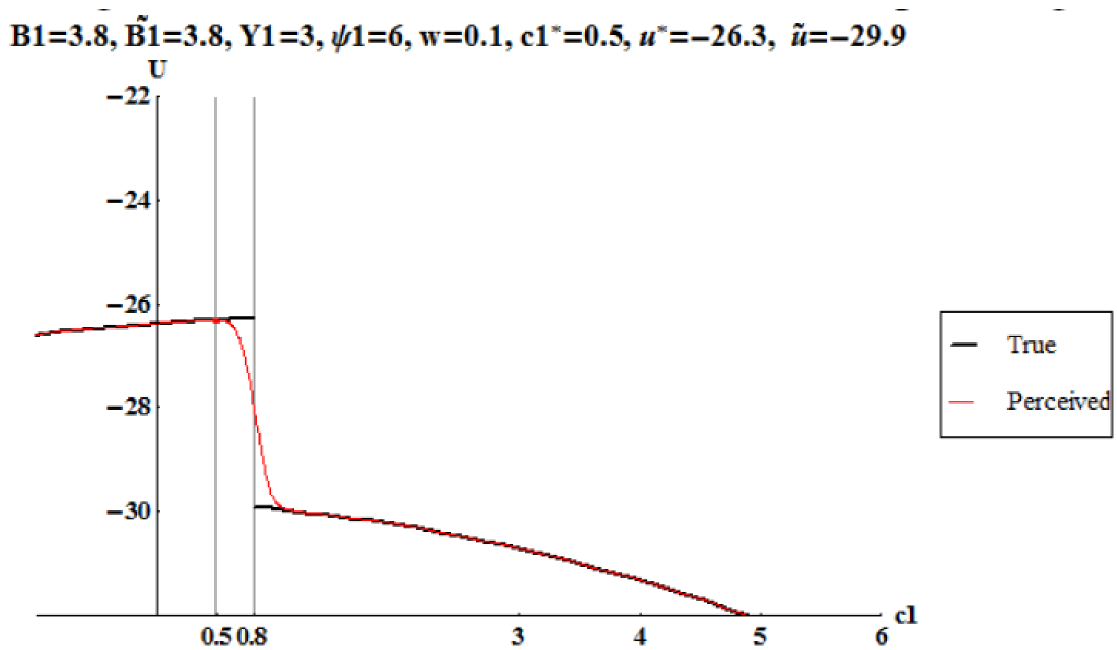
Figure 21 Inattention Leads to Overdraft--Balance Error (B_1) ̃>B_1



In a different scenario (Figure 21), if the consumer overestimates her balance to be 7 (her true balance is 3.8), i.e., she has a positive perception error regarding her true balance, then she would spend 2.8 which is the optimal amount based on this misperception. This perception error leads her to an overdraft.

3.4.8.3 Effect of Inattention--Error Constraints Spending

Figure 22 Inattention Leads to Underspending



Finally, we discover an interesting case where inattention may cause the consumer to spend less than her optimal spending level. This happens because the consumer knows that she is inattentive, i.e., she might overestimate her effective balance to run into overdraft. In order to prevent this, the consumer tends to constrain her spending. As shown in Figure 22, though the optimal spending is 0.8 as in the previous example (section 3.4.8.3), the inattentive consumer chooses to spend 0.5 to prevent overdraft. This example suggests a new revenue source for the bank. If the bank provides automatic alerts to consumers to inform them of their exact balances, the consumers won't have to take precautions to avoid overdrafts. As a consequence, consumers will spend more and the bank can benefit from the increased interchange fees.

3.5 Identification and Estimation

We now discuss the identification of the parameters and the estimation procedure.

3.5.1 Identification

The unknown structural parameters in the model include $\{\theta, \beta, \varsigma, \xi, \rho, \Upsilon\}$ where θ is the logarithm of the mean risk averse coefficient, β is the discount factor, ς is the standard deviation of the risk averse coefficient, ξ is the monitoring cost, ρ is the sensitivity of balance error variance to the lapsed time since last balance checking, and Υ is the dissatisfaction sensitivity. Next we provide an informal rationale for identification of each parameter.

First of all, as we know from Rust (1987), the discount factor β cannot be separately identified from the static utility parameter, which in our case, the risk aversion coefficient . The reason is that lowering θ tends to increase consumption/spending, an effect which can also be achieved by lowering β. As we are more interested in the consumers' time preference rather than risk preference, we fix the risk averse coefficient, which allows me to identify the discount factor[76]. This practice is also used in Gopalakrishnan, Iyengar, Meyer 2014. As to the risk averse coefficient, we choose $\theta = 0.74$, following the latest literature by Andersen et al. (2008) where they jointly elicit risk and time preferences[77]. After fixing θ, β_i can be well identified by the sequences of consumption (spending) within a pay period. A large discount factor (close to 1) implies a stable consumption stream while a small discount factor implies a downward sloping consumption stream. Because a discount factor is constrained above by 1, we do a transformation to set $\beta_i = \frac{1}{1+\exp(\lambda_i)}$ and estimate λ_i instead.

Second, the standard deviation of risk averse coefficient ς_i is identified by the variation of consumptions on the same day of the pay period but across different pay periods.

Moreover, according to the intertemporal tradeoff, the longer the consumer goes without checking her balance, the more likely she will be to overdraw due to the balance error. The observed data pattern of more overdraft fees paid longer after a balance checking inquiry can help pin down the structural parameters ρ_i .

Intuitively, the monitoring cost ξ_i is identified by the expected overdraft payment amount. Recall that the tradeoff regarding balance checking is that a consumer only checks balance when ξ_i is smaller than the expected overdraft payment amount. In the data we observe the balance checking frequency. Combining this with the calculated ρ_i we can compute the expected overdraft probability and further the expected overdraft payment amount, which is the identified ξ_i . Given ρ_i , a consumer with few balance checking inquiries must have a higher balance checking cost ξ_i .

Lastly, the dissatisfaction sensitivity parameter Υ_i can be identified by the data pattern that consumers' account closure probability varies with the ratio of overdraft fee over the overdraft transaction amount, as shown in section 3.3.1.

[76] We also tried to fix the discount factor (at 0.9997) and estimate the risk averse coefficients. The posterior mean of the estimated relative risk averse coefficient is 0.72. Other structural parameter estimates are not significantly unaffected under this specification. Our results confirm that the risk averse coefficient and the discount factor are mathematically substitutes (Andersen et al. 2008). Estimation results with fixed discount factor are available upon requests.

[7777] We also tried other values for the relative risk averse coefficient θ , the estimated discount factor β values change with different θ 's, but other structural parameter values remain the same. The policy simulation results are also robust with different values of θ 's.

Note that aside from these structural parameters, there is another set of parameters that govern the transition process. These parameters can be identified prior to structural estimation from the observed state variables in our data. The set includes $\{ \phi, G, X, p\}$.

In sum, the structural parameters to be estimated include $\{ \lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i\}$.

3.5.2 Likelihood

The full likelihood function is

$$Likelihood = L\left(\left\{\left\{C_{it}, Q_{it}, W_{it}; \widehat{S_{it}}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{f\left\{\widehat{S_{it}}|\widehat{S_{it-1}}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{\widehat{S_{i0}}\right\}_{i=1}^{I}\right)$$

where $\widehat{S_{it}} = \{B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_t\}$. As the likelihood for the optimal choices and that for the state transition process are additively separable when we apply log to the likelihood function, we can first estimate the state transition process from the data, then maximize the likelihood for the optimal choices. The likelihood function for the optimal choice is

$$L\left(\left\{\left\{C_{it}, Q_{it}, W_{it}; \widehat{S_{it}}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) = \prod_{i=1}^{I}\prod_{t=1}^{T} L\left\{C_{it}, Q_{it}, W_{it}; \widehat{S_{it}}\right\}$$

$$= \prod_{i=1}^{I}\prod_{t=1}^{T} L\left\{C_{it}; \widehat{S_{it}}\right\} L\left\{Q_{it}; \widehat{S_{it}}\right\} L\left\{W_{it}; \widehat{S_{it}}\right\}$$

$$= \prod_{i=1}^{I}\prod_{t=1}^{T} f\{\varepsilon_{it}|C_{it}\} Pr\{\chi_{it}|Q_{it}, C_{it}\} L\{\varpi_{it}|W_{it}, Q_{it}, C_{it}\}$$

where $f(\varepsilon_{it}|C_{it})$ is estimated from the normal kernel density estimator to be explained in section 3.5.3.1, $Pr(\chi_{it}|C_{it}, Q_{it})$ and $Pr(\varpi_{it}|C_{it}, Q_{it}, W_{it})$ follow the standard logit model given the choice specific value function in equation [eq:ChoiceSpeValFun]. In specific,

$$\Pr\left(Q_{it} = 1; \widehat{S_{it}}\right) = \int_{\varpi_{it}}\int_{\varepsilon_{it}}\int_{\eta_{it}} \frac{exp\left\{v\left(C_{it}, Q_{it} = 1, W_{it}; \widehat{S_{it}}\right)\right\}}{\sum_{Q_{it}} v\left(C_{it}, Q_{it}, W_{it}; \widehat{S_{it}}\right)} d\eta_{it} d\,\varepsilon_{it} d\varpi_{it}$$

$$\Pr\left(W_{it} = 1; \widehat{S_{it}}\right) = \int_{\chi_{it}}\int_{\varepsilon_{it}}\int_{\eta_{it}} \frac{exp\left\{v\left(C_{it}, Q_{it}, W_{it} = 1; \widehat{S_{it}}\right)\right\}}{\sum_{W_{it}} v\left(C_{it}, Q_{it}, W_{it}; \widehat{S_{it}}\right)} d\eta_{it} d\,\varepsilon_{it} d\chi_{it}$$

3.5.3 Estimation: Imai, Jain and Ching (2009)

3.5.3.1 Modified IJC

We use the Bayesian estimation method developed by Imai, Jain and Ching (2009) to estimate the dynamic choice problem with heterogeneous parameters. As our model involves a continuous choice variable, spending, we adjust the IJC algorithm[78]. to obtain the choice probability through kernel density estimation. We now show the details of the estimation procedure. The whole parameter space is divided into two sets ($\Omega = \{ \Omega_1, \Omega_2 \}$), where the first one contains hyper-parameters in the distribution of the heterogeneous parameters ($\Omega_1 = \{ \mu_\lambda, \mu_\varsigma, \mu_\xi, \mu_\rho, \mu_\Upsilon, \sigma_\lambda, \sigma_\varsigma, \sigma_\xi, \sigma_\rho, \sigma_\Upsilon \}$), and the second set contains heterogeneous parameters ($\Omega_2 = \{ \lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i \}_{i=1}^{I}$) . We allow all heterogeneous parameters (represented by ϑ_i) to follow a normal distribution with parameters mean μ_ϑ and standard deviation σ_ϑ . Let the observed choices be $O^d = \{ O_i^d \}_{i=1}^{I} = \{ C_i^d, Q_i^d, W_i^d \}$ where $C_i^d \equiv \{ C_{it}^d, \forall\, t \}$, $Q_i^d \equiv \{ Q_{it}^d, \forall\, t \}$ and $W_i^d \equiv \{ W_{it}^d, \forall\, t \}$.

Each MCMC iteration mainly consists of two blocks.

(i) Draw Ω_1^r, that is, draw $\mu_\vartheta^r \sim\ f_{\mu_\vartheta}(\vartheta | \sigma_\vartheta^{r-1}, \Omega_2^{r-1})$ and $\sigma_\vartheta^r \sim\ f_{\sigma_\vartheta}(\sigma_\vartheta | \mu_\vartheta^r, \Omega_2^{r-1})$

($\vartheta \in \{ \lambda, \varsigma, \xi, \rho, \Upsilon \}$, the parameters that capture the distribution of ϑ for the population) where f_{μ_ϑ} and f_{σ_ϑ} are the conditional posterior distributions.

(ii) Draw Ω_2^r , that is, draw individual parameters $\vartheta_i \sim\ f_i(\vartheta_i | O_i^d, \Omega_1^r)$ by the Metropolis-Hastings (M-H) algorithm.

More details of the estimation algorithm is presented in Appendix A3.2.

3.5.3.2 Parallel Computing: Neiswanger, Wang and Xing (2014)

We adopt the parallel computing algorithm by Neiswanger, Wang and Xing (2014) to estimate our model with data from more than 500,000 consumers. The logic behind this algorithm is that the full likelihood function is a multiplicative of the individual likelihood.

$$p(\vartheta | x^N) \propto p(\vartheta) p(x^N | \vartheta) = p(\vartheta) \prod_{i=1}^{N} p(x_i | \vartheta)$$

So we can partition the data onto multiple machines, and then perform MCMC sampling on each using only the subset of data on that machine (in parallel, without any communication). Finally, we can combine the subposterior samples to algorithmically construct samples from the full-data posterior.

[78] The IJC method is designed for dynamic discrete choice problems. Zhou (2012) also applied it to a continuous choice problem

In details, the procedure is:

(1) Partition data x^N into M subsets $\{x^{n_1}, \dots, x^{n_M}\}$.

(2) For $m = 1, \dots, M$ (in parallel):

(a) Sample from the subposterior p_m, where $p_m(\vartheta|x^{n_m}) \propto p(\vartheta)^{\frac{1}{M}} p(x^{n_m}|\vartheta)$

(3) Combine the subposterior samples to produce samples from an estimate of the subposterior density product $p_1 \dots p_M$, which is proportional to the full-data posterior, i.e. $p_1 \dots p_M(\vartheta) \propto p(\vartheta|x^N)$.

Given T samples $\{\vartheta_t\}_{t=1}^{T}$ from a subposterior p_m, we can write the kernel density estimator as $\widehat{p_m(\vartheta)}$,

$$\widehat{p_m}(\vartheta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K(\frac{||\vartheta - \vartheta_t||}{h})$$

$$= \frac{1}{T}\sum_{t=1}^{T}(2\pi\, h^2)^{-\frac{d}{2}}|I_d|^{-\frac{1}{2}} exp\left\{-\frac{1}{2h^2}(\vartheta - \vartheta_t)' I_d^{-1}(\vartheta - \vartheta_t)\right\}$$

$$= \frac{1}{T}\sum_{t=1}^{T} N(\vartheta|\vartheta_t, h^2 I_d)$$

where we have used a Gaussian kernel with bandwidth parameter h. After we have obtained the kernel density estimator $\widehat{p_m}(\vartheta)$ for M subposteriors, we define our nonparametric density product estimator for the full posterior as

$$\widehat{p_1 \cdots p_m}(\vartheta)$$

$$= \widehat{p_1} \cdots \widehat{p_m}(\vartheta)$$

$$= \frac{1}{T^M}\sum_{t_1=1}^{T} \cdots \sum_{t_M=1}^{T}\prod_{m=1}^{M} N\left(\vartheta|\vartheta_{t_m}^m, h^2 I_d\right)$$

$$\propto \sum_{t_1=1}^{T} \cdots \sum_{t_M=1}^{T} N\left(\vartheta|\overline{\vartheta_{t\cdot}}, \frac{h^2}{M} I_d\right)\prod_{m=1}^{M} N\left(\vartheta_{t_m}^m|\overline{\vartheta_{t\cdot}}, h^2 I_d\right)$$

$$= \propto \sum_{t_1=1}^{T} \cdots \sum_{t_M=1}^{T} w_t \cdot N\left(\vartheta|\overline{\vartheta_{t\cdot}}, \frac{h^2}{M} I_d\right)$$

This estimate is the probability density function (pdf) of a mixture of TM Gaussians with unnormalized mixture weights $w_t \cdot$ Here, we use $t \cdot= \{t_1, \dots, t_M\}$ to denote the set of indices for the M samples $\{\vartheta_{t_1}^1, \dots, \vartheta_{t_M}^M\}$ (each from one machine) associated with a given mixture component, and let

$$w_t \cdot= \prod_{m=1}^{M} N\left(\vartheta_{t_m}^m | \overline{\vartheta_{t.}}, h^2 I_d\right)$$

$$\overline{\vartheta_{t.}} = \frac{1}{M} \sum_{m=1}^{M} \vartheta_{t_m}^m$$

(4) Given the hierarchical Bayes framework, after obtaining the posterior distribution of the population parameter, use M-H algorithm once more to obtain the individual parameters (details in Appendix A3.2 Step 4)

The sampling algorithm is presented in Appendix A3.3.

3.6 Results

3.6.1 Model Comparison

Table 39 Model Comparison

	A: No Forward Looking	B: No Inattention	C: No Heterogeneity	D: Proposed
Log-Marginal Density	-2943.28	-3636.59	-2764.56	-1758.33
Hit Rate: Overdraft	0.499	0.351	0.504	0.870
Hit Rate: Check Balance	0.405	0.226	0.632	0.841
Hit Rate: Close Account	0.660	0.727	0.696	0.758

We compare our model against the other four benchmark models in order to investigate the contribution of each element of the structural model. Models A to C are our proposed model without forward-looking, inattention and unobserved heterogeneity respectively and model D is our proposed model. Table 3 shows the log-marginal density (Kass and Raftery 1995) and the hit rate for overdraft, check balance and close account incidences (We only consider when these events happen because no event takes place the majority of the time. The hit rates for non-incidences are shown in the appendix A3.4.). All four measures show that our proposed model significantly outperforms the benchmark models. Notably inattention contributes the most to model fit which is consistent with our conjecture in section 3.3.2.

3.6.2 Value of Parallel IJC

Table 40 Estimation Time Comparison

Size\Method (seconds)	Parallel IJC	IJC	CCP	FIML
1,000	518	1579	526	5,010
10,000	3,199	12,560	4,679	54,280
100,000	4,059	140,813	55,226	640,360
>500,000	5,308	788,294	399,337	3,372,660
	(1.5 hr)	(9 days)	(5 days)	(39 days)

Table 41 Monte Carlo Results when N=100,000

Var	True Value		Parallel IJC	IJC	CCP	FIML
μ_β	0.9	Mean	0.878	0.883	0.851	0.892
		Std	0.041	0.039	0.036	0.025
μ_ς	1.5	Mean	1.505	1.502	1.508	1.501
		Std	0.131	0.124	0.199	0.103
μ_ξ	0.5	Mean	0.482	0.507	0.515	0.502
		Std	0.056	0.039	0.071	0.044
μ_ρ	1	Mean	1.006	1.003	1.015	1.002
		Std	0.027	0.022	0.026	0.019
μ_Υ	5	Mean	5.032	5.011	4.943	4.987
		Std	0.023	0.01	0.124	0.008
σ_β	0.1	Mean	0.113	0.095	0.084	0.104
		Std	0.016	0.014	0.015	0.01
σ_ς	0.3	Mean	0.332	0.318	0.277	0.309
		Std	0.024	0.015	0.029	0.021
σ_ξ	0.1	Mean	0.112	0.091	0.08	0.09
		Std	0.055	0.029	0.025	0.025
σ_ρ	0.1	Mean	0.107	0.107	0.085	0.105
		Std	0.008	0.006	0.01	0.006
μ_Υ	0.1	Mean	0.092	0.109	0.111	0.1
		Std	0.014	0.013	0.021	0.009

We report the computational performance of different estimation methods in Table . All the experiments are done on a server with an Intel Xeon CPU, 144 cores and 64 GB RAM. The first column is the performance of our proposed method, IJC with parallel computing. We compare it with the original IJC method, the Conditional Choice Probability (CCP) method by Arcidiacono and Miller (2011)[79] and the Full Information Maximum Likelihood (FIML)

[79] We use the finite mixture model to capture unobserved heterogeneity and apply the EM algorithm to solve for the unobserved heterogeneity. More details of the estimation results can be obtained upon requests.

method by Rust (1987) (or Nested Fixed Point Algorithm)[80]. As the sample size increases, the comparative advantage of our proposed method is more notable. To run the model on the full dataset with more than 500,000 accounts takes roughly 1.5 hours compared to 9 days with the original IJC method[81]. The reason for the decrease in computing time is that our method takes advantage of multiple machines that run in parallel. We further run a simulation study to see if the various methods are able to accurately estimate all parameters. Table shows that different methods produce quite similar estimates and all mean parameter estimates are within two standard errors of the true values. The Parallel IJC method is slightly less accurate than the original IJC method.

The parallel IJC is almost 600 times faster than FIML. This happens because the full solution method solves the dynamic programming problem at each candidate value for the parameter estimates, whereas this IJC estimator only evaluates the value function once in each iteration.

3.6.3 Parameter Estimates

Table 42 Structural Model Estimation Results

Var	Interpretation	Mean (μ_ϑ)	Standard deviation (σ_ϑ)
β_i	Discount factor	0.9997	0.362
		(0.00005)	(0.058)
ς_i	Standard deviation of relative risk aversion	0.257	0.028
		(0.014)	(0.003)
ξ_i	Monitoring cost	0.708	0.255
		(0.084)	(0.041)
ρ_i	Inattention Dynamics–lapsed time	7.865	0.648
		(0.334)	(0.097)
Υ_i	Dissatisfaction Sensitivity	5.479	1.276
		(1.329)	(0.109)

Table presents the results of the structural model. We find that the daily discount factor is around 0.9997. This is equivalent to a yearly discount factor of 0.89, largely consistent with the literature (Fang and Wang 2014, Hartmann and Nair 2010). The standard deviation of the discount factor is 0.362. This suggests that some consumers have quite low discount factors--consistent with our heavy discounting hypothesis.

[80] We use the random coefficient model to capture unobserved heterogeneity. More details of the estimation results can be obtained upon requests.

[81] We keep 2000 total number of MCMC iterations, using the first 500 as burn-in. Convergence was assessed visually using plots of the parameters. We chose a store N=100 past pseudo-value functions. The bandwidth parameter is set to be $h = 0.01$.

The monitoring cost is estimated to be 0.708. Using the risk averse coefficient, we can evaluate the monitoring cost in monetary terms. It turns out to be $2.03. We also obtained the cost measure for each individual consumer.

The variance of the balance perception error increases with the lapsed time since the last time to check balance and with the mean balance level. Notably the variance of the balance perception error is quite large. If we take the average number of days to check the balance from the data, which is 9, then the standard deviation is 7.865*9=70.79 . This suggests a very widely spread distribution of the balance perception error.

The estimated dissatisfaction sensitivity parameter confirms our hypothesis that consumers can be strongly affected by the bank fee and close the account as a consequence of dissatisfaction. If we consider an average overdraft transaction amount at $33, then the relative magnitude of the effect of dissatisfaction is comparable to $171. This suggests that unless the bank would like to offer a $171 compensation to the consumer, the dissatisfied consumer will close the current account and switch. Moreover, consistent with the evidence in Figure 18, the dissatisfaction sensitivity is stronger for light overdrafters (whose average is 5.911) than for heavy overdrafters (whose average is 3.387). And keeping the average overdraft transaction amount as fixed, a 1% increase in the overdraft fee can increase the closing probability by 0.12%.

3.7 Counterfactuals

3.7.1 Pricing

The structurally estimated model allows us to examine the effect of changing the pricing structure on consumers' spending pattern and more importantly, their overdraft behavior. We test three alternative pricing schemes: a reduced per-item flat fee, a percentage fee, and a quantity premium.

Table 43 Overdraft Fee under Alternative Pricing

Pricing	Current	Reduced Flat	Percentage	Quantity Premium
	$31	$29.27	15.80%	8.5% *I (OD ≤ 10) + $31 *I (OD>10)
Overdraft Revenue	$18,654,510	$19,262,647	$19,982,711	$20,297,972
Overdraft Freq	544,997	590,093	610,288	631,325
% Δ Revenue	–	3.26%	7.12%	8.81%
% Δ Freq	–	2.77%	11.98%	15.84%
% Δ Check Balance	–	-3.58%	2.83%	3.31%
% Δ Close Account	–	-1.01%	-1.35%	-1.94%

Notice here that the underlying assumption for all these simulations is fungibility, i.e., consumers' reaction only depends on the fee amount rather than the fee structure. If two different fee structures result in the same fee amount, then the consumer should respond in the same fashion.

In the first scenario, we keep the per-item flat fee scheme but reduce it to $29.27 per item. Because of law of demand, there's a negative relationship between the per-item overdraft fee and overdraft frequency. So we further pursue an optimization task where we try to solve the optimal per-item fee. As we aggregate data to the daily level, we calculate the average transaction amount for each item, which is $44, and use it to derive the total overdraft fee. For example, if a consumer overspent $170, then the consumer had to pay four overdraft item fees. The optimization is a nested algorithm where in the outer loop we search for the per-item overdraft fee, and in the inner loop we solve the consumer's best response, including optimal spending, balance checking and account closing given the fee size. We found that the optimal per-item overdraft fee is $29.27 under which the bank's revenue will increase by 3.26%. This suggests that the current overdraft fee is too high because the bank fails to take into account consumer's negative reaction to the overdraft fee, which results in huge loss in the consumers' lifetime value (I calculate the lifetime value of a consumer in a conservative way by multiplying the accounts spendings by the interchange rate).

In the second scenario, the per-item flat fee is changed to a percentage fee of 15.8% (optimized in a similar way as described in the first scenario). This is lower than the 17% calculated from the ratio of the total fee paid over the total transaction amount that caused the fees in the data. Again this suggests that the bank might be charging a too high fee currently. Intuitively, the percentage structure should encourage consumers to overdraw on transactions of a small amount but deter them from overdrawing on transactions of a large amount. As there are more transactions of a small amount than transactions of a large amount, the total fees generated soars by 7.12%. Therefore, the percentage overdraft fee invites more consumers to use the overdraft service. It is this market expansion effect that increases the bank's overdraft revenue.

In the last scenario, a quantity premium structure is employed, where when a consumer overdraws no more than 10 times, she pays a 8.5% percentage fee and if she overdraws more than 10 times, she pays a flat fee at $31. This quantity premium can increase the bank’s revenue by 8.81%, because the quantity premium uses the second degree price discrimination to segment two types of overdrafters. The bank will earn more overdraft fee from the heavy overdrafters who are willing to pay for the flat fee while retaining the lifetime value for the light overdrafters who prefer the percentage fee (due to the high dissatisfaction sensitivity).

3.7.2 Alerts Benefit Consumers And the Bank

Although the changed pricing strategies can help the bank improve revenue, the bank is still exploiting consumer inattention and may exacerbate consumer attrition. In this counterfactual, we propose a new product design strategy (specific design to be introduced in section 3.7.3) to help consumers prevent overdrafts: sending automatic alerts to inform consumers about their balances. As alerts eliminate consumers' balance perception error, the total amount of overdraft fee paid by consumers decreases by 49.53% (Table . This is in comparison to the overdraft revenue under the optimal Quantity Premium pricing strategy in Table).

Table 44 Effect of Alerts on Bank's Revenue

	Amount	Percentage Change
Overdraft revenue	$10,243,529	-49.53%
Interchange revenue from increased spendings	$1,997,488	9.84%
Lifetime value from retained consumers	$8,430,424	41.53%
Total	$20,671,441	1.84%

Although alerts benefit consumers by helping them avoid the high overdraft fees, the bank might not have incentives to send out alerts as its objective is to earn more revenue. However, we find that alerts can benefit the bank too for two reasons. First of all, as shown in section 3.4.8.3, due to inattention consumers are constraining spendings to prevent overdrafts. With alerts, consumers' precautionary motive is relieved so that they will increase spendings. As a result, the bank can gain more interchange fees. We calculate this gain of more interchange fee from the increased amount of spending by multiplying the increased spending with an average interchange fee rate of 0.8%[82]. We find that sending alerts to consumers can offset 9.84% of the loss in overdraft fees because of the gain in the interchange fees. Moreover, without being dissatisfied by the overdraft fee, consumers are less likely to close their accounts. We find that alerts reduce the number of closed accounts from 16.37% to 8.25% which increases the bank's revenue by getting the lifetime value from these retained consumers. As shown in Table , the increased lifetime value from retained consumers and the increase in interchange fee from increased spendings not only offset the loss in overdraft revenue but increase it by 1.84%.

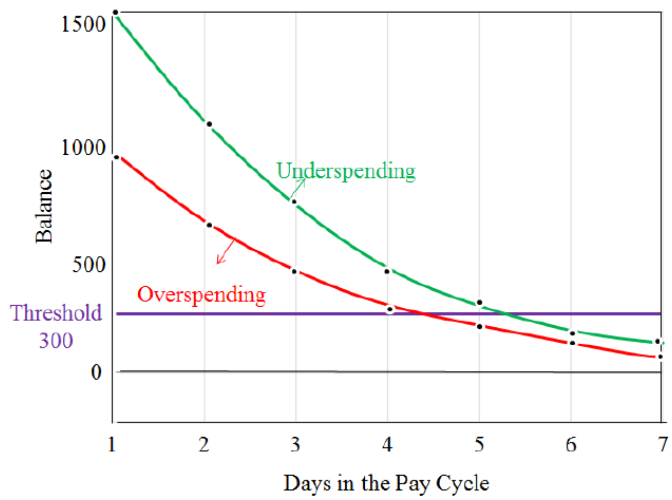
3.7.3 Optimal Alert Strategy

Finally, we explain how we design the optimal alert that can help the bank increase its revenue in section 3.7.2. We show the effect of the proposed alert with an example in Figure 23. Consider a consumer who receives a weekly wage of $2000. This consumer's discount factor is 0.8[83]. She sets a threshold alert at $300 originally thus will only receive the alert when the account balance is below $300. But our proposed alert will be triggered both when the consumer is overspending and underspending. As shown in the figure 23, as long as the consumer's spending falls out of the range between the overspending and underspending lines, an alert will be received. So when the consumer's balance is below $700 on day 2, she

[82] http://www.federalreserve.gov/paymentsystems/regii-average-interchange-fee.htm
[83] For the ease of exposition, we choose a relatively small discount factor.

will receive an alert although the threshold is not reached yet. The optimal alert is earlier than the threshold alert to give the consumer more time to adjust her spending rather than to wait until the last moment when she can hardly make any immediate change. On the other hand, if the consumer's balance is below $300 on day 5, the threshold alert will be triggered while the consumer is still in a safe zone. Receiving the threshold alert doesn't help consumers because her perception error accumulates too fast to make day 6 and 7 danger days prone to overdrafts again. Therefore, the dynamic alert can correct the defects of the threshold alerts of being either too late or too early.

Figure 23 Dynamic Optimal Alert Notifies Overspending and Underspending



Another imbedded feature of the dynamic alert is that it accounts for consumers' disutility to receive too many alerts. In reality, consumers dislike frequent alerts that spam their mailboxes. We incorporate this alert-averse effect into an optimization task where we choose the optimal timing to send the alerts given the estimated structural parameters. The objective function is as follows

$$\max_{\{A_{it}\}} \sum_{i=1}^{N} \sum_{t=1}^{\infty} \beta^{t-1} \left[U_{it}\left(C_{it}^{*}, W_{it}^{*}; \widehat{S_{it}} \right) - \kappa_i \right]$$

$$\widehat{S_{it}} = A_{it} S_{it} + (1 - A_{it}) \widetilde{S_{it}}$$

is a binary choice of whether to send an alert to the consumer i at time t. The second equation means that if the alert is sent, the consumer knows the exact balance and coming bills, denoted as the true state variable S_{it} ; if not, the consumer only knows the distribution of the perceived balance and coming bills, denoted as $\widetilde{S_{it}}$. The consumers' disutility of receiving the alert is summarized by a time invariate the parameter κ_i . We solve the optimization problem in a nested algorithm where in the outer loop we test for all combinations of alert opportunities, and in the inner loop we solve the consumer's best response, including optimal spending and account closing given the alert profile. (We assume that consumers don't have to make the balance checking decision because of the automatic alerts.)

We first test the optimal alert strategy assuming that all consumers have the same structural parameters (we use the posterior mean of the hyper-distribution parameters). We set this disutility as the inverse of the estimated monitoring cost (μ_ξ) because the consumer who incurs a high monitoring cost might not know how to use online banking or call centers so automatic message alerts are favored. As Table reports, this alert service increases total consumer utility by 1.11% when the threshold rule of $300 is applied and 2.85 when the dynamic rule is applied.

We further allow all structural parameters to be heterogeneous across consumers and solve the optimal alert timing specific to each individual. We find that targeted alerts can increase consumer utilities six times more than the uniform threshold alert (6.65%).

Table 45 Utility Impact of Different Types of Alerts

Alert Type	Alert Timing	Utility Gain
Uniform	Threshold	1.11%
	Dynamic	2.85%
Targeted	Threshold	4.39%
	Dynamic	6.65%

3.8 Contributions and Limitations

The $32 billion dollar annual overdraft fee has caused consumer attrition and may induce potentially tighter regulation. However there is little quantitative research on consumers' financial decision making processes that explains their overdraft behaviors. The lack of well-calibrated models prevent financial institutions from designing pricing strategies and improving financial products. With the aid of Big Data associated with consumers' spending patterns and financial management activities, banks can use adverse targeting (Kamenica, Mullainathan, and Thaler 2011) to help consumers know themselves better and make better financial decisions.

In this paper we build a dynamic structural model of consumer daily spending that incorporates inattention to rationalize consumers' overdraft behavior. We quantify the discount factor, monitoring cost and dissatisfaction sensitivity for each consumer and use these to design new strategies. First we compare the current pricing scheme with several alternative pricing strategies. We find that a percentage fee structure can increase the bank's revenue through market expansion and the quantity premium structure can increase the bank's revenue because of second degree price discrimination. More importantly, we propose an alert strategy to make the incentive of the bank and the incentive of the consumers better aligned. The optimal alert can be sent to the right consumer at the right time to prevent overdrafts. This customized dynamic alert product can be six times more effective than a uniform threshold alert. Not only does this alert benefits consumers, it can also benefit the bank through increased interchange fees and lower consumer attrition.

We calibrated our model at an individual level on a sample of more than 500,000 accounts. This Big Data provide great value for our analysis. First of all, an overdraft is still a relatively rare event compared to numerous other transactions. Without a large amount of data, we cannot detect these rare but detrimental events, let alone their diverse causes. Second, as summarized by Einav and Levin (2014), Big Data contain rich micro-level variation that can used to identify novel behavior and develop predictive models that are harder with smaller samples, fewer variables, and more aggregation. We leverage the variation in consumer daily spending and balance checking behaviors to evaluate the effect of heterogeneous policy instruments. These evaluations can be useful for bank managers to design new products and policy makers to create new regulation rules at a much more refined fashion than before.

In order to estimate a complicated structural model with Big Data, we adopt parallel computing techniques in combination with the Bayesian estimation algorithm developed by Imai, Jain and Ching (2009). This new method significantly reduces the computation burden and could be used for other researchers and marketers who would like to use structural models to solve real-world large-scale problems.

There are several limitations of the current study that call for future work. First, we don't observe consumers' existing alert settings. Some consumers may have already received alerts to help them make financial decisions. In our policy simulations, we made bold assumptions about consumers' disutility for reading alerts. These assumptions could be tested if we had the alerts data. The current alerts are set by consumers who might fail to consider their spending dynamics. Future field experiments are needed to test the effect of our proposed alert strategy. Second, we don't have the data about consumers' decision on whether to opt-in for overdraft protection by ATM/POS transactions. We only know that if ATM/POS transactions caused an overdraft, then the consumer must have opted-in. If no such transactions happened, we do not know the consumer's opt-in status. Had we known this information, we could have provided an informative prior in our the Bayesian model. The logic is that a consumer who has opted in probably has stronger needs for short term liquidity due to fluctuations in the size and arrival time of income and expenditures. Finally, we only model consumers' non-preauthorized spending in the checking account. In reality, consumers usually have multiple accounts, like savings, credit cards and loans, with multiple financial institutions. A model to capture consumers' decisions across all accounts for both short-term and long-term finances will provide a more complete picture of consumers' financial management capabilities and resources so that the bank can design more customized products.

Bibliography

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3), 583-618.

Andrews, D. W., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics*, *101*(1), 123-164.

Arcidiacono, P., & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. Econometrica, 79(6), 1823-1867.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485-1509.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297.

Arrow, K. J. 1963. Liquidity preference. Lecture VI in Lecture Notes for Economics 285, *The Economics of Uncertainty*, pp. 33-53, undated, Stanford University.

Bass, F. M. 1969. A new product growth model for consumer durables. *Management Sci.* 15(5):215-227.

Bellman, Richard 1957. *Dynamic Programming*, Princeton, NJ: Princeton University Press.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. *Marketing science*, 17(1), 45-65.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, *1*(2), 141-162.

Brumberg, R. & Modigliani, F. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. In Kenneth K. Kurihara, ed., *Post-Keynesian Economics*, 388-436. New Brunswick, NJ: Rutgers University Press.

Card, D., DellaVigna, S., & Malmendier, U. (2011). The role of theory in field experiments. *The Journal of Economic Perspectives*, 25(3), 39-62.

Chakravarty, A., Liu, Y., & Mazumdar, T. (2010). The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation. *Journal of Interactive Marketing*, *24*(3), 185-197.

Chen, Y., Wang, Q., & Xie, J. (2011). Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning. *Journal of Marketing Research*, *48*(2), 238-254.

Cheng, Z., Caverlee, J., & Lee, K. (2010, October). You are where you Tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 759-768). ACM.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4), 1145-1177.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 345-354.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, *29*(5), 944-957.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375-1388.

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*,*27*(4), 293-307.

Derdenger, D. and V. Kumar 2013. The dynamic effect of bundling as a product strategy. *Marketing Sci.* forthcoming.

Dewan, S., & Ramaprasad, J. (2012). Research Note—Music Blogging, Online Sampling, and the Long Tail. *Information Systems Research*, *23*(3-Part-2), 1056-1067.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), 300-307.

Dutt, P., & Padmanabhan, V. (2011). Crisis and consumption smoothing. *Marketing Science*, *30*(3), 491-512.

Einav, L., & Levin, J. D. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy*, forthcoming.

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, *53*(6), 881-893.

Erdem, T., M. P. Keane, T. S. and J. Strebel. 2005. Learning about computers: An analysis of information search and technology choice. *Quant. Marketing Econom*. 3(3):207-246.

Evans, D., Litan, R., & Schmalensee, R. (2011). Economic Analysis of the Effects of the Federal Reserve Board's Proposed Debit Card Interchange Fee Regulations on Consumers and Small Businesses. Available at SSRN 1769887.

Fang, H., & Silverman, D. (2009). Time-inconsistency and welfare program participation: Evidence from the nlsy. *International Economic Review*, 50(4), 1043-1077.

Fang, H., & Yang W. (2014). Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions. *International Economic Review*, forthcoming.

Farrell, Joseph, and Paul Klemperer, "Coordination and lock-In: Competition with switching costs and network effects," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, Volume III (Amsterdam: North-Holland, 2005).

Farrell, J. and Simcoe, T. 2012. Four Paths to Compatibility. *Oxford Handbook of the Digital Economy*. Oxford University Press.

Fishman, M. J., & Hagerty, K. M. (2003). Mandatory versus voluntary disclosure in markets with informed and uninformed customers. *Journal of Law, Economics, and Organization*, 19(1), 45-63.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. The Quarterly Journal of Economics, 121(2), 505-540.

Gentzkow, M. 2007. Valuing new goods in a model with complementarity: Online newspapers. *Amer. Econom. Rev.* 97(3):713-744.

Ghose, A., Ipeirotis, P. G., & Sundararajan, A. (2007, June). Opinion mining using econometrics: A case study on reputation systems. In *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS* (Vol. 45, No. 1, p. 416).

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on*, *23*(10), 1498-1512.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), 493-520.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, *23*(4), 545-560.

Gopalakrishnan, A., Iyengar, R., & Meyer, R. J. (2014). Consumer dynamic usage allocation and learning under multi-part tariffs: theory and empirical evidence. Working paper, Wharton School, University of Pennsylvania.

Gopinath, S., Chintagunta, P. K., & Venkataraman, S. (2013). Blogs, advertising, and local-market movie box office performance. *Management Science*, *59*(12), 2635-2654.

Gordon, B. R. 2009. A dynamic model of consumer replacement cycles in the PC processor industry. *Marketing Sci.* 28(5):846-867.

Gowrisankaran, G., M. Rysman. 2012. Dynamics of consumer demand for new durable consumer goods. *J. of Polit. Econ*, 120:1173-1219.

Grubb, M. D. (2014). Consumer inattention and bill-shock regulation. *The Review of Economic Studies*, forthcoming.

Grubb, M. D., & Osborne, M. (2014) Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock. *American Economic Review*, forthcoming.

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005, August). The predictive power of online chatter. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 78-87). ACM.

Guadagni, P. M., J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203-238.

Halko, N. P. (2012). *Randomized methods for computing low-rank approximations of matrices* (Doctoral dissertation, University of Colorado).

Hall, R. (1978). Stochastic implications of the life-cycle permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86, 971-987.

Hartmann, W. R., Harikesh S. N. 2010. Retail competition and the dynamics of demand for tied goods. *Marketing Science*, 29(2), 366–386.

Haselhuhn, M. P., Pope, D. G., Schweitzer, M. E., & Fishman, P. (2012). The impact of personal experience on behavior: Evidence from video-rental fines. *Management Science*, 58(1), 52-61.

Hirshleifer, D., & Hong Teoh, S. (2003). Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management*, 9(1), 25-66.

Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, *77*(6), 37-53.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In*Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Imai, S., Jain, N., & Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, 77(6), 1865-1899.

Jiang, L. (2012). The welfare effects of "bill shock" regulation in mobile telecommunication markets. Working Paper, University of British Columbia.

Kahneman, D., J. Knetsch, and R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98(6), 1325–1348.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: *Journal of the Econometric Society*, 263-291.

Kamakura, Wagner and Gary Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *J. of Marketing Res.* 26 (November):379-90.

Kamenica, E., Mullainathan, S., & Thaler, R. (2011). Helping consumers know themselves. *The American Economic Review*, 101(3), 417-422.

Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution

and box office revenue. *International Journal of Research in Marketing*, *28*(1), 62-74.
Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). Getting to the top of mind: How reminders increase saving (No. w16205). National Bureau of Economic Research.
Kass, R., and A. Raftery. (1995). Bayes Factors. *Journal of the American Statistical Association* 90 (430): 773-95.
Katz, M. L., C. Shapiro. 1985. Network externalities, competition, and compatibility. *Amer. Econom. Rev*. 75(3):424-440.
Keaveney, S. M. (1995). Customer switching behavior in service industries: an exploratory study. *The Journal of Marketing*, 71-82.
Lee, T. Y., & BradLow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881-894.
Lee, D., Hosanagar, K., Nair, H. S., & Stanford, G. S. B. (2013). The Effect of Advertising Content on Consumer Engagement: Evidence from Facebook. *Working Paper*
Liu, H., P. Chintagunta and T. Zhu 2010. Complementarities and the demand for home broadband internet services. *Marketing Sci*. 29(4):701-720.
Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 74-89.
Liu, Y., Huang, X., An, A., & Yu, X. (2007, July). ARSA: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 607-614). ACM.
Magnac, T., & Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2), 801-816.
Matutes, Carmen, and Pierre Regibeau, (1988), "Mix and Match: Product Compatibility Without Network Externalities," *Rand J. Econom*, vol. 19 (2):219-234.
Matzler, K., Würtele, A., & Renzl, B. (2006). Dimensions of price satisfaction: a study in the retail banking industry. *International Journal of Bank Marketing*, 24(4), 216-231.
Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science*, 22(1), 58-84.
Melnikov, O. 2013. Demand for differentiated durable products: The case of the U.S. computer printer industry. *Econom. Inquiry* 51(2):1277–1298.
Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PloS one*, *8*(8), e71226.
Mishne, G., & Glance, N. S. (2006, March). Predicting Movie Sales from Blogger Sentiment. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs* (pp. 155-158).
Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, *48*(3), 444-456.
Nair, H. 2007. Intertemporal price discrimination with forward looking consumers: Application to the U.S. market for console video games. *Quant. Marketing Econom*. 5(3):239-292.
Neiswanger, W., Wang, C., & Xing, E. (2014). Asymptotically exact, embarrassingly parallel MCMC. Proceedings of the 30th International Conference on Conference on Uncertainty in Artificial Intelligence.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*,*31*(3), 521-543.
NewYork Times (2015), "Study of TV Viewers Backs Twitter's Claims to be Barometer of Public Mood," March 8.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the*

Econometric Society, 1417-1426.
Norton, J. A., F. M. Bass. 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Sci*. 33(9):1069-1086.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.
O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *ICWSM*, *11*, 122-129.
Onishi, H., & Manchanda, P. (2012). Marketing activity, blogging and sales. *International Journal of Research in Marketing*, *29*(3), 221-234.
Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563-602.
Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals". *Journal of Health Economics*, 28(6), 1154-1165.
Reis, R. (2006). Inattentive consumers. *Journal of monetary Economics*, 53(8), 1761-1800.
Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, *9*(1), 86-136.
Rossi, Peter E., Allenby, G. M., & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. J. Wiley & Sons.
Rust, J. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*. 55(5):999-033.
S Sadikov, E., Parameswaran, A. G., & Venetis, P. (2009, March). Blogs as Predictors of Movie Success. In *ICWSM*.
Seetharaman, P. B., S. Chib, A. Ainslie, P. Boatwright, T. Chan, S. Gupta, N. Mehta, V. Rao, A. Strijnev. 2005. Models of multicategory choice behavior. *Marketing Lett*. 16(3--4):239-254.
Sims, C. A. (1998, December). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 49, pp. 317-356). North-Holland.
Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3), 665-690.
Sethuraman, R., Srinivasan, V., Kim, D. 1999. Asymmetric and neighborhood cross-price effects: Some empirical generalizations. *Marketing Sci*. 18(1):23-41.
Sinha, S., Dyer, C., Gimpel, K., & Smith, N. A. (2013). Predicting the NFL using Twitter. *arXiv preprint arXiv:1310.6998*.
Song, I., P. K. Chintagunta. 2003. Measuring cross-category price effects with aggregate store data. *Management Sci*. 52(10):1594-1609.
Stango, V., & Zinman, J. (2014). Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *Review of Financial Studies*, forthcoming.
Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research*, *49*(5), 624-639.
Sriram, S., P. K. Chintagunta, M. K. Agarwal. 2009. Investigating consumer purchase behavior in related technology product categories. *Marketing Sci*. 29(2):291-314.
Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198-215.
Train, Kenneth. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press
Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). 'Estimating the Dynamic Effects of Online Word-of-Mouth on Member Growth of a Social Network Site. *Journal of Marketing*, *73*(5), 90-102.

Tuli, K. R., & Bharadwaj, S. G. (2009). Customer Satisfaction and Stock Returns Risk. *Journal of Marketing*, *73*(6), 184-197.
Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, *10*, 178-185.
Yao, S., Mela, C. F., Chiang, J., & Chen, Y. (2012). Determining consumers' discount rates with field studies. *Journal of Marketing Research*, 49(6), 822-841.
Zhou, Y. (2012). Failure to Launch in Two-Sided Markets: A Study of the US Video Game Market. Working paper, SUNY at Stony Brook.

Appendix for Chapter 1

A1.1. Model Details

Table A1.1 Summary of Mean Utility Functions

1.only c1: $u_{it}^{1,0} = \alpha_i^1 + \phi_i QC_t^1 + \lambda_i PC_t^1 + \kappa_i * I(1 = \bar{c}_{it}) + \gamma_i^1 \sum_{k=1}^{Invm_{it}^1} (\theta_i^1 + \psi_i * QM_{PT_{1k}}^1)$
2.only c2: $u_{it}^{2,0} = \alpha_i^2 + \phi_i QC_t^2 + \lambda_i PC_t^2 + \kappa_i * I(2 = \bar{c}_{it}) + \gamma_i^2 \sum_{k=1}^{Invm_{it}^2} (\theta_i^2 + \psi_i * QM_{PT_{2k}}^2)$
3.only c3: $u_{it}^{3,0} = \alpha_i^3 + \phi_i QC_t^3 + \lambda_i PC_t^3 + \kappa_i * I(3 = \bar{c}_{it}) + \gamma_i^2 \sum_{k=1}^{Invm_{it}^2} (\theta_i^2 + \psi_i * QM_{PT_{2k}}^2)$
4.only c4: $u_{it}^{4,0} = \alpha_i^4 + \phi_i QC_t^4 + \lambda_i PC_t^4 + \kappa_i * I(4 = \bar{c}_{it}) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
5.only c5: $u_{it}^{5,0} = \alpha_i^5 + \phi_i QC_t^5 + \lambda_i PC_t^5 + \kappa_i * I(5 = \bar{c}_{it}) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
6.only c6: $u_{it}^{6,0} = \alpha_i^6 + \phi_i QC_t^6 + \lambda_i PC_t^6 + \kappa_i * I(6 = \bar{c}_{it}) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
7.only c7: $u_{it}^{7,0} = \alpha_i^7 + \phi_i QC_t^7 + \lambda_i PC_t^7 + \kappa_i * I(7 = \bar{c}_{it}) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
8.only m1: $u_{it}^{0,1} = \theta_i^1 + \psi_i * QM_t^1 + \lambda_i PM_t^1 + I(\bar{c}_{it} = 1)(\alpha_i^{\bar{c}_{it}} + \phi_i * QC_{PT_{\bar{c}_{it}}}^{\bar{c}_{it}}) + \gamma_i^1 I(\bar{c}_{it} = 1) * \sum_{k=1}^{Invm_{it}^1} (\theta_i^1 + \psi_i * QM_{PT_{1k}}^1)$
9.only m2: $u_{it}^{0,2} = \theta_i^2 + \psi_i * QM_t^2 + \lambda_i PM_t^2 + \sum_{j=2}^{3} I(\bar{c}_{it} = j)(\alpha_i^{\bar{c}_{it}} + \phi_i * QC_{PT_{\bar{c}_{it}}}^{\bar{c}_{it}}) + \gamma_i^2 \sum_{j=2}^{3} I(\bar{c}_{it} = j) * \sum_{k=1}^{Invm_{it}^2} (\theta_i^2 + \psi_i * QM_{PT_{2k}}^2)$
10.only m3: $u_{it}^{0,3} = \theta_i^3 + \psi_i * QM_t^3 + \lambda_i PM_{3t} + \sum_{j=4}^{7} I(\bar{c}_{it} = j)\, (\alpha_i^{\bar{c}_{it}} + \phi_i * QC_{PT_{\bar{c}_{it}}}^{\bar{c}_{it}}) + \gamma_i^3 \sum_{j=4}^{7} I(\bar{c}_{it} = j) * \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
11.c1 & m1: $u_{it}^{1,1} = \alpha_i^1 + \phi_i QC_t^1 + \kappa_i * I(1 = \bar{c}_{it}) + \theta_i^1 + \psi_i QM_t^1 + \lambda_i (PC_t^1 + PM_t^1) + \gamma_i^1 \sum_{k=1}^{Invm_{it}^1} (\theta_i^1 + \psi_i * QM_{PT_{1k}}^1)$
12.c2 & m2: $u_{it}^{2,2} = \alpha_i^2 + \phi_i QC_t^2 + \kappa_i * I(2 = \bar{c}_{it}) + \theta_i^2 + \psi_i QM_t^2 + \lambda_i (PC_t^2 + PM_t^2) + \gamma_i^2 \sum_{k=1}^{Invm_{it}^2} (\theta_i^2 + \psi_i * QM_{PT_{2k}}^2)$
13.c3 & m2: $u_{it}^{3,2} = \alpha_i^3 + \phi_i QC_t^3 + \kappa_i * I(3 = \bar{c}_{it}) + \theta_i^2 + \psi_i QM_t^2 + \lambda_i (PC_t^3 + PM_t^2) + \gamma_i^2 \sum_{k=1}^{Invm_{it}^2} (\theta_i^2 + \psi_i * QM_{PT_{2k}}^2)$
14.c4 & m3: $u_{it}^{4,3} = \alpha_i^4 + \phi_i QC_t^4 + \kappa_i * I(4 = \bar{c}_{it}) + \theta_i^3 + \psi_i QM_t^3 + \lambda_i (PC_t^4 + PM_t^3) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
15.c5 & m3: $u_{it}^{5,3} = \alpha_i^5 + \phi_i QC_t^5 + \kappa_i * I(5 = \bar{c}_{it}) + \theta_i^3 + \psi_i QM_t^3 + \lambda_i (PC_t^5 + PM_t^3) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
16.c6 & m3: $u_{it}^{6,3} = \alpha_i^6 + \phi_i QC_t^6 + \kappa_i * I(6 = \bar{c}_{it}) + \theta_i^3 + \psi_i QM_t^3 + \lambda_i (PC_t^6 + PM_t^3) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
17.c7 & m3: $u_{it}^{7,3} = \alpha_i^7 + \phi_i QC_t^7 + \kappa_i * I(7 = \bar{c}_{it}) + \theta_i^3 + \psi_i QM_t^3 + \lambda_i (PC_t^7 + PM_t^3) + \gamma_i^3 \sum_{k=1}^{Invm_{it}^3} (\theta_i^3 + \psi_i * QM_{PT_{3k}}^3)$
18.no purchase: $u_{it}^{0,0} = \sum_{j=1}^{7} I(\bar{c}_{it} = j)\, (\alpha_i^{\bar{c}_{it}} + \phi_i * QC_{PT_{\bar{c}_{it}}}^{\bar{c}_{it}}) + \sum_{m=1}^{M} \gamma_i^m I(\bar{c}_{it} \sim m) * \sum_{k=1}^{Invm_{it}^m} (\theta_i^m + \psi_i * QM_{PT_{mk}}^m)$

Table A1.2 Transition Matrix of Inventory Process

Choice	$\bar{c}_t$	$Invm_{t+1}^1$	$Invm_{t+1}^2$	$Invm_{t+1}^3$
(0,0)	$\bar{c}_t$	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(1,0)	*1*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(2,0)	*2*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(3,0)	*3*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(4,0)	*4*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(5,0)	*5*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(6,0)	*6*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(7,0)	*7*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3$
(0,1)	$\bar{c}_t$	$Invm_t^1 + 1$ *Or* $Invm_{1t}$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$	$Invm_t^2$	$Invm_t^3$
(0,2)	$\bar{c}_t$	$Invm_t^1$	$Invm_t^2 + 1$ *Or* $Invm_t^2$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$	$Invm_t^3$
(0,3)	$\bar{c}_t$	$Invm_t^1$	$Invm_t^2$	$Invm_t^3 + 1$ *Or* $Invm_t^3$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$
(1,1)	*1*	$Invm_t^1 + 1$ *Or* $Invm_{1t}$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$	$Invm_t^2$	$Invm_t^3$
(2,2)	*2*	$Invm_t^1$	$Invm_t^2 + 1$ *Or* $Invm_t^2$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$	$Invm_t^3$
(3,2)	*3*	$Invm_t^1$	$Invm_t^2 + 1$ *Or* $Invm_t^2$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$	$Invm_t^3$
(4,3)	*4*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3 + 1$ *Or* $Invm_t^3$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$
(5,3)	*5*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3 + 1$ *Or* $Invm_t^3$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$
(6,3)	*6*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3 + 1$ *Or* $Invm_t^3$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$
(7,3)	*7*	$Invm_t^1$	$Invm_t^2$	$Invm_t^3 + 1$ *Or* $Invm_t^3$ *when* $\sum_{k=1}^{M} Invm_{It}^k \geq \bar{B}$

A1.2. Identification

When attempting to evaluate the effect of add-ons on the base product, we face the challenge to identify the add-on-to-base effect, especially when we incorporate state dependence and heterogeneous consumer brand preference at the same time. In fact, the data pattern of persist purchasing the same brand of camera might be due to either cost of switching from the memory card compatibility constraint, loyalty derived from state dependence, or high intrinsic preference. Below we explain how to separately identify add-on-to-base effect from state dependence and heterogeneous brand preference.

First, theoretically speaking, the identification of add-on-to-base effect, state dependence and brand intercept comes from distinct sources of data variation. The A-to-B effect is identified by variantion of inventory of memory card conditional on the compatible camera inventory (past purchase). For example, let's consider two consumers (X and Y) who both adopted a Sony camera. X has 1 Memory Stick and Y has 2 Memory Sticks. Our model implies that Y is more likely to continue purchasing Sony cameras than X. This suggests that when we control for state dependence, the A-to-B effect still exists and can be identified given enough variation of memory card inventory level. On the other hand, state dependence is manifested in persistent purchases of the same brand of cameras regardless of the memory card inventory accumulation. Again, consider two consumers (X and Y), neither of whom has a memory card. X has a Sony camera and Y has no camera. If X is more likely to buy another Sony camera than Y, state dependence factor is identified. Furthermore, as is well documented in marketing and economics literature (Dubé et al. 2010, Paulson 2012) that structural state dependence can be separately identified from unobserved heterogeneity if we know consumers' initial brand choice and there exists enough price variation to induce switching behavior. In our case, we are fortunate that the sample was collected at the beginning of the digital camera and memory card market. Hence we observe consumers' initial brand choice directly from the data. After carefully taking care of unobserved heterogeneity by segmenting consumer brand preferences into multiple classes (latent class approach), we are able to identify both state dependence and brand intercepts. In summary, different dimensions of data variation allow us to pin down add-on-to-base effect, state dependence and unobserved brand preference.

Second, we conduct a Monte Carlo simulation to show the ability of our model to separately identify add-on-to-base effect from state dependence and brand preference. The market structure is set to be different from real data: there are two brands of cameras and two compatible memory card standards. Our simulation scheme is as follows: First, we simulate price and quality series data, based on the following transition probability $pc_t = \begin{bmatrix} 0.76 & 0.28 \\ 0.16 & 0.56 \end{bmatrix} * pc_{t-1} + \eta_{pt}$,$\eta_{pt} \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.75 \end{bmatrix}\right)$,$pm_t = \begin{bmatrix} 0.72 & 0.24 \\ 0.16 & 0.40 \end{bmatrix} * pm_{t-1} + \iota_{pt}$,$\iota_{pt} \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.75 \end{bmatrix}\right)$, $qc_t = \begin{bmatrix} 0.96 & 0.36 \\ 0.24 & 0.92 \end{bmatrix} * qc_{t-1} + \eta_{qt}$, $\boldsymbol{\eta_{qt}} \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.75 \end{bmatrix}\right)$,$qm_t = \begin{bmatrix} 1.078 & 0.245 \\ 0.147 & 1.029 \end{bmatrix} * qm_{t-1} + \iota_{qt}$, $\boldsymbol{\iota_{qt}} \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.75 \end{bmatrix}\right)$.

We generate the price and quality series for 8 time periods. We use the utility specification as in part 4.2. Given the price/quality series, we compute the observable part of the value functions. We then generate the value function by simulating the Type I extreme value error term $\varepsilon_{it}^{c,m}$. We simulate the purchasing behavior of 1200 individuals. Using the computed values of the $V_{it}'s$, we decide the timing of purchase by comparing $V_{it}^{c,m}$ with $V_{it}^{0,0}$ (outside option of no purchase). We generate 50 data sets for the same values of the parameters.

The results are shown in Table A3. All estimates are within two standard deviations from the true values. This result demonstrates the ability of our model to recover the quality, inventory, and price coefficients as well as state dependence.

The simulation result reveals that our model can separately identify add-on-to-base effect from state dependence. Essentially, the A-to-B effect is identified by variance of inventory of memory card conditional on the camera inventory. For example, let's consider two consumers (X and Y) who both adopted a Sony camera. X has 1 Memory Stick and Y has 2 Memory Sticks. Our model implies that Y is more likely to continue purchasing Sony cameras than X. This suggests that when we control for state dependence, the A-to-B effect still exists and can be identified given enough variation of memory card inventory level.

Table A1.3 Simulation Results

Parameters	True Value	Estimates	Std
Camera Intercept 1	6.8	7.189	0.479
Camera Intercept 2	2	1.794	0.595
Memory Intercept 1	4.2	4.518	0.177
Memory Intercept 2	1.2	1.411	0.122
Camera Quality	1.7	1.961	0.175
Memory Quality	0.3	0.314	0.026
Camera Inventory	1.1	1.334	0.136
Memory Inventory	0.5	0.526	0.028
Price	-3	-3.160	0.094
State Dependence	1.5	1.684	0.197

Third, there exists rich variation in our data to identify all three types of parameters. As we know from previous literature (Dube et al. 2010), state dependence is identified from persistent choices of less-favored brands (The brand chosen by a consumer at the regular price (rather than discounted price) is the favored brand) after price promotion ends. In our sample there are 91 consumers who have switched from the favorite brand to the less favorite brand of camera and stayed with the less favorite brand for the replacement choice. The purchase incidences of these consumers help identify the state dependence effect. In addition, for all the 126 consumers who have camera replacement purchases, the number of memory cards owned ranges from 0 (11.90%), 1 (14.29%), 2 (25.40%), 3 (38.89%), 4 (4.76%), to 5 (4.76%). The variance in the memory card inventory can help identify the add-on-to-base effect. Different from CPG, consumers' adoption and replacement of durable goods is relatively infrequent. Should we have a larger sample, we could have achieved higher statistical power for the state-dependence parameter (as you can see from Table 5 of the paper, the standard deviation of the state dependence parameter is rather large). But our estimates show that the data is still rich enough to separately identify add-on-to-base effect from state dependence. The last but not least, brand preference is identified by the average market share. All purchase incidences in the sample contribute to identification of the brand preference.

A1.3. Focal Store Assumption

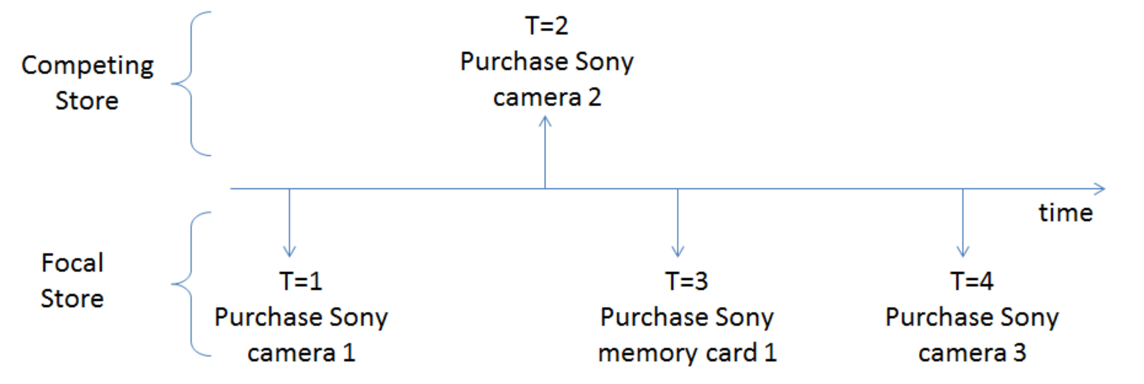
To justify that our observed data pattern from this store is representative of the industry average, we collected evidence which suggests that consumers generally replace cameras every three years. For example, Walmart has a 3 Year Replacement Plan for a Camera/Camcorder (http://www.walmart.com/ip/3-Year-Replacement-Plan-for-a-Camera-Camcorder/10227063). Our sample ranges from 4th quarter of 1998 to 4th quarter of 2004. However, from 1998 to 2000, only 115/1059=10.9% camera transactions took place. So the majority of the transactions happened within the four year range between 2001 and 2004. In this sense, most consumers (75.00%) only bought one camera and one memory card is consistent with the industry average. We collect additional information from consumer online forums to confirm that the purchase frequency observed in our data set is consistent with reality. In our sample, the average length for a consumer to replace a camera is 4.67 years. From four digital camera forums[84], we obtain a sample of 26 data points. The average replacement cycle is 4.30[85] years and the standard deviation is 2.28 years. This gives us more confidence that our data pattern is consistent with the reality.

We also want to emphasize that unlike most existing studies of durable goods that rely on aggregate sales of manufacturers (brands), we have individual panel data on consumer purchase history from a retailer. This unique dataset allows us to conduct the micro study of consumer decision process of cross-category purchases. We expect the results to be more prominent if we have complete purchases of camera and memory cards from all the retailers. Even if we do miss some consumer purchases of cameras from competing stores, we believe that this won't bias our add-on-to-base effect parameter. We explain the logic with the following example.

In the timeline below, if we interpret the purchase of a Sony camera at T=4 is because of a memory card purchase at T=3 rather than an unobserved purchase at a competing store at T=2, we might misattribute state dependence effect to the add-on-to-base effect. However, we argue that this doesn't cause much of a problem because in our dataset, we never encounter a case when the state variable, inventory of camera is 0, the consumer chose to purchase memory card. In other words, a consumer's purchase sequence never starts with a "Only Memory Card" case. Therefore, there's always another purchase occasion of a compatible camera purchase before that. In the timeline, it implies that the T=1 occasion of purchasing a Sony camera. As a consequence, although the consumer purchased camera 2 at a competing store, this doesn't change her camera inventory state variable (her camera inventory is still 1 and all else are kept the same). More specifically, the state dependence term $\kappa_i * I(c = \bar{c}_{it})$ in equation (3) are the same no matter we observe the T=2 event or not. With the state dependence effect taken into account, we can identify the add-on-to-base effect based on the variation in the memory card inventory.

[84] http://www.dpreview.com/forums/thread/3085393, http://www.travelblog.org/Topics/18677-1.html, http://forums.steves-digicams.com/general-discussion/44234-how-often-do-you-buy-new-digital-camera.html#b, http://www.twopeasinabucket.com/mb.asp?cmd=display&thread_id=3092220

[85] One thing to know is that many consumers on these forums are photographers who replace more frequently than the average consumers



That said, we also want to repeat that our data are collected at the beginning of the digital camera industry and from the chain store that has the largest market share. So the missing data problem is probably negligible.

A1.4. Other Utility Specifications

We assumed an additive utility function. However, it could be possible that cameras and memory cards are strict complement for some consumers so a standard alone camera doesn't provide any utility. We test this hypothesis with the following two specifications, conditional additive and multiplicative (Yalcin, Ofek, Koenigsberg and Biyalogorsky 2013) utility functions. However, we find that our proposed additive model fits the data better (measured by AIC and BIC) than the other two alternatives (Table A4-1). We show the details of the two specifications in equations A4-1 to A4-8. The parameter estimates are shown in Table A4-2 and A4-3.

Table A1.4-1 Model Comparison

	Additive	**Conditional Additive**	**Multiplicative**
-LL	6390.99	6411.87	6613.58
AIC	12858.68	13752.01	14209.32
BIC	13500.59	13901.85	14177.94

Conditional Additive

$$\bar{U}_{it}^{c,0} = \underbrace{U_{it}^{c}}_{camera\ new} * I(INVM_{it} > 0) + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory} * I(c\sim m) - \lambda_i PC_t^c \tag{A4-1}$$

$$\bar{U}_{it}^{0,m} = \underbrace{U_{it}^{\bar{c}}}_{camera\ inventory} + \left[\underbrace{u_{it}^{m}}_{memory\ new} + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(\bar{c}\sim m) - \lambda_i PM_t^m \tag{A4-2}$$

$$\bar{U}_{it}^{c,m} = \underbrace{U_{it}^{c}}_{camera\ new} + \left[\underbrace{u_{it}^{m}}_{memory\ new} + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(c \sim m) + \lambda_i (PC_t^c + PM_t^m) \quad \text{(A4-3)}$$

$$\bar{U}_{it}^{0,0} = \underbrace{U_{it}^{\bar{c}}}_{camera\ inventory} * I(INVM_{it} > 0) + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory} * I(\bar{c} \sim m) \quad \text{(A4-4)}$$

Table A1.4-2 Estimation Results-Conditional Additive

	Proposed Dynamic Model					
	One Segment		Two Segments			
			Seg.1 (90.5%)		Seg.2 (9.5%)	
Parameters	Est.	SE	Est.	SE	Est.	SE
Intercept: Sony ($\boldsymbol{\alpha^1}$)	-0.219	(0.090)	-0.333	(0.044)	3.086	(0.062)
Intercept: Oly ($\boldsymbol{\alpha^2}$)	-0.633	(0.142)	-0.355	(0.046)	0.586	(0.052)
Intercept: Fuji ($\boldsymbol{\alpha^3}$)	-1.048	(0.041)	-0.824	(0.073)	-0.296	(0.024)
Intercept: Kodak ($\boldsymbol{\alpha^4}$)	-0.532	(0.104)	-0.464	(0.065)	0.855	(0.014)
Intercept: Canon ($\boldsymbol{\alpha^5}$)	-0.647	(0.100)	-0.532	(0.057)	0.731	(0.024)
Intercept: HP ($\boldsymbol{\alpha^6}$)	-1.814	(0.088)	-1.858	(0.115)	0.279	(0.065)
Intercept: Nikon ($\boldsymbol{\alpha^7}$)	-1.615	(0.084)	-1.758	(1.072)	1.314	(0.073)
Intercept: Std1 ($\boldsymbol{\theta^1}$)	-2.114	(0.059)	-2.224	(0.274)	-0.156	(0.049)
Intercept: Std2 ($\boldsymbol{\theta^2}$)	-0.451	(0.143)	-0.412	(0.338)	0.441	(0.079)
Intercept: Std3 ($\boldsymbol{\theta^3}$)	5.232	(0.087)	-0.274	(0.254)	-0.035	(0.055)
Cquality ($\boldsymbol{\phi}$)	0.601	(0.076)	0.462	(0.084)	1.167	(0.028)
Mquality ($\boldsymbol{\psi}$)	0.114	(0.028)	0.170	(0.034)	0.665	(0.002)
A-to-B: Std1 ($\boldsymbol{\gamma^1}$)	0.391	(0.049)	0.318	(0.061)	0.175	(0.010)
A-to-B: Std2 ($\boldsymbol{\gamma^2}$)	0.151	(0.077)	0.156	(0.081)	0.069	(0.003)
A-to-B: Std3 ($\boldsymbol{\gamma^3}$)	0.301	(0.081)	0.200	(0.039)	0.108	(0.007)
Price ($\boldsymbol{\lambda}$)	-2.354	(0.007)	-2.406	(0.001)	-0.647	(0.001)
State Dep ($\boldsymbol{\kappa}$)	0.044	(0.021)	0.032	(0.008)	0.051	(0.033)

Multiplicative

$$\bar{U}_{it}^{c,0} = \left[\underbrace{U_{it}^{c}}_{camera\ new}\right] * \left[\underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(c \sim m) - \lambda_i PC_t^c \quad \text{(A4-5)}$$

$$\bar{U}_{it}^{0,m} = \left[\underbrace{U_{it}^{\bar{c}}}_{camera\ inventory}\right] * \left[\underbrace{u_{it}^{m}}_{memory\ new} + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(\bar{c} \sim m) - \lambda_i PM_t^m \quad \text{(A4-6)}$$

$$\overline{U}_{it}^{c,m} = \left[\underbrace{U_{it}^{c}}_{camera\ new}\right] * \left[\underbrace{u_{it}^{m}}_{memory\ new} + \underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(c \sim m) - \lambda_i(PC_t^c + PM_t^m) \quad \text{(A4-)}$$

$$\overline{U}_{it}^{0,0} = \left[\underbrace{U_{it}^{\overline{c}}}_{camera\ inventory}\right] * \left[\underbrace{\gamma_i^m * u_{it}^{INVM}}_{memory\ inventory}\right] * I(\overline{c} \sim m) \quad \text{(A4-8)}$$

Table A1.4-3 Estimation Results-Multiplicative

	Proposed Dynamic Model					
	One Segment		Two Segments			
			Seg.1 (90.5%)		Seg.2 (9.5%)	
Parameters	Est.	SE	Est.	SE	Est.	SE
Intercept: Sony (α^1)	2.741	(0.098)	2.500	(0.047)	8.990	(0.065)
Intercept: Oly (α^2)	2.138	(0.140)	2.628	(0.041)	3.646	(0.052)
Intercept: Fuji (α^3)	1.385	(0.038)	1.842	(0.083)	2.598	(0.025)
Intercept: Kodak (α^4)	2.446	(0.124)	2.096	(0.063)	4.600	(0.014)
Intercept: Canon (α^5)	1.923	(0.111)	2.427	(0.071)	4.040	(0.022)
Intercept: HP (α^6)	0.267	(0.097)	0.295	(0.105)	3.481	(0.070)
Intercept: Nikon (α^7)	1.340	(0.081)	0.441	(0.820)	4.748	(0.068)
Intercept: Std1 (θ^1)	0.714	(0.057)	0.316	(0.239)	2.779	(0.043)
Intercept: Std2 (θ^2)	2.274	(0.142)	2.358	(0.398)	3.504	(0.076)
Intercept: Std3 (θ^3)	8.406	(0.085)	2.543	(0.256)	2.958	(0.057)
Cquality (ϕ)	0.487	(0.081)	0.369	(0.085)	0.946	(0.029)
Mquality (ψ)	0.103	(0.031)	0.151	(0.037)	0.588	(0.003)
A-to-B: Std1 (γ^1)	0.312	(0.046)	0.254	(0.056)	0.140	(0.010)
A-to-B: Std2 (γ^2)	0.123	(0.075)	0.124	(0.096)	0.056	(0.003)
A-to-B: Std3 (γ^3)	0.243	(0.088)	0.159	(0.041)	0.085	(0.007)
Price (λ)	-1.499	(0.007)	-1.537	(0.000)	-0.412	(0.001)
State Dep (κ)	0.043	(0.021)	0.031	(0.007)	0.050	(0.029)

A1.5. Procedure for Decomposition Analysis

Below we use the $23 cost of switching from Sony to Fujifilm as an illustration of the procedure of decomposition.

1) Price/Quality difference between brands: We assume that Fujifilm and Sony form a similar user experience and they are compatible with the same Standard 1 memory card. Therefore, after switching to Fujifilm, the consumer keeps her state dependence effect and

the add-on-to-base effect (memory card inventory) in her utility. The only difference is that Fujifilm charges a different price at a different quality level from Sony.

2) State dependence: We assume that Fujifilm matches Sony's price and quality strategy and is compatible with the Standard 1 memory card that Sony uses. But because Fujifilm provides a different user experience than Sony, the state dependence term in the consumer's utility function vanishes.

3) Add-on-to-base effect: We assume that Fujifilm provides a similar user experience as Sony's and matches price and quality with Sony. However, it is compatible with the Standard 2 memory card rather than Standard 1. When switching to Fujifilm, the consumer can no longer use her Standard 1 memory card in inventory.

We repeat the same exercise for all the other brands.

Appendix for Chapter 2

A2.1 Alternative Models

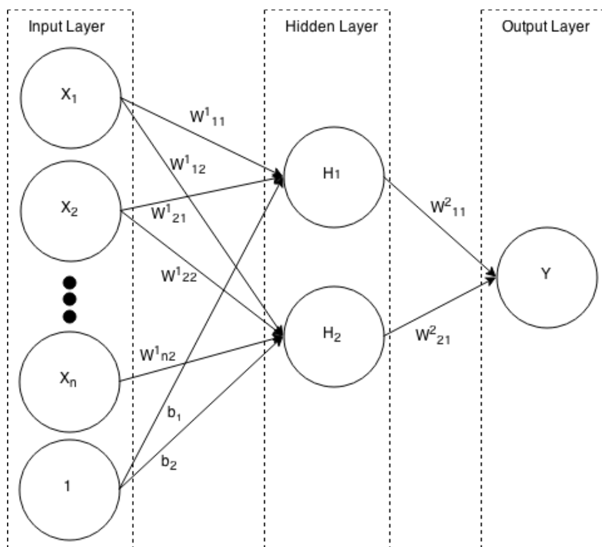
Autoregression X

We specify an autoregressive (AR) with exogenous variables model that accounts for endogeneity. We estimated the AR model of Equation 1 with two lags (optimal lag length selected by the Akaike information criterion)

$$Y_t = \beta X_t + \sum_{j=1}^{J} \phi_j Y_{t-j} + \epsilon_t$$

Because VAR model parameters are not interpretable on their own (Sims 1980), effect sizes and significance are determined through the analysis of impulse response functions (IRFs) and elasticities computed on the basis of the model (for details, see the Appendix).

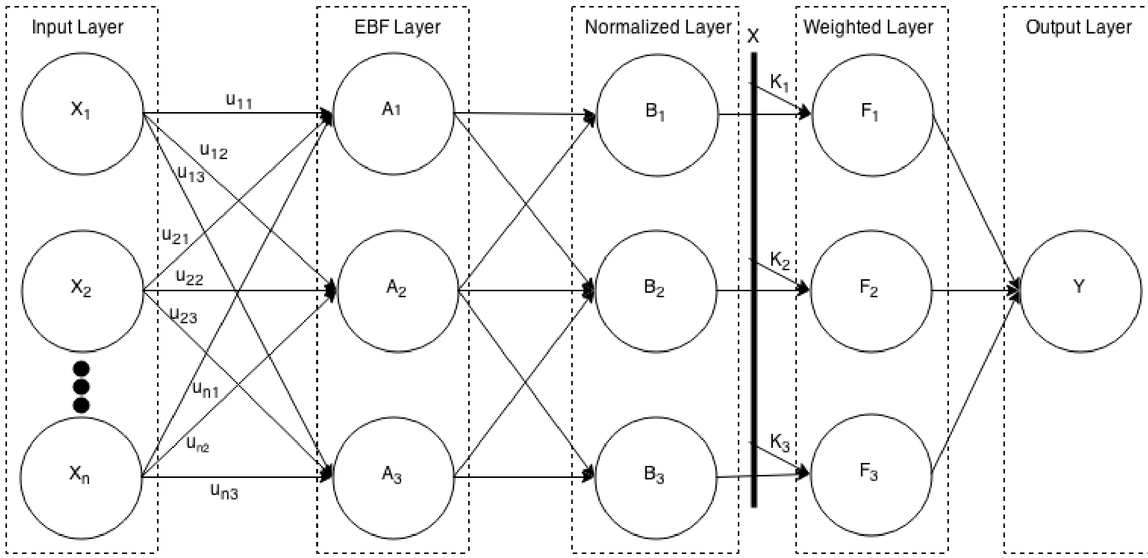
Figure A2.1 Multi-layered Feedforward Neural Networks



Feedforward neural networks is a widely used non-linear machine learning model. It is one type of artificial neural network (ANN) that was motivated by the structure of a real brain. Each neuron in the network is able to receive input signals, process them and send an output signal.

As shown in Figure A2.1, this feedforward neural network has 3 layers of neurons, including inputs $X_1, X_2, \dots X_n$, hidden neurons H_1, H_2 and output Y. Information flows in only one direction, forward, from the input neurons to the hidden neurons and to the output neuron. The connection between the i_{th} and j_{th} neuron is characterized by the weight coefficient W_{ij} and the i_{th} neuron by the threshold coefficient b_i. Each neuron performs a weighted summation of the inputs, which then passes a nonlinear activation function. For example, $H_1 = f(b_1 + \sum_{i=1}^{n} W_{i1}^{1} X_i)$, where f is the activation function. Specifically, we use the activation function $f(t) = \frac{1}{1+exp(-t)}$. The network was trained for 100 epochs with the back-propagation algorithm based on mean squared errors (Rumelhart et al., 1986). We use cross-validation to decide the number of hidden layers and the number of neurons in each hidden layer. The optimal structure of the network is two neurons in one hidden layer.

Figure A2.2 Self-Organizing Fuzzy Neural Network (SOFNN) Model



The self-organizing fuzzy neural network (SOFNN) is the five-layer fuzzy neural network shown in Figure A2.2. The five layers are the input layer, the ellipsoidal basis function (EBF) layer, the normalized layer, the weighted layer, and the output layer. It has the ability to self-organize its own neurons in the learning process.

The input layer is composed of the input vector $x = [x_1, x_2, \dots, x_n]$.

In the EBF layer, each neuron is a T-norm of Gaussian fuzzy membership functions belonging to the inputs of the network. Every neuron has both a center vector (c_{ij}) and a width vector (σ_{ij}), and the dimensions of these vectors are the same as the dimension of the input vector. Specifically, the membership function is

$$u_{ij} = exp\left[-\frac{\left(x_i - c_{ij}\right)^2}{2\sigma_{ij}^2}\right], i = 1, .., n; j = 1, \dots, u$$

and the output of the EBF layer is

$$A_j = exp\left[-\sum_{i=1}^{n}\frac{\left(x_i - c_{ij}\right)^2}{2\sigma_{ij}^2}\right], j = 1, \dots, u$$

Layer 3 is the normalized layer with output B_j as follows.

$$B_j = \frac{A_j}{\sum_{k=1}^{u} A_k} = \frac{exp\left[-\sum_{i=1}^{n}\frac{\left(x_i - c_{ij}\right)^2}{2\sigma_{ij}^2}\right]}{\sum_{k=1}^{u} exp\left[-\sum_{i=1}^{n}\frac{(x_i - c_{ik})^2}{2\sigma_{ik}^2}\right]}, j = 1, \dots, u$$

Layer 4 is the weighted layer. The neuron in this layer has two inputs, the weighted bias w_{2j} and the output of layer 3, B_j. The output F_j is the product of these two inputs, i.e.,

$$F_j = w_{2j}B_j$$

The weighted bias is an inner product of a row vector k_j and the column vector $X = [1, x_1, \dots, x_n]$, as in Takagi & Sugeno 1985.

$$w_{2j} = k_j X = k_{j0} + k_{j1}x_1 + \cdots + k_{jn}x_n, j = 1, \dots, u$$

Layer 5 is the output layer. Each neuron is a summation of incoming signals from layer 4. Thus,

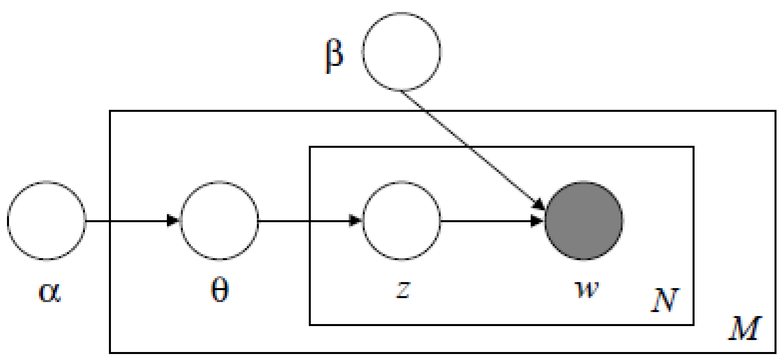
$$y(x) = \sum_{j=1}^{u} F_j = \frac{\sum_{j=1}^{u} w_{2j} exp\left[-\sum_{i=1}^{n} \frac{\left(x_i - c_{ij}\right)^2}{2\sigma_{ij}^2}\right]}{\sum_{k=1}^{u} exp\left[-\sum_{i=1}^{n} \frac{(x_i - c_{ik})^2}{2\sigma_{ik}^2}\right]}$$

The learning process of the SOFNN includes both the structure learning (find an economical network size using the self-organizing approach) and the parameter learning (using an on-line recursive least square algorithm developed by Leng et al. 2004).

LDA (Latent Dirichlet Allocation)

Latent Dirichlet Allocation (LDA, Blei, Ng and Jordan 2003) is a generative probabilistic model that uses an underlying set of "topics" to classify or summarize text documents. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words.

Figure A2.3 Latent Dirichlet Allocation Graphical Model



Specifically, it assumes that each document (composed of N words) in a collection (of a total of M documents) is generated based on the following process (Figure A2.3):

1. Choose the joint distribution of a topic mixture $\theta \sim Dir\ (\alpha)$
2. For each word w_n, $n = 1,2, \dots, N$
   a. Choose a topic $z_n \sim Multinomial(\theta)$
   b. Choose a word w_n from $p(w_n|z_n, \beta)$, which is a multinomial probability conditioned on the topic z_n.

To estimate the model, we wish to find parameters a and b that maximize the (marginal) log likelihood of the data:

$$l(\alpha, \beta) = \sum_{d=1}^{M} log[p(w_d|\alpha, \beta)]$$

We use Collapsed Gibbs sampling (Steyvers and Griffiths 2004) methods to make inferences based on the model (as implemented in the R package "lda"). After estimation, we use the topic distribution to summarize the content of the documents. Specifically, we calculate

$$p(\theta|w,\alpha,\beta) = \frac{p(\theta,w|\alpha,\beta)}{p(w|\alpha,\beta)}$$

We are also interested in the distribution of words in each topic:

$$p(w|\theta,z,\beta)$$

For more detailed information about lda, please refer to Blei, Ng and Jordan 2003.

Griffiths, T.; Steyvers, M. (2004). "Finding scientific topics". *Proceedings of the National Academy of Sciences* **101** (Suppl 1): 5228–35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.

Leng, G., Prasad, G., & McGinnity, T. M. (2004). An on-line algorithm for creating self-organizing fuzzy neural networks. *Neural Networks*, *17*(10), 1477-1493.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, (1), 116-132.

Sims, Christopher A. (1980), "Macroeconomics and Reality," Econometrica, 48 (1), 1–48

A2.2 Regression and Prediction Results for the One-Week Window

Table A2.1 Regression (with more omitted vars) 1-Week

Rating	1	2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Rl	T	G	W	I	H	Rl + T	RI + G	T + Sen	Rl+T+Sen	Topic	Rl+Topic	RI+T+G+W+Topic	PC	Rl+PC	RI+T+G+W+PC
Rating_lag	0.459						0.552	0.412		0.441		0.461	0.394		0.400	0.392
p_value	<.001						<.001	<.001		<.001		<.001	<.001		<.001	<.001
Tweets		0.003					0.001		5.38E-04	7.03E-04			4.85E-04			0.002
p_value		<.001					0.012		0.027	0.161			0.026			0.027
Google			1.87E-07					3.33E-06					1.60E-06			3.71E-07
p_value			0.026					<.001					<.001			0.000
Wiki				1.87E-05									7.13E-06			2.59E-05
p_value				<.001									0.102			<0.001
IMDB					-0.122											
p_value					0.389											
Huffington						0.059										
p_value						0.194										
Tweet_Pos									0.001	-0.001						
p_value									0.269	0.420						
Tweet_Neg									0.008	0.002						
p_value									0.462	0.853						
Tweet_PC1														0.358	0.543	0.440
p_value														<.001	<.001	<.001
Tweet_PC2														0.602	0.563	0.844
p_value														<.001	<.001	<.001
Tweet_PC3														0.968	0.404	0.629
p_value														<.001	<.001	<.001
Tweet_PC4														0.968	0.888	1.001
p_value														<.001	<.001	<.001
Tweet_T1											0.421	0.410	0.468			
p_value											<.001	<.001	<.001			
Tweet_T2											1.220	0.767	0.514			
p_value											<.001	<.001	<.001			
Tweet_T3											0.567	0.631	0.892			
p_value											<.001	<.001	<.001			
Tweet_T4											1.472	0.473	1.460			
p_value											<.001	<.001	<.001			
Tweet_T5											2.006	0.907	0.673			
p_value											<.001	<.001	<.001			
Premier	0.369	0.345	0.356	0.271	0.431	0.375	0.323	0.232	0.269	0.285	0.368	0.365	0.184	0.336	0.363	0.385
p_value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001	<.001	<.001
Finale	-0.008	-0.037	-0.022	-0.077	-0.166	-0.098	-0.003	-0.087	-0.049	-0.056	-0.021	-0.005	-0.109	-0.019	-0.024	-0.043
p_value	0.743	0.262	0.618	0.034	0.004	0.087	0.988	0.075	0.321	0.100	0.528	0.909	0.005	0.717	0.608	0.288
Age	-0.119	-0.137	-0.186	-0.165	-0.236	-0.142	-0.052	-0.153	-0.183	-0.210	-0.153	-0.133	-0.130	-0.101	0.035	-0.149
p_value	<.001	<.001	<.001	<.001	<.001	<.001	0.002	<.001	<.001	0.001	<.001	0.019	0.001	0.006	<.001	<.001

R2**	0.756	0.067	0.051	0.121	0.034	0.029	0.761	0.780	0.068	0.779	0.814	0.833	0.836	0.856	0.860	0.864
Wald Chi2	244.914	151.014	113.397	203.413	108.363	79.748	274.045	274.531	152.769	279.413	284.262	300.824	521.527	404.451	422.980	442.243
p-value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
AR(1)	-3.058						-2.883	-3.343		-3.086		-3.124	-3.312		-3.195	-3.147
p-value	0.001						0.002	0.000		0.001		0.001	0.000		0.001	0.001
AR(2)	0.047						1.536	0.277		0.451		0.209	-0.533		0.439	0.281
p-value	0.481						0.062	0.390		0.325		0.417	0.298		0.331	0.388
Sargan Chi2	24.762						20.623	20.659		23.233		17.739	7.200		25.313	21.292
p-value	1						1	1		1		1	1		1	1
MMSC-BIC	-11738						-3466	-11636		-11520		-11768	-11546		-11973	-11973

Table A2.2 Prediction: TV Series with Data from One-Week Window

Model		MAPE	MSE
	1	0.1269	0.0864
	2	0.4678	0.3362
	3	0.4834	0.3473
	4	0.4344	0.3106
	5	0.4736	0.3459
	6	0.4958	0.3857
	7	0.1180	0.0802
	8	0.1071	0.0717
	9	0.4657	0.3336
	10	0.1094	0.0733
	11	0.0947	0.0641
	12	0.0801	0.0551
	13	0.0827	0.0559
	14	0.0779	0.0516
	15	0.0719	0.0485
1 wk	16	0.0703	0.0484

A2.3 Correlation among Independent Variables

Table A2.3 Correlation

Cor	R	T	G	W	I	H	pos	zero	neg	T1	T2	T3	T4	T5	PC1	PC2	PC3	PC4	premiere	finale	age
R	1																				
T	0.09	1																			

	0.00																				
G	0.01	0.24	1																		
	0.53	0.00																			
W	0.35	0.44	0.54	1																	
	0.00	0.00	0.00																		
I	-0.01	-0.01	0.01	0.02	1																
	0.42	0.55	0.33	0.11																	
H	0.00	0.00	0.00	0.01	0.00	1															
	0.87	0.84	0.69	0.25	0.73																
pos	0.10	0.89	0.22	0.39	0.00	0.01	1														
	0.00	0.00	0.00	0.00	0.73	0.08															
zero	0.08	0.98	0.24	0.43	-0.02	0.01	0.78	1													
	0.00	0.00	0.00	0.00	0.09	0.09	0.00														
neg	0.11	0.62	0.05	0.32	-0.01	-0.01	0.55	0.59	1												
	0.00	0.00	0.02	0.00	0.09	0.07	0.00	0.00													
T1	0.44	0.18	0.27	0.27	0.03	0.01	0.22	0.16	0.52	1											
	0.00	0.00	0.00	0.00	0.06	0.09	0.00	0.00	0.00												
T2	0.22	0.15	0.11	0.30	-0.02	-0.02	0.30	0.47	0.16	0.12	1										
	0.00	0.00	0.00	0.00	0.06	0.07	0.00	0.00	0.00	0.00											
T3	0.25	0.25	0.38	0.18	0.00	0.00	0.37	0.18	0.39	0.12	0.13	1									
	0.00	0.00	0.00	0.00	0.87	0.92	0.00	0.00	0.00	0.00	0.00										
T4	0.12	0.32	0.15	0.16	0.01	-0.03	0.21	0.34	0.22	0.12	0.15	0.11	1								
	0.00	0.00	0.00	0.00	0.55	0.06	0.00	0.00	0.00	0.00	0.00	0.00									
T5	0.25	0.28	0.23	0.36	0.01	0.00	0.25	0.27	0.45	0.61	0.10	0.15	0.11	1							
	0.00	0.00	0.00	0.00	0.72	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00								
PC1	0.48	0.46	0.57	0.39	0.01	0.02	0.64	0.44	0.28	0.39	0.11	0.62	0.13	0.11	1						
	0.00	0.00	0.00	0.00	0.71	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00							
PC2	0.57	0.59	0.61	0.24	0.00	0.01	0.19	0.26	0.46	0.13	0.13	0.13	0.38	0.18	0.16	1					
	0.00	0.00	0.00	0.00	0.91	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
PC3	0.14	0.39	0.16	0.22	0.00	-0.01	0.12	0.60	0.16	0.13	0.09	0.12	0.10	0.14	0.11	0.13	1				
	0.00	0.00	0.00	0.00	0.96	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
PC4	0.12	0.24	0.48	0.49	0.00	0.01	0.26	0.21	0.23	0.12	0.12	0.56	0.32	0.10	0.17	0.11	0.11	1			
	0.00	0.00	0.00	0.00	0.89	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
premiere	0.06	0.16	0.07	0.07	-0.01	0.19	0.13	0.16	0.14	0.00	-0.01	0.01	0.05	0.01	-0.01	0.00	0.01	-0.02	1		
	0.01	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.09	0.06	0.07	0.02	0.05	0.05	0.09	0.08	0.02			
finale	-0.03	0.04	0.02	0.05	0.01	0.01	0.04	0.04	0.04	-0.02	-0.01	-0.01	0.01	-0.01	0.02	0.01	0.00	-0.01	-0.05	1	
	0.09	0.04	0.42	0.03	0.10	0.20	0.04	0.05	0.04	0.04	0.06	0.07	0.08	0.08	0.05	0.05	0.09	0.08	0.01		
age	0.26	0.11	-0.10	0.23	0.01	0.00	0.04	0.13	0.06	-0.02	0.01	-0.02	0.00	-0.03	-0.02	0.03	-0.03	0.04	-0.01	-0.02	1
	0.00	0.00	0.00	0.00	0.48	0.92	0.08	0.00	0.00	0.04	0.06	0.04	0.09	0.02	0.04	0.01	0.01	0.07	0.07	0.05	

A2.4 Details of Implementation of EMR

Machines: Amazon EC2 and EMR

Amazon Elastic Compute Cloud (EC2)[86] is one of the Amazon Web Services (AWS) that provides resizable computing capacity in the cloud. To use EC2, we simply run an Amazon Machine image using various configurations (called instance types) of memory, CPU, storage, and operating system that are optimal for our purposes. For example, we choose the m1.large instance type[87] with 7.5 GB of memory, 860 GB of local storage and 4 EC2 Compute Unit on a 64-bit Unix platform. Because it uses a pay-as you go pricing strategy, we only pay for the computing capacity based on the configuration and time of usage. EMR is Hadoop MapReduce running on EC2 instances. We use EMR because it is easily accessible and cheap compared with an onsite cluster. In addition, Mahout supports running on EMR, and all we need to do is to provide the customized java program files for our algorithms.

Cost: spot instance

To further reduce the cost of running programs on EMR, we choose the Spot Instances (aws.amazon.com/ec2/spot-instances) option of AWS. Instead of paying the high on-demand price, we bid on unused Amazon EC2 capacity. The bidding price (0.026/hour) is almost 1/10 of the on-demand price (0.24/hour).

Implementation

To select relevant Tweets, we obtained 1874 files (daily Tweets) with a maximum size of 22 GB for a single file. Thus, we provisioned a cluster of 1 master node (m3.medium) and 160 m3.2 large (m3.2xlarge) core nodes, each of which allows for 12 mappers. Note that Amazon has recommended using m3 instances, which have better performance than the previous generation m1 instances. This task takes approximately 20 minutes.

To select relevant Wikipedia pages, we obtained 51,636 files (hourly page statistics) with a maximum size of 123 MB. We continued using the provisioned cluster of 1 master node (m3.medium) and 160 m3.2 large (m3.2xlarge) core nodes. We allow each mapper to take 27 files. This task takes approximately 30 minutes.

In this revision, we choose the m3.x2 large instance, which has 30 GB of memory, because for the task to select relevant Tweets, the largest file size is 22 GB.

To perform the PCA job, we use the SSVD function in Mahout. We set the rank parameter at 100 and the number of reduce tasks at 4. The oversampling parameter is set at the default value of 15. The task finishes in 6 minutes for the TV show dataset.

[86] aws.amzon.com/ec2.

[87] For both the master node and the core nodes.

Appendix for Chapter 3

A3.1 Overdraft Fees at Top US Bank

Table A3.1 Overdraft Fees at Top U.S. Banks

Bank	Overdraft Fee	Max Fees per Day	Overdraft Protection Transfer	Continuous Overdraft Fee	Grace Period
Bank of America	$35	4	$10.00	$35	5
BB&T	$36	6	$12.50	$36	5
Capital One	$35	4	$10.00		
Capital One 360	$0	N/A	N/A		
Chase	$34	3	$10.00	$15	5
Citibank	$34	4	$10.00		
PNC	$36	4	$10 .00	$7	5
SunTrust	$36	6	$12.50	$36	7
TD Bank	$35	5	$10.00	$20	10
US Bank*	$36	4	$12.50	$25	7
Wells Fargo	$35	4	$12.50		

A3.2 Estimation Algorithm: Modified IJC

Detailed steps

1. Suppose that we are at iteration r. We start with $H^r = \left\{ \left\{ \widetilde{S}_i^k, \widetilde{V}^k \left(\hat{S}_i^k, \widetilde{S}_i^k; \vartheta_i^k \right) \right\}_{i=1}^{I} \right\}_{k=r-N}^{r-1}$ where I is the number of consumers; N is the number of past iterations used for the expected future value approximation; $\vartheta_i = \{\lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i\}$.

2. Draw μ_ϑ^r (population mean of ϑ_i) from the posterior density (normal) conditional on σ_ϑ^{r-1} and $\{\vartheta_i^{r-1}\}_{i=1}^{I}$. $\mu_\vartheta^r \sim N\left(\frac{\sum_{i=1}^{I} \vartheta_i^{r-1}}{I}, \sigma_\vartheta^{r-1}\right)$

3. Draw σ_ϑ^r (population variance of ϑ_i) from the posterior density (inverted gamma) conditional on μ_ϑ^r and $\{\vartheta_i^{r-1}\}_{i=1}^{I}$.$\sigma_\vartheta^r \sim IG\left(\frac{I}{2}, \frac{\sum_{i=1}^{I}\left(\vartheta_i^{r-1} - \mu_\vartheta^r\right)^2}{2}\right)$

4. For each $i = 1, ..., I$, draw ϑ_i^r from its posterior distribution conditional on $(C_i^d, Q_i^d, W_i^d, \mu_\vartheta^r, \sigma_\vartheta^r)$, which is

$$f_i\left(\vartheta_i | C_i^d, Q_i^d, W_i^d, \mu_\vartheta^r, \sigma_\vartheta^r\right) \propto \pi\left(\vartheta_i | \mu_\vartheta^r, \sigma_\vartheta^r\right) \rho_i\left(C_i^d | \vartheta_i\right) \rho_i\left(Q_i^d | \vartheta_i\right) \rho_i\left(W_i^d | \vartheta_i\right)$$

Since there is no easy way to draw from this posterior, we use the M-H algorithm.

(a) Draw ϑ_i^{*r} from the proposal distribution $q\left(\vartheta_i^{r-1}, \vartheta_i^{*r}\right)$ (e.g., $\vartheta_i^{*r} \sim N\left(\vartheta_i^{r-1}, \sigma^2\right)$ where ϑ_i^{*r} is a candidate value of ϑ_i^r .

(b) Compute the pseudo-likelihood for consumer i at ϑ_i^{*r}, i.e., $\rho_i^r\left(C_i^d | \vartheta_i^{*r}\right)$, $\rho_i^r\left(Q_i^d | \vartheta_i^{*r}\right)$ and $\rho_i\left(W_i^d | \vartheta_i^{*r}\right)$. Since there is no closed form solution to the optimal strategy profile, a likelihood function based on observed C_{it} becomes infeasible. Instead, we implement a numerical approximation method to establish a simulated likelihood function for estimation. For each C_{it}

observed in the data and its corresponding state point $\hat{S}_{it}$, we use the following steps to simulate its density:

i. First assume the unobserved state variables are $\widetilde{S}_{it} = \{\epsilon_{it}, \eta_{it}, \chi_{it}, \varpi_{it}\}$. Draw nr=1000 random shocks $\widetilde{S}_{it} = \{\epsilon_{it}, \eta_{it}, \chi_{it}, \varpi_{it}\}$ from

$$\eta_{it} \sim N(0, \omega_i^2), \ \ \epsilon_{it} \sim N(0,1), \chi_{it} \tilde{\ } \mathtt{EVI}^{1}, \varpi_{it} \tilde{\ } \mathtt{EVI};$$

i. For each balance checking decision $Q = \{1,0\}$and account closing decision $W = \{1,0\}$, each random draw of $\widetilde{S}_{it}$ and the observed $\hat{S}_{it}$, calculate the optimal consumption by solving the following equations

$$\begin{aligned} C_{it}^{*}\left(\hat{S}_{it}, \widetilde{S}_{it}|Q, W\right) &= arg \max_{C_{it}} \widetilde{v}^{r}\left(Q, W; \hat{S}_{it}, \widetilde{S}_{it}, \vartheta_i^{*r}\right) \\ &= arg \max_{C_{it}} U\left(C_{it}, Q, W; \hat{S}_{it}, \widetilde{S}_{it}, \vartheta_i^{*r}\right) + \beta \hat{E}^{r}_{S_{it+1}}\left\{V\left(\hat{S}_{it+1}, \widetilde{S}_{it+1}; \vartheta_i^{*r}\right)|\hat{S}_{it}, \widetilde{S}_{it}\right\} \end{aligned}$$

iii. Using the calculated nr = 1000 optimal $C_{it}^{*}\left(\hat{S}_{it}, \widetilde{S}_{it}\right)$, simulate $\rho_i^r\left(C_{it}^d|\vartheta_i^{*r}\right)$, the density of the observed C_{it}^d, using a Gaussian kernel density estimator. (This simulation borrows an idea from Yao, Mela, Chiang and Chen (2012)) . Moreover,

$$\rho_i\left(Q_{it}^d|\vartheta_i^{*r}\right) = \frac{1}{nr}\sum_{\eta,\varepsilon,\varpi}\frac{exp\left\{\overline{v}_{\chi_{it}}\left(C_{it}^{*}, Q=Q_{it}, W_{it}^{*}; \widetilde{B}_{it}, R_{it}, \Psi_{it}, Y_{it}, DL_{it}, D_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \varepsilon_{it}, \varpi_{it}, \vartheta_i^{*r}\right)\right\}}{\sum_{Q\in\{1,0\}} exp\left\{\overline{v}_{\chi_{it}}\left(C_{it}^{*}, Q, W_{it}^{*}; \widetilde{B}_{it}, R_{t}, \Psi_{it}, Y_{it}, DL_{it}, D_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \varepsilon_{it}, \varpi_{it}, \vartheta_i^{*r}\right)\right\}}$$

and

$$\rho_i\left(W_{it}^d|\vartheta_i^{*r}\right) = \frac{1}{nr}\sum_{\eta,\varepsilon,\chi}\frac{exp\left\{\overline{v}_{\varpi}\left(C_{it}^{*}, Q_{it}^{*}, W=W_{it}; \widetilde{B}_{it}, R_{it}, \Psi_{it}, Y_{it}, DL_{it}, D_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \varepsilon_{it}, \chi_{it}, \vartheta_i^{*r}\right)\right\}}{\sum_{W\in\{1,0\}} exp\left\{\overline{v}_{\varpi}\left(C_{it}^{*}, Q_{it}^{*}, W; \widetilde{B}_{it}, R_{t}, \Psi_{it}, Y_{it}, DL_{it}, D_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \varepsilon_{it}, \chi_{it}, \vartheta_i^{*r}\right)\right\}}$$

$$\begin{aligned} \rho_i\left(O_i^d|\vartheta_i^{*r}\right) &= \rho_i\left(C_i^d|\vartheta_i^{*r}\right)\rho_i\left(Q_i^d|\vartheta_i^{*r}\right)\rho_i\left(W_i^d|\vartheta_i^{*r}\right) \\ &= \prod_{t=1}^{T}\rho_i\left(C_{it}^d|\vartheta_i^{*r}\right)\rho_i\left(Q_{it}^d|\vartheta_i^{*r}\right)\rho_i\left(W_{it}^d|\vartheta_i^{*r}\right) \end{aligned}$$

.

To obtain $\widetilde{v}^r\left(\hat{S}_{it}, \widetilde{S}_{it}, \vartheta_i^{*r}\right)$, we need $\hat{E}^r_{S'}\left\{V\left(\hat{S}'_i, \widetilde{S}'_i; \vartheta_i^{*r}\right)|\hat{S}_i, \widetilde{S}_i\right\}$, which is obtained by a weighted average of $\left\{\widetilde{V}^k\left(\hat{S}_i^k, \widetilde{S}_i^k; \vartheta_i^{*k}\right)\right\}_{k=r-N}^{r-1}$, treating ϑ_i as one of the parameters when computing the weights. In the case of independent kernels, for all $\hat{S}_i = \{B_i, \psi_i, Y_i, DL_i, OD_i, \Gamma_i, \Xi_i\}$, because B_i,Ξ_i are continuous and evolves deterministically, ψ_i and OD_i are continuous and evolve stochastically, and Y_i, DL_i, Γ_i are discrete so

$$\begin{aligned} &\hat{E}^r_{S'}\left\{V\left(B'_i, \psi'_i, Y'_i, DL'_i, OD'_i, \Gamma'_i, \Xi'_i, \widetilde{S}'_i; \vartheta_i^{*r}\right)|B_i, \psi_i, Y_i, DL_i, OD_i, \Gamma_i, \Xi_i, \widetilde{S}_i\right\} \\ = &\sum_{k=r-N}^{r-1}\widetilde{V}^k\left(B_i^k, \psi_i^k, Y'_i, DL'_i, OD_i^k, \Gamma'_i, \Xi'_i, \widetilde{S}_i^k; \vartheta_i^{*k}\right)\frac{K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*k}\right)K_{hs}\left(B'_i-B_i^k\right)f\left(\psi_i^k|\phi_i, G_i\right)f\left(OD_i^k|X_i, p_i\right)K_{hs}\left(\Xi'_i-\Xi_i^k\right)}{\sum_{l=r-N}^{r-1}K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*l}\right)K_{hs}\left(B'_i-B_i^l\right)f\left(\psi_i^l|\phi_i, G_i\right)f\left(OD_i^l|X_i, p_i\right)K_{hs}\left(\Xi'_i-\Xi_i^l\right)} \end{aligned}$$

[1]Type I Extreme Value Distribution

We repeat the same step and obtain the pseudo-likelihood $(\rho_i^r(O_i^d|\vartheta_i^{r-1}))$ conditional on (ϑ_i^{r-1}). Then, we determine whether or not to accept ϑ_i^{*r}. The acceptance probability, Λ, is given by

$$\Lambda = min\left(\frac{\pi\left(\vartheta_i^{*r}|\mu_\vartheta^r,\sigma_\vartheta^r\right)\rho_i^r\left(O_i^d|\vartheta_i^{*r}\right)q\left(\vartheta_i^{*r},\vartheta_i^{r-1}\right)}{\pi\left(\vartheta_i^{r-1}|\mu_\vartheta^r,\sigma_\vartheta^r\right)\rho_i^r\left(O_i^d|\vartheta_i^{r-1}\right)q\left(\vartheta_i^{r-1},\vartheta_i^{*r}\right)},1\right)$$

where $\pi(\cdot)$ denotes the prior distribution.

(c) Repeat (a) & (b) for all i.

5. Computation of the pseudo-value function, $\left\{\widetilde{V}^r\left(\hat{S}_i^r,\widetilde{S}_i^r;\vartheta_i^{*r}\right)\right\}_{i=1}^I$

(a) Make one draw of the unobserved state variables $\widetilde{S}_i^r$ from $\eta_i \sim N(0,\omega_i^2)$, $\epsilon_i \sim N(0,\varsigma_i^2)$, χ_i ~ EVI[2], and ϖ_i ~ EVI;

(b) Compute the pseudo expected future value at ϑ_i^{*r}.

$$\hat{E}_{S_i'}^r\left\{V\left(\hat{S}_i',\widetilde{S}_i';\vartheta_i^{*r}\right)|\hat{S}_i^r,\widetilde{S}_i^r\right\} = \sum_{k=r-N}^{r-1}\widetilde{V}^k\left(B_i^k,\psi_i^k,Y_i',DL_i',OD_i^k,\Gamma_i',\Xi_i',\widetilde{S}_i^k;\vartheta_i^{*k}\right)\frac{K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*k}\right)K_{hs}\left(B_i^r-B_i^k\right)f\left(\psi_i^k|\phi_i,G_i\right)f\left(OD_i^k|X_i,\right.}{\sum_{l=r-N}^{r-1}K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*l}\right)K_{hs}\left(B_i^r-B_i^l\right)f\left(\psi_i^l|\phi_i,G_i\right)f\left(OD_i^l\right.}$$

(c) Compute $\widetilde{V}^r\left(\hat{S}_i^r,\widetilde{S}_i^r;\vartheta_i^{*r}\right)$, using the pseudo expected future values computed in (b) and the optimal choices C_i^*,Q_i^*,W_i^*.

$$\widetilde{V}^r\left(\hat{S}_i^r,\widetilde{S}_i^r;\vartheta_i^{*r}\right) = U\left(C_i^*,Q_i^*,W_i^*;\hat{S}_i^r,\widetilde{S}_i\right)+\beta\hat{E}_{S_i'}^r\left\{V\left(\hat{S}_i',\widetilde{S}_i';\vartheta_i^r\right)|\hat{S}_i^r,\widetilde{S}_i^r\right\}$$

where C_i^*,Q_i^*,W_i^* satisfy

$$\widetilde{V}^r\left(\hat{S}_i^r,\widetilde{S}_i^r;\vartheta_i^{*r}\right) = \max_{C_i,Q_i,W_i} U\left(C_i^*,Q_i^*,W_i^*;\hat{S}_i^r,\widetilde{S}_i^r\right)+\beta\hat{E}_{S'}^r\left\{V\left(\hat{S}_i',\widetilde{S}_i{}';\vartheta_i^r\right)|\hat{S}_i^r,\widetilde{S}_i^r\right\}$$

(d) Repeat (a-c) for all i.

6. Go to iteration r + 1.

A3.3 Parallel MCMC Sampling Algorithm

Table A3.2 Algorithm: Asymptotically Exact Sampling via Nonparametric Density Product Estimation

Input: Subposterior samples, $\{\vartheta_{t_1}\}_{t_1=1}^T \sim p_1(\vartheta),\ldots,\{\vartheta_{t_M}\}_{t_M=1}^T \sim p_M(\vartheta)$

Output: Posterior samples (asymptotically, as $T\to\infty$), $\{\vartheta_i\}_{i=1}^T \sim p_1\ldots p_M(\vartheta) \propto p(\vartheta|x^N)$

Set $h=1$.
Draw $t\cdot = \{t_1,\ldots,t_M\} \overset{iid}{\sim} Unif(\{1,..,T\})$
Set $c\cdot = t\cdot$.
Draw $\vartheta_1 \sim N\left(\bar{\vartheta}_{t\cdot},\frac{h^2}{M}I_d\right)$.
for i = 2 to T do
for m = 1 to M do
Set $t\cdot = c\cdot$.
Draw $t_m \sim Unif(\{1,..,T\})$
Set $h = i^{-\frac{1}{(4+d)}}$.
Draw $u \sim Unif([0,1])$.
if $u < \frac{w_{t\cdot}}{w_{c\cdot}}$ then
Draw $\vartheta_t \sim N\left(\bar{\vartheta}_{t\cdot},\frac{h^2}{M}I_d\right)$.
Set $c\cdot = t\cdot$.
else
Draw $\vartheta_t \sim N\left(\bar{\vartheta}_{c\cdot},\frac{h^2}{M}I_d\right)$.
end if
end for
end for

A3.4 Model Comparison--Hit Rates for Non-incidences

Table A3.3 Model Comparison

	A: No Forward Looking	B: No Inattention	C: No Heterogeneity	D: Proposed
Hit Rate: Overdraft	0.893	0.81	0.925	0.939
Hit Rate: Check Balance	0.766	0.659	0.804	0.897
Hit Rate: Close Account	0.885	0.853	0.901	0.916

A3.5 Alert Data

I use the alert data to verify my assumptions and counterfactuals. Since I have estimated the model in a hierarchical Bayesian fashion in which each consumer has her own set of parameters. The estimated parameters should have reflected whether the consumer uses an alert or not. For example, a consumer with low monitoring cost is more likely to have used an alert than a consumer with high monitoring cost. Following this logic, I use the alert data to do the following analysis:

1) Verify that alerts help light overdrafters reduce the overdraft frequency
2) Verify that optimal alerts should be heterogeneous
3) Verify that optimal alerts should be dynamic

Table A3.4 Overdraft Frequency Comparison: Alert vs No Alert * Heavy vs Light

	Alert					NoAlert					
	Count	Mean	Std	Min	Max	Count	Mean	Std	Min	Max	T-stat
H	8,749	**31.57**	30.94	11	337	18,432	**29.85**	28.26	11	342	**5.59***
L	33,397	**3.09**	2.48	0	10	50,069	**3.2**	2.58	0	10	**-6.37***

Note: *: p-value<0.001

For 1), as shown in Table A3.4, I found that for light overdrafters, those who use alerts are less likely to overdraw. It confirms my finding that light overdrafters are more likely to overdraw because of inattention. And alerts can help light overdrafters reduce their overdraft frequency. In contrast, for heavy overdrafters, those who use alerts are more likely to overdraw. This may seem counterintuitive. My conjecture is that there's a selection bias. Extremely heavy overdrafters are very aware of the overdraft issue. So they tend to set up alerts to help themselves better manage their accounts. However, since they are overdrawing because of heavy discounting, alerts cannot help them much.

Table A3.5 Overdraft Frequency Comparison: Alert vs No Alert * MC High vs MC Low

	Alert			No Alert			
Monitoring Cost	Count	Mean	Std	Count	Mean	Std	T-stat
High	13,448	**3.02**	2.47	20,932	3.2	2.61	**-6.61*****
Low	19,949	**3.13**	2.49	29,137	3.19	2.56	**-2.68****

Note: ***: p-value<0.001, **: p-value<0.01

For 2), I find that alerts should work better for consumers with high monitoring cost than those with low monitoring costs. Specifically, Table A3.5 shows that consumers with low monitoring costs have more alerts. And alerts help reduce the overdraft frequency more for consumers with high monitoring cost than those with low monitoring cost. This is consistent with our proposed strategy that alerts should be heterogeneous; more alerts should be sent to consumers with high monitoring cost.

Table A3.6 Effect Comparison for Different Types of Alerts

	AlertBalanceLessThanX + AccountOverdraft	AccountOverdraft	T-stat
Mean	**2.83**	**3.09**	
Std	2.49	2.48	
Count	13,627	28,127	**-10.37*** **

Note: ***: p-value<0.001

For 3), I look at the overdraft frequency for consumer who only use the "AccountOverdraft" type of alert versus those who use both the "AccountOverdraft" alert and "AlertBalanceLessThanX" alert. The idea is that "AccountOverdraft" only alert a consumer when there's zero balance. But "AlertBalanceLessThanX" alerts consumers before reaching the zero-threshold. As shown in Table A3.6, I find that sending an alert before the zero-balance threshold is reached can significantly help consumers reduce their overdraft fee payments. This is consistent with our proposed dynamic alert strategy because I suggest that the last minute threshold alert is too late for a consumer to take any action to correct the mistake.

All in all, the alert data help me confirm my model estimates and counterfactuals that light overdrafters are more likely to overdraw due to inattention, and the proposed heterogeneous and dynamic alerts should outperform the uniform threshold alert.

A3.6 Predict Overdrafting

Instead of conditioning on overdrafting, I examine what factors can predict whether a consumer is going to overdraw and whether he/she is a heavy overdrafter or a light overdrafter.

The below logistic regressions in Table A3.7 show that:

• Being younger or a student is more likely to overdraw and be a light overdrafter.
• Having low income is more likely to overdraw and be a heavy overdrafter.
• Having longer tenure/direct deposit/more debit/credit/mortgage accounts is less likely to overdraw or be a heavy overdrafter.
• Having more debit card transactions is more likely to overdraw and be a light overdrafter.
• Checking balances frequently or a steep spending slope will be less likely to be light overdrafters.

Table A3.7 Predict Overdrafting

Dep Overdraft=1 (N=275,843, R2=0.1561)		
Var	Est.	Std.
Age	-0.0045***	0.0004
Low Income	0.2617***	0.0117
Student	0.6510***	0.0284
Tenure	-0.0023***	0.0001
Direct Deposit	-1.0780***	0.0143
Debit Card Acct	-0.0693***	0.0049
Credit Card Acct	-0.5911***	0.0127
Mortgage Acct	-0.0981***	0.0281
Debit Card #Txn	0.0053***	0.0002
Online Transfer #Txn	-0.0957***	0.0140
Balance Checking Freq	0.0030	0.0020
Spending slope	0.0236***	0.0168

Dep: OD Freq (N=56,362, R2=0.03)		
Var	Est.	Std.
Age	0.0097***	0.0008
Low Income	0.0593*	0.0237
Student	-0.3744***	0.0737
Tenure	-0.0010***	0.0001
Direct Deposit	-0.2895***	0.0357
Debit Card Acct	-0.0480***	0.0105
Credit Card Acct	-0.6464***	0.0359
Mortgage Acct	-0.2773***	0.0683
Debit Card #Txn	-0.0007*	0.0003
Online Transfer #Txn	-0.0226	0.0278
Balance Checking Freq	0.0010***	0.0000
Spending slope	0.0507***	0.0088

Note: ***: p-value<0.001, **: p-value<0.01, *: p-value<0.05

A3.7 One Period A Head Model

Following Gabaix et al. 2006's directed cognition (DC) model, I solve the problem by evaluating the utility as if each evaluation operation were the last evaluation operation. To apply directed cognition, I calculate the expected benefit and cost of each available choice alternative as if this operation were the last one executed before a final choice is taken. I call this model one-period-a-head model.

I compare the model fit and parameter estimates of the three models: myopic model, one-period-a-head model and fully forward-looking model.

Table A3.8 Model Comparison

	A: Myopic	B: One Period A Head	E: Fully Forward- Looking
Log-Marginal Density	-2943.28	-2482.09	-1758.33
Hit Rate: Overdraft	0.499	0.657	0.87
Hit Rate: Check Balance	0.405	0.705	0.841
Hit Rate: Close Account	0.66	0.691	0.758

Table A3.8 reports that the one-period-a-head model has a better model fit than the myopic model but worse than the fully forward-looking model. This suggests that when doing dynamic budget allocations, consumers have foresights by more than one day.

Table A3.9 Structural Model Estimation Results Comparison

Var	Interpretation	One Period A Head	Fully Forward-Looking
β_i	Discount factor $\frac{1}{1+exp(\lambda_i)}$	0.9998	0.9997
σ_{β_i}	Standard deviation of discount factor	0.397	0.362
ς_i	Standard deviation of the preference shock	0.285	0.257
ξ_i	Monitoring cost	0.820	0.708
ρ_i	Inattention sensitivity to lapsed time	7.491	7.865
Υ_i	Dissatisfaction sensitivity	9.366	8.061

Besides, I find that (in Table A3.9) in the one-period-a-head model, the estimated discount factor is higher than that in the fully forward-looking model. Failing to account for the full dynamics can also lead to overestimated standard deviation of the preference shock, monitoring cost and dissatisfaction sensitivity as well as underestimated inattention sensitivity to lapsed time.