Temporal Implications and Optimization of Digital Marketing Interventions

by

Jangwon Choi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2023

Doctoral Committee:

Professor Fred Feinberg, Chair
Associate Professor Inyoung Chae, Sungkyunkwan University
Assistant Professor Walter Dempsey
Associate Professor Stefanus Jasin
Associate Professor Eric M. Schwartz

Jangwon Choi

jangwonc@umich.edu

ORCID iD: 0000-0002-1316-5870

© Jangwon Choi 2023

DEDICATION

I dedicate this dissertation to my family and friends, who have never given up on me despite the numerous blunders and mistakes I have made throughout my life.

ACKNOWLEGEMENTS

I keep telling anyone who is not tired of listening that having met Fred Feinberg as my advisor is the best thing that could have ever happened to my academic life. It is truly not an overstatement that this dissertation would not have seen the light of day without his illuminating guidance and unwavering support. He has been there for me throughout the entirety of my PhD journey, not only as an academic mentor and a coauthor, but more importantly as a friend, a confidant, and someone I can turn to when I needed some life advice. Countless Zoom meetings with him in the evening that revolved around Bayesian nonparametrics, cats (one *very* special one and some others), nonsensical estimation results, video games, and the job market were some of the most defining and fun moments during my six years in Ann Arbor. Of the many great qualities he possesses as an advisor, it is his patience and empathy toward his students that I hope to be able to carry with me as I embark on my new academic career.

I would also like to express my deepest gratitude to each member of my doctoral committee. I am very grateful to Inyoung Chae who has provided me with invaluable datasets that have become the foundations of two essays of this dissertation, as well as her feedback and suggestions during our countless research meetings over the past few years. Walter Dempsey has introduced me to the field of mobile health, on which the essay on dynamic factor model is based. I also thank Eric Schwartz for making me aware of the concept of active learning, which has also helped me with the ideation of the final essay, and Stefanus Jasin for his keen insights on e-commerce and optimization. I also appreciate inputs from the marketing faculty including

but not limited to Puneet Manchanda, Justin Huang, Yesim Orhun, Anocha Aribarg, S. Sriram, and Jessica Fong.

I would be remiss if I did not acknowledge many of the friends I have met here in Michigan. My time here would have been far less enjoyable without the conversations and laughs I had with them. I am thankful to all of them, but especially Gwen Ahn for being a reliable job market buddy (i.e., willing to listen to my rants) even during the toughest moments. I will surely keep reminiscing the times I spent with Longxiu Tian, Rebecca Chae, Dana Turjeman, Prashant Rajaram, Tim Doering, Hayoung Cheon, Bruno Castelo Branco, Varad Deolankar, Yiqi Li, Yu Song, Celeste Yi, and Aishwarya Joshi.

Last but not least, I would like to thank my parents Dong Shik Choi and Miran Song, and my sister Se Hyun Choi for their love, encouragement, and support from 6,500 miles away. They have always motivated me to challenge anything life throws at me and become the best version of myself. I hope I have made them proud, as much as I am of myself at this moment.

TABLE OF CONTENTS

Dedication ii
Acknowledgements iii
List of Tables vii
List of Figures ix
List of Appendices xi
Abstract xii

Chapter 1: Introduction 1

Chapter 2: Wait For Free: A Consumption-Decelerating Promotion for Serialized Digital Media 5
2.1 Abstract 5
2.2 Introduction 6
2.3 Literature Review 10
2.3.1 Digital Media Marketing 10
2.3.2 Consumption of Serialized Digital Content 11
2.3.3 Interactions between Products: Assortment Variety and Cannibalization 12
2.3.4 Waiting For Free 12
2.4 Data and Model-Free Evidence 13
2.4.1 Data 13
2.4.2 Model-Free Evidence 16
2.5 Comic-Level Analyses 18
2.5.1 Difference-in-Differences 18
2.5.2 Causal Inference: Corrections and Checks 19
2.5.3 Estimation Results 23
2.5.4 Heterogeneous Promotional Effect Across User Groups 24
2.5.5 Time Trend of Promotional Lift 27
2.6 Episode-Level Analyses 29
2.6.1 Model Specification 29
2.6.2 Estimation and Results 33
2.6.3 Robustness Checks 38
2.7 Optimization 39
2.7.1 Optimization Settings and Objective Function 39
2.7.2 Optimization Results 40
2.8 Individual-Level Analyses 42
2.8.1 Model Specification 42

2.8.2 Data Sampling and Reweighting 45
2.8.3 Results 46
2.9 Discussion 51

Chapter 3: Comparing the Effectiveness of Retargeting and Acquisition Online Banner Ads: A Flexible Approach to Estimating Ad Stock 55
3.1 Abstract 55
3.2 Introduction 56
3.3 Literature Review 60
3.3.1 Advertising Effectiveness: From Offline to Online and Retargeting 60
3.3.2 Ad Stock 64
3.3.3 Nonparametric Estimation in Marketing 67
3.4 Data and Descriptive Analyses 70
3.4.1 Data Description 70
3.4.2 Ad Impressions 73
3.4.3 Visits 75
3.4.4 Conversions 77
3.5 Empirical Analyses Using a Bayesian Parametric Approach 78
3.5.1 Model Specification 78
3.5.2 Data Construction and Estimation 83
3.5.3 Results 85
3.6 Simulation Studies with Gaussian Process Priors 94
3.6.1 Simulation Data Generation 94
3.6.2 Gaussian Process Priors 97
3.6.3 Simulation Study Results 100
3.7 Conclusion and Future Directions 105

Chapter 4: To Whom, When, and What to Ask?: Assisting Smoking Cessation and Detecting Relapse with Real Time Interventions 110
4.1 Abstract 110
4.2 Introduction and Literature Review 111
4.3 Data 115
4.4 Exploratory Analyses on Response Quality 118
4.4.1 Correlations between Items 118
4.4.2 Straight-Line Responses 119
4.5 Bayesian Dynamic Factor Model 121
4.5.1 Model Specification 121
4.5.2 Preliminary Empirical Analyses 123
4.6 Future Directions and Discussion 127

Chapter 5: Summary and Outlook 130

Appendices 132
Bibliography 160

LIST OF TABLES

Table 2.1: Summary Statistics for the Main Dataset 15
Table 2.2: Platform-wide Consumption Metrics 2-Weeks Before and After the WFF Adoption.17
Table 2.3: Comparison of Attributes between WFF and Non-WFF Comics 20
Table 2.4: Average Promotional Effect of WFF on Free and Paid Views, With and Without Non-WFF Promotional Controls.....25
Table 2.5: Heterogeneous WFF Promotional Effects by User Group 27
Table 2.6: Coefficients for Individual-Level Cox Proportional Hazards Model.....47
Table 3.1: Distribution of Number of Visits to the Target Firm's Website Across Users 76
Table 3.2: Distribution of Number of Conversions Across Users.....77
Table 4.1: Sample EMA Items with the Strongest Positive and Negative Correlations.....119
Table A1: WFF Adoption Dates and Number of Comics for Each Wave.....132
Table A2: Variable Operationalization for Episode and Comic Characteristics 133
Table A3: Logistic Regression Results for PSM and IPW 133
Table A4: Comparison of Comic-Level Promotional Effect Estimates using Random Effects vs. Fixed Effects for α_c and α_t.....134
Table A5: Promotional Effect Estimates using 10 Random Seeds for Match Selection during CEM.....134
Table A6: Summary Statistics of Comic Attributes for WFF and Non-WFF Comics, Before and After CEM 135
Table A7: Estimates of Promotional Effect for Paying vs. Non-Paying Users Using Various Endogeneity Correction Methods 135
Table A8: Estimates of Promotional Effect for Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods 135
Table A9: Estimates of Promotional Effect for Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods, Using 5 Visit Days as Cutoff for Frequent User 136
Table A10: Estimates of Promotional Effect for Frequent (Cont.) vs. Infrequent (Cont.) Users Using Various Endogeneity Correction Methods.....136
Table A11: Estimates of Promotional Effect for Paying vs. Non-Paying and Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods 137
Table A12: Lags and Correlated Random Effects in Episode-Level Analyses.....138
Table A13: WFF Effect Time Trends Parameter Estimates and Model Fit Measures using Various Model Specifications for Episode-level Analyses 139
Table A14: Observed Heterogeneity in WFF Effects across Genres 140

Table A15: Set of Rules Applied to WFF Structure of Each Comic for Optimization 140
Table A16: Average and Median Characteristics of Optimal Promotion Set vs. Platform-chosen Promotion Set.....141
Table A17: Optimized Comic Characteristics, Revenue vs. No. Paid Views 142
Table A18: Cox Proportional Hazards Negative Sampling Simulation Study 142
Table B1: Estimation Time for 6 MCMC Chains in the Empirical Analysis.....151
Table B2: Effective Sample Sizes and Rhats for Key Parameters in the Empirical Analysis.....151
Table B3: List of Decay Functions Used in the Simulation Study 152
Table B4: Estimation Time for 6 MCMC Chains in the Simulation Study.....153
Table B5: Effective Sample Sizes and Rhats for Key Parameters in the Simulation Study.....153
Table C1: EMA Items with the Strongest Positive and Negative Correlations.....156

LIST OF FIGURES

Figure 2.1: Comparison of WFF and Non-WFF Series 9
Figure 2.2: Example of Non-WFF (Left) Vs. WFF (Right) Comics 14
Figure 2.3: Viewership Comparison After Vs. Before WFF 16
Figure 2.4: Distribution of Propensity Scores for WFF and Non-WFF Comics, Before and After IPW 22
Figure 2.5: Time Trend of Promotional Lift of WFF on Free and Paid Viewership 28
Figure 2.6: Comic and Episode-Level Time Trends of Promotional Lift of WFF on Free and Paid Viewership 34
Figure 2.7: Promotional Effect, 3rd vs. 1st Quartile, for Comic and Episode Characteristics 36
Figure 2.8: Relative Cumulative Revenue of Optimized and Platform-Chosen Sets vs. No WFF 41
Figure 3.1: Growth of Internet Ad Revenues from 2010 to 2020 57
Figure 3.2: An Example of a Retargeting Ad 59
Figure 3.3: Examples of Online Banner Ads Used by the Target Firm 71
Figure 3.4: Histogram of Number of Ad Impression Across Users 74
Figure 3.5: Histogram of Log Number of Recorded Online Activities Across Users 76
Figure 3.6: Posterior Distributions of Decay Rate Parameters 86
Figure 3.7: Posterior Median and 95% Credible Intervals of γ_d 89
Figure 3.8: Posterior Distributions of Initial Impacts of Ad Creatives c_c 90
Figure 3.9: Resulting Ad Decay Functions for Acquisition and Retargeting Ad Creatives 91
Figure 3.10: Posterior Distributions for User Heterogeneity-Related Parameters 92
Figure 3.11: Posterior Distributions for Control Variable Coefficients 93
Figure 3.12: 10 Decay Functions Used in the Simulation Study 97
Figure 3.13: True Decay Functions (Blue) and Posterior Median and 95% Credible Bands (Pink) Estimated for 10 Ad Creatives 102
Figure 3.14: Posterior Distributions and True Values of Heterogeneity-Related Parameters 103
Figure 3.15: Posterior Distributions and True Values of Control Variable Coefficients 104
Figure 3.16: Posterior Distribution and True Value of Zero-Inflation Parameter π_0, and Posterior Distribution for GP Hyperparameters η and ρ 105
Figure 4.1: Histogram of Number of Completed EMAs Across Participants by Type 116
Figure 4.2: Scatterplot of Proportions of EMAs with Positive Reports of Recent Smoking Against the Number of Completed EMAs (Left) and Scatterplot of Average of Log Number

of Cigarettes Smoked Against the Number of Completed EMAs Across Participants (Right) 117
Figure 4.3: Distribution of Responses of a User Engaging in an Almost-Complete Straight-Line Responses....120
Figure 4.4 Posterior Probabilities of Each Item Belonging to Each Latent Factor in a 10-Factor Structure....125
Figure A1: New Users Per Week Over Data Window 143
Figure A2: Time Trend of Promotional Lift of WFF on Free and Paid Viewership (All Endogeneity Correction Methods)....144
Figure A3: Distribution of Propensity Scores for WFF and Non-WFF Comics, Before and After PSM 144
Figure A4: Robustness Check: Promotional Effects, IPW-Weighted (solid) vs. No Correction (dashed)....145
Figure A5: Robustness Check: Promotional Effects, Free and Paid Models Estimated Jointly (solid) vs. Separately (dashed)....146
Figure A6: Robustness Check: Promotional Effects, None (solid) vs. Error Correlation (dashed) 147
Figure A7: Robustness Check: Promotional Effects, 800 (solid) vs. 602 (dashed) Comics Included....148
Figure A8: Robustness Check: Promotional Effects, With (solid) vs. Without (dashed) Outlier Correction 149
Figure A9: Revenue and Paid Views Convergence Plots for Genetic Algorithm Optimization.150
Figure B1: Trace Plots for Key Parameters in the Empirical Analysis 154
Figure B2: Trace Plots for Key Parameters in the Simulation Study 155
Figure C1: Posterior Probabilities of Each Item Belonging to Each Latent Factor in a 5-Factor Structure....159

LIST OF APPENDICES

Appendix A: Appendix to Chapter 2 132
Appendix B: Appendix to Chapter 3 151
Appendix C: Appendix to Chapter 4 156

ABSTRACT

This dissertation consists of three essays that focus on evaluating the impact of online behavioral interventions over time and optimizing them where the objective functions might relate to non-immediate distal outcomes. Contexts studied in this work vary from promotions on digital contents platforms and online banner advertisements to mobile questionnaires designed to improve health outcomes.

The first essay examines the promotional lift of a unique type of promotion applied to serialized digital content – which we call "Wait For Free" (WFF) – that rewards *delayed* consumption, not accelerated consumption as typical promotions do. Specifically, it offers free access to the next episode if users are willing to wait a predetermined time after consuming an episode of the target series. We construct comic-level and episode-level models that account for the time-varying nature of the promotional lift over time, while also accounting for the spillover effects and observed heterogeneity of the treatment effect in the latter. Using our model estimates, we find the optimal promotion set that maximizes revenues over an extended time horizon using a genetic algorithm. Finally, we utilize a Cox proportional hazards model that directly examines the changes in individual users' behavior along the temporal dimension. Our results suggest that this novel consumption-decelerating promotion not only boosts paid viewership for the target comic, but can also do so at the platform level, even after accounting for potential spillover effects. The main mechanism behind the success of this promotion seems

to be an increased rate of free trial among the user base and a partial conversion of such users into paying ones.

The second essay of the dissertation proposes a novel framework for accounting for the non-immediate impact of advertisements, or any marketing interventions in general. Compared to the Nerlove-Arrow framework (Nerlove and Arrow 1962) of Ad Stock that has been customarily used to link ad exposures to observable customer behaviors such as site visits or purchases, our more flexible ad stock specification allows for a deviation from the restrictive "geometric decay" assumption with a single decay rate parameter. More specifically, when applying our semiparametric model to an empirical setting where a financial firm ran acquisition and retargeting online banner ad campaigns, we find that the assumption of constant decay rate over time and identical decay rates for the two types of ads are violated. We also propose a Bayesian nonparametric approach of modeling the ad stock using Gaussian Process (GP) priors, where a subsequent simulation study shows promising results for recovering the ad decay functions.

The final essay outlines a dynamic Bayesian factor model that can be used to optimize the frequency and content of Ecological Momentary Assessments (EMAs) designed to mitigate negative health outcomes such as smoking. The goal is to individualize which items in the survey are included based on the previous responses of each participant, where a Bayesian dynamic factor model allows one to weed out redundant questions that provide little additional information on top of another and those the responses to which remain mostly invariant over time and hence need not be asked frequently.

CHAPTER 1

Introduction

Over the past decade, both the number of users with internet access and the time spent online have seen dramatic increases. These trends are expected to accelerate even further given the adoption of remote work and online shopping during the recent COVID-19 pandemic, which has largely restricted in-person gatherings and retail activities. Specifically, the number of internet users was estimated to be 4.9 billion globally in 2021, compared to only 2.17 billion in 2011, and is estimated to grow to 7.5 billion in 2030. (Zippia 2023). Similarly, in the United States, the average time spent on digital media per day has more than doubled from 2011 to 2021: from 214 minutes to 485 minutes (Statista 2023a). Meanwhile, global retail e-commerce revenues amounted to $5.2 trillion in 2021, compared to $1.3 trillion in 2014 (Statista 2023b). Given the increased traffic online and the explosion in transactions in the digital market, more and more marketers are naturally shifting the focus of their interventions from traditional spaces and media to online counterparts. A recent survey by Gartner (2022) reveals that Chief Marketing Officers were planning to spend the majority of their marketing budget in online channels as opposed to more traditional ones like radio, television, and print.

The emphasis of marketing practitioners on digital marketing, coupled with the increased availability and granularity of consumer behavior data, has increasingly attracted the attention of scholarly work within the marketing literature as well. Some key topics of literature on digital marketing include, among others, the consumption and monetization of digital media (Ameri et

al. 2019, Nevskaya and Albuquerque 2019, Pattabhiramaiah et al. 2019), consumer response to and optimization of online advertising (Manchanda et al. 2006, Aribarg and Schwartz 2019, Chae et al. 2019), and customer segmentation and targeting (Ansari and Mela 2003, Dubé et al. 2017, Ascarza et al. 2018).

While the existing work on digital marketing has uncovered interesting patterns in consumer behaviors and tools to optimize marketing mix variables accordingly, there remains a relative underappreciation of a key dimension of marketing interventions: time. Understanding and taking into account the temporal implications of any behavioral interventions is critical, as many marketing activities online do not engender an immediate response from consumers or may become detectable only after a substantial amount of time. Not properly accounting for such longer-term ramifications might lead to unprofitable actions that harm both the firm and the consumer. Fortunately, due to the increased availability of individual user-level data, firms are better equipped to than ever to tailor the frequency and types of their interactions with each consumer over time (Schwartz et al. 2017, Rafieian 2022). Relatedly, digital channels more readily allow marketers to use time as a key configuration tool for promotion or product design, as evidenced by the recent increase in popularity on creator platforms such as Spotify granting early access for new content and merch for a specific group of users.

My dissertation consists of three essays that focus on estimating the effects of and optimizing online behavioral interventions over time, collectively contributing to the digital marketing literature, which has seen a dearth of discussion on these important topics. Interventions examined in the three essays span disparate contexts such as a unique consumption-decelerating promotion on a serialized digital contents platform, acquisition and retargeting online banner ads, and ecological momentary assessments (EMAs) used in a smoking

cessation program. They utilize a wide range of statistical methodologies ranging from a mixed-effects log-linear model to Gaussian Process priors for Bayesian nonparametrics and dynamic Bayesian factor models.

The first essay evaluates the effectiveness of a unique and innovative promotion implemented for a selection of series on an online platform for serialized digital content. It differs from the vast majority of traditional promotions in that it explicitly rewards its users for decelerating their consumption: users are able to access the following episode of a comic series if they are willing to wait for a predetermined time after viewing a series. Although a series is treated once, different episodes that make up the series effectively have their own time of treatment, as users would need to progress through the series to take advantage of the promotion. Given the unique nature of the promotion, time is an extremely important dimension to consider when evaluating the effectiveness of this promotion. We first identify the direct causal effects of this promotion on the number of free and paid views using a comic-level Difference-in-Difference (DiD) specification, accompanied by a wide range of endogeneity correction methods. Then, we build a more granular episode-level log linear model that encapsulates the spillover effects of the promotion as well as the observed heterogeneity of the promotional lift, which enables us to use a combinatorial genetic algorithm to search over the optimal set of comics to treat that maximize revenue at the platform level. Our results indicate that this type of promotion can not only boost the viewership of the target comics, but also at the platform level even after accounting for potential cannibalization. Finally, we use a Cox proportional hazards model to examine how the treatment has changed the "stickiness" of individual users' consumption within comics and their moving between comics, patterns not easily uncovered via an aggregate approach.

In the second essay, we model the effectiveness of online banner ads of two different types, acquisition ads aimed at new customers, and retargeting ads focusing on users who have visited the focal website. Many advertisements are not expected to generate instantaneous behavioral reaction from the target audience. Rather, only after repeated exposure to these ads do most customers engage in observable behaviors. As such, traditional advertising literature has used the concept of "Ad Stock," a time-decaying weighted average of previous exposure, in order to relate past ad exposure or spending to the current or future impact of them. One commonly used, yet very restrictive assumption in evaluating the ad stock is that effect of previous ad exposure decays exponentially over time. We attempt to break free from this restriction by using a Bayesian semiparametric approach to estimating the decay function. We find evidence that the data we have imply a strong violation of the traditional geometric decay assumption, and that the decay is more dramatic in the initial days after the ad exposure. We also propose the use of Gaussian Process (GP) priors, which serve as a nonparametric means of recovering the decay function, and present a simulation study that demonstrates GPs' ability to recover decay functions within various families of functional forms.

The third and final essay aims to optimize when to send out ecological momentary assessments (EMAs) to each participant of a smoking cessation program and which questions to include in each. While EMAs are a useful tool for detecting the participants with the highest likelihood of lapsing, repeatedly bombarding them with the same set of questions might overwhelm them, which might act as a trigger for nonresponse. As such, we attempt to optimally spread out the activation of EMAs for each participant across time based on their responses to the previous ones, and also identify the set of questions that have the most information value at those moments.

CHAPTER 2

Wait For Free: A Consumption-Decelerating Promotion for Serialized Digital Media

2.1 Abstract

Promotions for digital goods have typically focused on enticing users to hasten their consumption. Here, we investigate the effects of a novel consumption-decelerating promotion, "Wait For Free" (WFF), applied to serialized digital content – sequences of interconnected episodes – monetized via episode-level paywalls. Specifically, customers can sample early episodes of promoted series for free, and can continue to do so by waiting a pre-specified time; or, for those unwilling to wait, by paying. We analyze large-scale viewership data from a platform that enacted WFF for a wide range of web comics. A comic-level Difference-in-Differences (DiD) analysis employing distinct causal inference approaches provides robust evidence that WFF boosts paid viewership for the promoted series, and the degree of lift varies substantially by user type and over time. A more granular episode-level analysis incorporating inter-comic spillovers and promotional lift heterogeneity, alongside combinatorial genetic algorithm utilizing its results, reveals that WFF can boost net-of-cannibalization revenue at the platform level; specifically, model-guided set of promoted comics performs up to 17.9% better than the firm-enacted one and 25.3% better than the no-promotion baseline. Finally, to understand individual user-level promotional effects, we employ a proportional hazards framework to identify the degree of within-series inertia in content consumption and switching

behavior across-series, and how these are affected by WFF.

Keywords: serialized digital content, promotion optimization, Wait For Free, digital content monetization, consumption deceleration

2.2 Introduction

The global market for online entertainment, $367B in size as of 2022 and projected to grow over 20% annually through 2028 (IMARC Group 2023), owes its rapid expansion to a surge in "smart" devices – including phones, TVs, and tablets – allowing ubiquitous access to streaming media. The digital content economy includes such linchpin platforms as Netflix for films and TV shows; Audible for audiobooks; and Roblox for developer-exclusive games. A critical but often overlooked aspect of much of this digital content is that it is *serialized*: a cohesive topic or storyline broken across multiple episodes, helping maintain user engagement through a series and "keeping them coming back for more."

Platforms for digital media employ payment schemes of varying granularity to monetize their serialized content. Several, like Netflix and Spotify, have adopted fixed-fee "all you can eat" subscription models that offer unrestricted access for all series on their sites. At the other end of this granularity spectrum lie *episode-level* paywalls, which allow users to purchase individual episodes in a series via micro-payments. Contemporary examples of such episode-level paywalls include Amazon Kindle Vella, which hosts "serial stories to read one short episode at a time" (and charges for successive episodes after users have exhausted the initial free trial episodes) and the "No Sleep Podcast," which serves horror fiction series consisting of multiple episodes that are individually for sale.

Firms monetizing serialized digital content through episode-level paywalls have a pronounced incentive to promote *paid* content consumption: unlike in "all you can eat" settings based on a fixed per-customer fees, increased paid content consumption directly translates to higher revenues. A typical approach taken by platforms to stimulate paid consumption is to offer the first few "trial" episodes in each series for free, while requiring users to unlock subsequent ones, under the theory that they may become sufficiently hooked to progress through the series by paying.

Here, we introduce "Wait For Free" (WFF), a promotional strategy that (perhaps counterintuitively) rewards *delayed* consumption by offering users additional *free* content. Since its adoption, WFF has steadily gained traction in those quintessential examples of serialized digital content – online comics and novels – where numerous large-scale domestic and multinational platforms have adopted it, one example being Tapas,[1] which serves both online comics and novels throughout North America. Traditional wisdom that promotions typically incentivize consuming revenue-generating content sooner might question the viability of WFF, even deeming it destined for failure, since it seemingly does the literal opposite.

WFF is a promotion applicable to serialized media,[2] one that gives users the option to wait a pre-specified time after viewing particular episodes to unlock the following one for free, instead of paying for it. Figure 2.1 depicts how a WFF series differs from a non-WFF counterpart, where each box in a horizontal group represents an episode in that series. At the beginning of a non-WFF series, a pre-specified number of episodes – common to all users but varying across series – are available for free views by default (navy boxes in the top panel). The remaining episodes are paid-only (red boxes in the top panel), where the only option to unlock a particular one is to purchase

[1] Because an NDA prevents disclosure of our data-providing firm, we emphasize that Tapas serves strictly as an example of an WFF-enacting business model.

[2] In our forthcoming empirical setting, WFF is applied to "completed" series, i.e., those for whom all available episodes have been uploaded to the platform.

it.

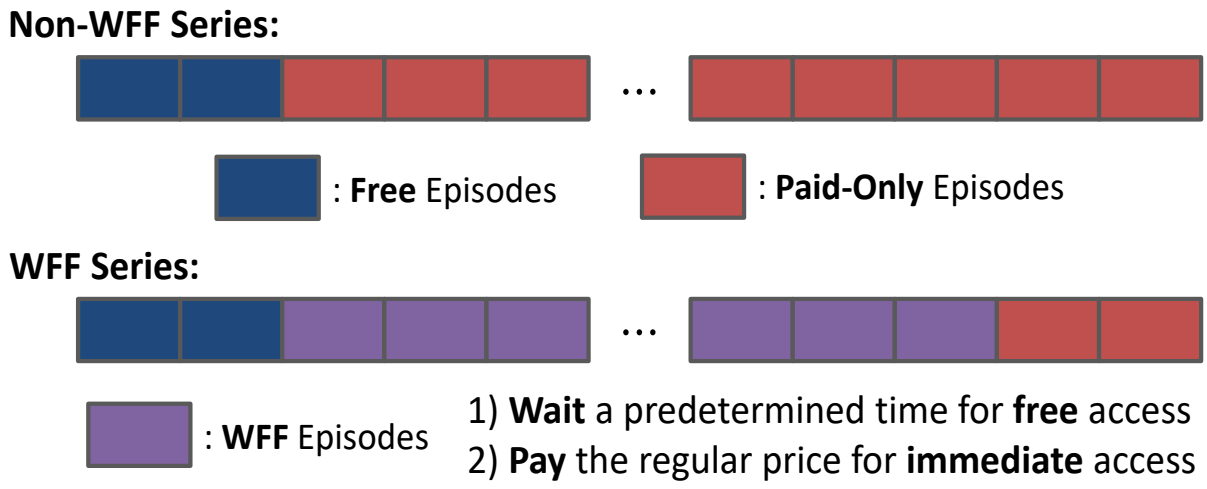
By contrast, for WFF series, the vast majority of episodes after the first few free ones are WFF ones (purple boxes in the bottom panel), where the users can either (1) wait a predetermined time after having watched the previous one to unlock for free, or (2) pay to unlock immediately. There are also a number of episodes near the end of the series accessible only if one pays for them. For instance, for the WFF series depicted at the bottom of Figure 2.1, a user may start out by viewing the first two[3] (navy) episodes for free without waiting, but the others would remain locked. A timer starts when the user starts viewing the second episode (the last of the free ones); the user may unlock the third episode (the first of the WFF ones) either by waiting for a given amount of time (e.g., 24 hours) for free access or by paying the regular price for immediate access before the timer ends. Users can continue progressing through the series using combinations of paying for immediate access or "waiting for free" until they reach the final (red) episodes, where their only option is to pay.

Allowing users to consume certain portions of a series for free via delayed consumption can entail a wide variety of downstream implications for the target series. For example, users who would have gladly paid for the content can now avail of some of it for free if they are willing to wait, leading to potential revenue loss. By contrast, users reluctant to start a series knowing the free episodes would not take them far might do so given the option of nearly completing the series for free, thereby generating a new pool of users who might potentially switch to paying for immediate access after "getting hooked." That is, WFF profitability depends on revenue loss from the first type of user being offset by gains among those who avail themselves of the free sample and choose to continue via payment. At the same time, individual series do not exist in a vacuum:

[3] The number of free, WFF, and paid-only episodes in the series depicted in the figure should be taken as illustrative examples, not suggestions of typical or best practices.

users availing of free consumption of one series may affect consumption patterns of other comics. Because users partaking in WFF may co-consume other series on the platform to differential degrees, a detailed understanding of *spillover* effects is required to determine whether WFF-type promotions are profitable at the platform-level.

Figure 2.1 Comparison of WFF and Non-WFF Series



Note: The key defining difference between WFF and non-WFF series is, for the former, the existence of the promoted episodes in the middle, which can be unlocked either by waiting a predetermined time for free, or immediately by paying.

In this paper, we examine whether the recent rollout of this unconventional approach to promoting serialized content is justified, in the context of a digital platform serving online comics. Using detailed viewership data from a large-scale platform that enacted it, we examine WFF's overall effectiveness and its implications for promotion optimization. To that end, we first develop a comic-level model focused on robust estimation of the promotional impact of designating a target series "Wait-For-Free," as well as how this varies across various user types and evolves over time. The comic-level analysis is followed by a more granular episode-level model – intended to more accurately reflect the nature of *serialized* digital media – that additionally accounts for a variety of potential spillover effects as well as across-comic-type treatment effect heterogeneity. Estimates from the episode-level model inform a combinatorial optimization for *which* comics to promote

for maximal revenue; that is, prescriptive recommendations as to what sorts of digital content appear most amenable to WFF.

Although we focus on a single digital content provider, similar practices are readily applicable to any platform serving serialized content via episode-level paywalls. As such, our findings regarding the varying nature of the promotional effects for different types of users and over time, observed heterogeneity in them, and the complex substitution patterns between content, can directly guide platforms considering similar promotional vehicles.

2.3 Literature Review

Our analysis is related to multiple streams of extant research, especially in the realm of digital media marketing; specifically, the use of free samples, consumption of serialized digital content, and interaction among products in an assortment. Here, we examine key constructs in each of these three lines of research and highlight novel aspects of our specific context (i.e., Wait For Free), so as to inform model-building and interpret eventual substantive findings.

2.3.1 Digital Media Marketing

As the demand for digital media increases, platforms have actively sought to monetize the content they serve using a variety of approaches (Lambrecht et al. 2014). While some, such as CNN or YouTube, offer (nearly all) content for free while using advertising as their main source of revenue, others, such as Netflix and the Washington Post, adopt a paywall model (Pattabhiramaiah et al. 2019, Aral and Dhillon 2021) to directly charge consumers for access. However, many attempts to ask consumers for payment have met fierce resistance, owing largely to reluctance to pay for digital content that appears freely available (Shampanier et al. 2007, Lin et al. 2013, Atasoy and Morewedge 2018).

Consequently, digital content platforms have often deployed various promotions to stimulate paid consumption. Offering portions of content as "free samples" is a common approach for information goods, as they allow users to evaluate the value of such "experiential" goods costlessly (Li et al. 2019), which may accelerate their potential purchase decisions (Bawa and Shoemaker 2004). For example, Chae et al. (2022) demonstrate that a temporary paywall suspension (i.e., offering news content for free) during a critical national event yields a significant increase in subsequent news subscriptions. Similarly, the "freemium" model for mobile apps allows consumers to self-select into accessing some content for free under various restrictions (e.g., limited selection, needing to watch ads), or using the paid version without such restrictions (Liu et al. 2014, Gu et al. 2018, Deng et al. 2022). By offering users the opportunity to evaluate its functionality for free, an app's provider hopes a sufficient proportion of trial users eventually convert to paid ones.

2.3.2 Consumption of Serialized Digital Content

In regard to the increasing prominence of *serialized* digital content, prior literature has revealed users' tendency to overindulge. For example, many consumers are prone to "binge-watching," consuming a great deal of content in one sitting or a condensed period of time (Schweidel and Moe 2016). Whereas sequentially-oriented content (e.g., film sequels; Basuroy et al. 2006) does exist in traditional media as well, the readily or instantly accessible nature of *serialized* digital media lends itself to a more temporally concentrated consumption pattern. Such behaviors can manifest when consumers find it difficult to resist the (unanticipated) temptation to keep watching successive episodes of a series, a tendency enhanced by the extensive use of "cliffhangers" in serialized content (Berger et al. 2021). For example, Zhang et al. (2021) reveal that behaviors of a substantial portion of users on an online fiction platform imply an inability to forecast future, or curtail ongoing, consumption of serialized digital content.

2.3.3 Interactions between Products: Assortment Variety and Cannibalization

Extensive research has examined the multiple ways in which one product in a line or a category might interact with others. For example, offering a diverse set leads to increased consumption (Rolls et. al 1981), particularly when the perception of variety is critical (Hoch et al. 1999, Kahn and Wansink 2004). In this sense, the mere presence of one product may indirectly benefit others and enhance total consumption of the set as a whole.

Obviously, a focal product need not always benefit from the presence of others, as they are constantly competing for share, e.g., promotions applied to one product often reduce share of others (Gupta 1988, Bell et al. 1999, van Heerde et al. 2004). Fong et al. (2019) uncover similar dynamics for digital products: while targeted promotions on e-books boost sales of the targeted and similar products, they crowd out the sales of products in a different genre.

2.3.4 Waiting For Free

To recap, the defining aspect of the focal promotion examined here, Wait For Free (WFF), is its incentivizing decelerated consumption to users willing to wait until the end of a preset time window. This approach to demand stimulation appears to defy typical, or even definitional, promotional logic (Neslin et al. 1985, Aggarwal and Vaidyanathan 2003, Bawa and Shoemaker 2004). Specifically, practices such as volume discounts (lower unit prices for buying in bulk) and limited-time deals (reduced prices for purchasing in a given time window) are commonly available in various product categories.

At the same time, WFF aligns with aspects of intertemporal price discrimination (IPD). For example, Brecko (2022) contrasts the practice ubiquitous in durable goods markets – where consumers can be "sorted" based on their tradeoff between willingness-to-pay for novel functionality vs. value-for-time – and software markets where this has proved more elusive. Similarly, Li (2019) demonstrates how IPD can be especially potent when a durable good like an

e-reader is coupled with complementary ones like e-books. As documented by Nair (2007), IPD doubtless plays a role in optimizing (dynamic) pricing for infrequently-purchased goods. Yet a deceleration-incenting promotion like WFF applied to *serialized* content may be an effective way of engendering inertia: an urge to continue consuming additional episodes shortly after finishing one, ultimately generating additional demand *above and beyond simply redistributing existing demand intertemporally*. While some users may successfully resist the temptation to binge, others, likely those who are more emotionally invested in the series, may not be as patient (Su 2007). Indeed, it may be possible to use WFF as an effective tool to differentially monetize across user types, and we investigate empirically these and related issues via a series of comic-level analyses conducted for different types of users.

A critical pragmatic task for managers implementing WFF is carefully identifying series to promote *as a group* that most benefit the platform overall, while accounting for potential interactions between series included in the promotion set and the target comics set's influence on viewership of other comics. While some extant work addresses the optimal *amount* of free content on a digital platform, or *whether* to do so at all (e.g., Prasad et al. 2003, Halbheer et al. 2014, Lambrecht and Misra 2017), relatively little concerns *which*, out of the platform's serialized content, should be offered for free. We address this key question using an episode-level viewership model that incorporates potential cannibalization and spillover from/to non-promoted ones moderated by inter-product similarity (Fong et al. 2019) combined with combinatorial optimization.

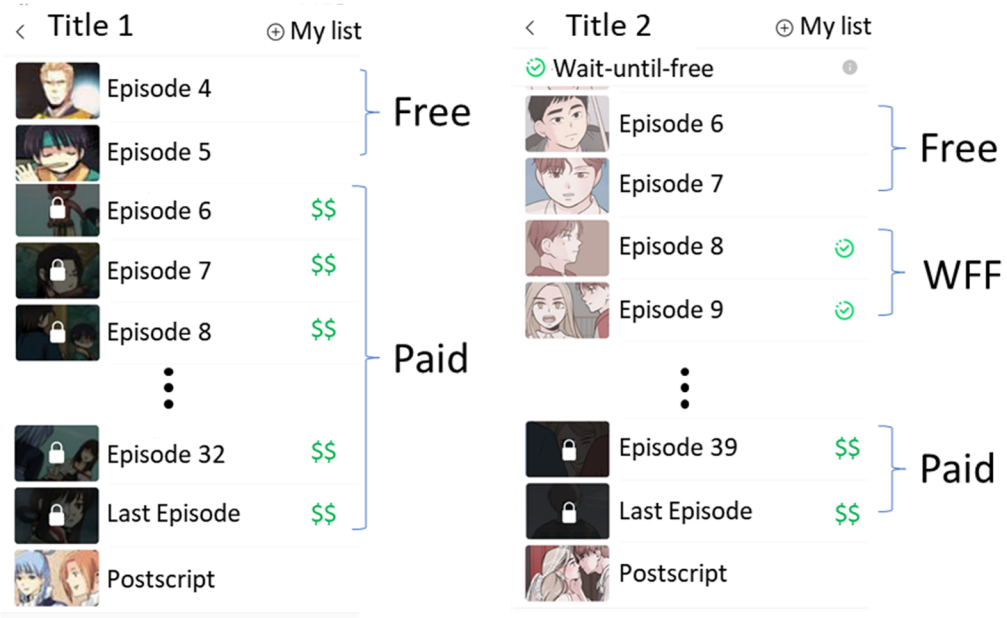
2.4 Data and Model-Free Evidence

2.4.1 Data

We analyze the effects of WFF on downstream consumption using data provided by an online

entertainment platform operating globally in North America, Europe, and Asia, focusing on their digital comic section in one of their regional markets. The platform's sole source of revenue is from the sales of its content, as it neither offers a subscription option nor generates any revenue from advertisements. The global digital comic industry has grown rapidly, with a market size of $7.9B in 2022 and projected CAGR of 6.2% for the next four years (Research and Markets 2022). The data-providing firm adopted WFF for some of its digital comics starting in November 2018 (Figure 2.2 depicts an illustrative case), and since that time similar promotions have become a common industry practice, providing some degree of empirical evidence that managers view WFF as potentially beneficial if implemented effectively.

Figure 2.2 Example of Non-WFF (Left) Vs. WFF (Right) Comics



Our data consist of four major types of information:

- User-level viewership data: records of users' viewing behavior on the platform. Includes user ID, comic and episode names, timestamp, whether or not the view was paid for.
- Comic and episode information: comic and episode names, comic genres, episode published date, artist name, and the price per episode.
- WFF information: which comics were subject to WFF, when WFF was first applied to each comic, the number of free episodes in the beginning, the number of paid-only episodes in

the end,[4] and the wait time needed to unlock the next episode (typically 24 hours).

- Non-WFF Promotion information: start and end dates of two types of non-WFF promotions applied to each comic: 1) discounts, and 2) more free episodes.

Our user-level viewership data, spanning July 16th, 2015 through May 6th, 2019, encompass the time periods both before and after WFF was first adopted by the platform. Specifically, the platform introduced WFF sequentially in smaller batches of comics across multiple waves; during our observable periods, there are 12 waves with a total of 59 comics eventually subject to WFF.[5] According to the data provider, they were 'somewhat experimental' in promoting the comics, as WFF was a new type of promotional event and they were unable to avail of performance-based feedback. To mitigate potential selectivity issues – i.e., the firm's utilizing experience gained from previous forays into WFF that were analyzed to inform subsequent comic selection – which we further address later, we focus on the first 8 waves for estimation purposes, including 35 WFF comics between June 26th, 2018 and February 25th, 2019: 19 weeks before the platform's first adoption of WFF and 16 weeks after.

The resultant dataset includes 8,327 comics with 249,313,617 total views from 1,871,557 users. Table 2.1 summarizes other pertinent aggregate information.

Table 2.1 Summary Statistics for the Main Dataset

	Mean	SD
Number of episodes per comic	21.1	34.0
Number of unique comics per user	14.8	34.9
Number of unique episodes per user	133.2	256.6
Number of users per week	402,787	36,592
Number of views per week	7,123,246	1,877,378

To arrive at a more manageable dataset without sacrificing critical model-informing information,

[4] Note that WFF was only applied to comics that have finished uploading all their episodes to the platform, possibly due to the existence of paid-only episodes towards the end of the series.

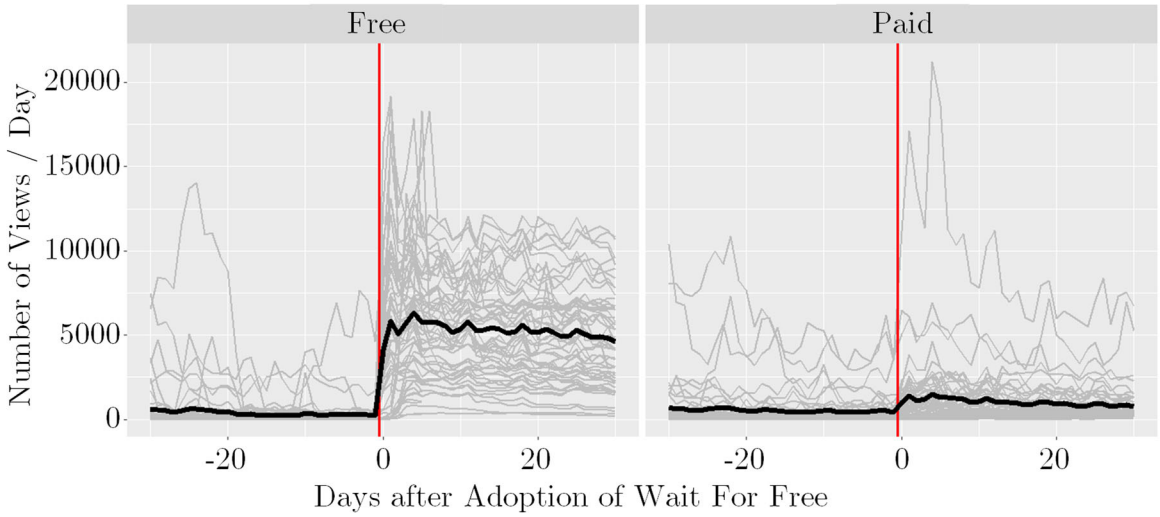
[5] The WFF adoption date and number of comics included within each wave can be found in Table A1 of the Appendix.

we focus on the top (i.e., most-viewed) 800 comics, including all 35 WFF ones for our analyses, which constituted approximately 93% of platform-level revenue during the 1-year period leading up to the platform's first adoption of WFF.

2.4.2 Model-Free Evidence

Our primary goal is to examine the impact of WFF on content consumption, both for WFF series and at the platform level. Before attempting to statistically elucidate the phenomenon, it helps to take stock of it in terms of observables. Specifically, we first explore content consumption patterns for promoted comics, comparing before vs. after WFF, for both paid and free views. Figure 2.3 depicts the changes in the per-day views of the WFF comics one month before and after the promotion, where the thick black line represents the viewership trajectory of the "average" WFF comic, and the thinner gray line the viewership for each of the WFF comics. Overall, the "average" WFF comic gets a noticeable boost in both free and paid viewership, with a more pronounced impact on the former.

Figure 2.3 Viewership Comparison After Vs. Before WFF



Note: The thin gray lines represent the viewership trajectory of individual promoted comics before and after their adoption of WFF, and the thicker black line represents the average of these individual lines.

Beyond boosting the overall readership for the WFF comics themselves, which we later document more rigorously using our comic-level analyses, we also examine whether WFF generated new demand or simply shifted it from the other non-WFF comics on the platform to the WFF ones. Table 2 compares several key metrics at the platform-level for the two weeks before and after the promotion was first adopted. Note that there is actually a *decrease* in the number of unique active users, whereas the number of views and paid views remains largely unchanged. With the caveat that this may be too brief a window to judge its full efficacy, Figure 2.3 and Table 2.2 collectively suggest that whether WFF is capable of generating new demand from its user base *in excess of cannibalization* is an open question, which we address via an episode-level viewership model that directly accounts for a variety of substitution (e.g., spillover) effects. While a small group of customers may be drawn to the platform due to the new promotion, the aggregate metrics of Table 2.2 by themselves fail to provide strong evidence for modeling such a phenomenon, as can be further verified visually in Figure A1 of the Appendix.

Table 2.2 Platform-wide Consumption Metrics 2-Weeks Before and After the WFF Adoption

	No. Unique Users	Avg. Episode Views per User	Avg. Paid Views per User	% Paid Views
2 Weeks Before	527,568	29.45	9.05	30.7%
2 Weeks After	511,300	30.09	9.27	30.8%

To reiterate, the platform's overarching goal in deploying WFF is to increase viewership across all comics on their platform (i.e., not just the WFF comics), but especially so *aggregate paid viewership*, which directly translates to revenue and ultimately profit. As such, in the next section, we first build a comic-level model that utilizes a Difference-in-Differences (DiD) approach along with a variety of causal inference (i.e., matching, weighting, multiple periods, etc.; due to the fact that the "treatment" occurs at the comic level, not the episode level) methods for

robust identification of WFF's (treatment) effects on target comics, how they evolve over time, and variation across different *a priori* subsets of users.

2.5 Comic-Level Analyses

Our initial and in some sense primary goal is to rigorously evaluate whether Wait For Free (WFF) is indeed beneficial for the promoted series in terms of generating additional viewership and revenue. Considering that the distinction between free and paid consumption is critical in evaluating the performance of any promotion, and that these quantities can be readily converted into key managerial metrics such as total viewership and revenue, free and paid views serve as dual dependent variables. We first estimate a static (i.e., time-averaged) measure of the mean overall promotional effect of WFF and for a variety of user segments, and then examine its variation over time.

2.5.1 Difference-in-Differences

We utilize Difference-in-Differences (DiD) as our main identification strategy for the *average* comic-level promotional lift of WFF. In this framework, WFF comics correspond to treated units while the non-WFF serve as control units. To estimate the average effect of WFF across all treated comics (i.e., the "average treatment effect", or ATE) and over time, we use a two-way random effects DiD model specification, separately for the number of free (F superscript) and paid views (P superscript), Y_{ct}^F and Y_{ct}^P respectively, for comic c on week t, that takes the following form (analogous for paid views):

$$\log(Y_{ct}^F + 1) = \beta_0^F + \beta^F \cdot 1\{c \in WFF\} \cdot 1\{t \in post\ WFF_c\} + \boldsymbol{X}_{ct}\boldsymbol{\gamma}^F + \alpha_c^F + \alpha_t^F + \epsilon_{ct}^F,$$
$$\alpha_c^F \sim N(0, \sigma_{cF}^2),\ \alpha_t^F \sim N(0, \sigma_{tF}^2) \qquad \text{(Eqn. 1)}$$

$\boldsymbol{X}_{ct}$ is a matrix containing a set of control variables, as elaborated below. Our focal parameter β^F captures the average lift of WFF for free views, and β^P analogously for paid views. Given that the viewership of each comic is affected by various comic-specific characteristics, such as the author and genre, as well as seasonality unrelated to WFF, we introduce comic-specific and week-specific intercepts, α_c and α_t, respectively. For estimation stability, we adopt a random effects specification for these intercepts, regularizing the comic- and week-level intercepts using a normal distribution.[6] (As per Table A4 in the Appendix, a fixed vs. random effects specification had minimal impact on the parameter estimates of interest.)

It may also be of interest from the platform's perspective to compare the promotional lifts between different types of promotions, including non-WFF ones that have occurred sporadically throughout our observation window; examples include "Discounts" (D) on paid-only episodes and "More Free Episodes" (MFE) that allow users to enjoy more episodes in the beginning of the target comics for free. As such, we include these as control variables, operationalized as the (log) number of days each of the two promotions was applied to the corresponding comic during each week.

2.5.2 Causal Inference: Corrections and Checks

A perpetual concern when estimating the effectiveness of any type of promotion using observational data is that the target might not be selected randomly, for example, in order to optimize some metric unknown to the researcher. Based on conversations with the platform, because WFF was novel and there were no data allowing for differential targeting, treated comics were not selected based on clearly articulated criteria. Although it is one of many possible selection criteria, genre classifications did not differ significantly for the WFF vs. the non-WFF comics (X^2 = 21.9, *df* = 17, $p > 0.18$).

[6] All models in this section are estimated using REML (REstricted Maximum Likelihood).

However, as a prelude to formal causal inference, we directly compare several key characteristics between WFF and non-WFF comics. These variables were chosen specifically due to both their relevance to our specific modeling context and on theoretical grounds: consumption clumpiness and number of episodes that are related to users' 'binging behavior (Schweidel and Moe 2016); conversion rate, which serves as a proxy for item quality (Liu et al. 2019); as well as price and popularity, which influence the success of a promotion in general (Ailawadi et al. 2006). Table 2.3 suggests nontrivial differences in terms of key comic characteristics, e.g., number of weeks since completion and popularity (as defined and operationalized in Table A2 of the Appendix), which calls to question any causal inference based on comparing the treated and control groups directly. Hence, we adopt a range of approaches devised to restore this comparability and explore their substantive robustness in terms of treatment effects. Specifically, we employ two matching-based methods (Propensity Score and Coarsened Exact Matching); Inverse Propensity Weighting; and doubly robust estimator especially suited to staggered treatments (Callaway and Sant'Anna 2021). Although one cannot claim fully causal inference in the absence of a randomized experiment, consistent results across distinct methodologies lend credence to direction, magnitude, and significance of model-based treatment effect estimates. We provide brief summaries of each method and more detailed information below and in the Appendix.

Table 2.3 Comparison of Attributes between WFF and Non-WFF Comics

	WFF		Non-WFF	
Consumption Clumpiness	-2.392	(0.462)	-2.351	(0.784)
No. Episodes	4.148	(0.378)	3.768	(0.765)
Weeks since Completion	3.436	(1.385)	2.542	(1.909)
Conversion Rate	-1.314	(0.916)	-1.369	(1.012)
Price	1.378	(0.049)	1.380	(0.233)
Popularity	8.226	(1.288)	7.071	(1.677)

Propensity Score Matching (PSM). PSM (Rubin and Waterman 2006; e.g., He et al. 2022) models the likelihood of a comic's being included in the portfolio, then matches a control comic with a similar propensity score to each treated one, here using a Logistic regression on whether each comic is treated against the six comic-level characteristics in Table 2.3.[7] Figure A3 of the Appendix superimposes propensity score densities for WFF and control comics with and without PSM; whereas the distributions are notably dissimilar prior to matching, they are remarkably so after.

Coarsened Exact Matching (CEM). Unlike PSM, CEM pairs treated and control comics based on observable characteristics directly (e.g., Yan et al. 2022). Here, we use median-split the first five variables of Table 2.3; for "popularity," which differs systematically between WFF comics and the controls, we use a more granular categorization of 5 bins.[8] Table A6 of the Appendix compares the summary statistics for WFF and control comics before and after CEM. While we don't see drastic differences in any of the attributes between the two groups of comics even before matching, we can readily notice a tighter comparability between the two groups of comics after matching.

"Staggered" Diff-in-Diff (Callaway and Sant'Anna: C&S). One distinct feature of our empirical context is that there were multiple waves of WFF (i.e., staggered) adoption in which a small group of comics adopted WFF, spaced across weeks, as opposed to all promoted comics adopting it at the same time. Callaway and Sant'Anna (2021) recently proposed a DiD estimator[9] that (1) directly incorporates the staggered structure of treatment, avoiding the use of yet-to-be-

[7] The logistic regression results we use for propensity score estimation are shown in Table A3 of the Appendix.

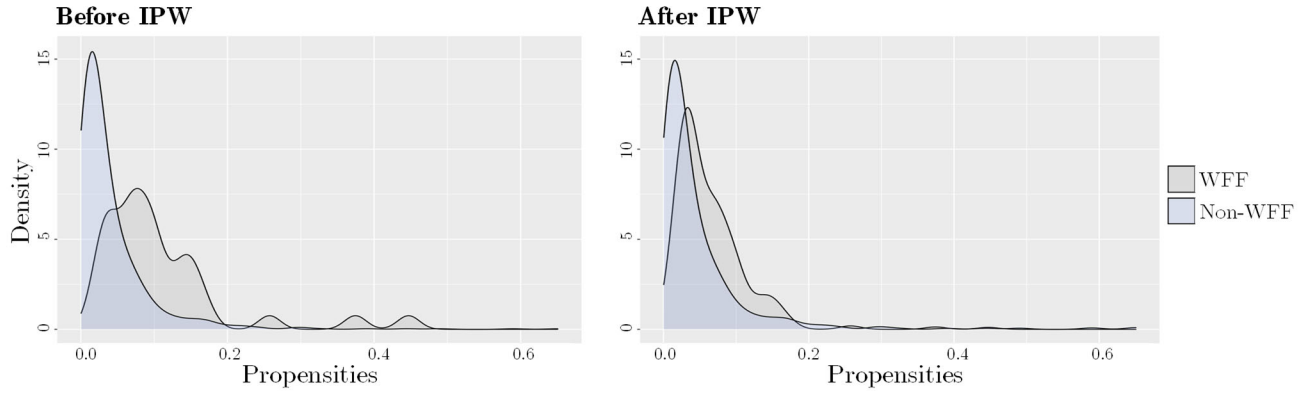
[8] When more than one control matched perfectly to a WFF comic, we randomly chose one with equal probability. Table A5 in the Appendix shows that such random inclusion does not substantively sway our results.

[9] Estimation was done via **R** package **did**, with a fixed effects specification for comic and week intercepts.

treated units as controls, and (2) is "doubly robust" to misspecification of the first-stage (for treatment propensity) and second-stage (for the outcome variable) regressions. We use this estimator with and without our 6 comic-level attributes that were used for PSM and CEM as model covariates in both stages.

Inverse Propensity Weighting (IPW). Lastly, IPW (e.g., Gordon et al. 2019, Kumar et al. 2019) mitigates the effects of potential selection with a propensity model like PSM, but probabilistically weighting (w_c) all control units using their propensities ($\hat{p}_c$): $w_c \propto \hat{p}_c^{-1}$ for $c \in$ WFF and $w_c \propto (1-\hat{p}_c)^{-1}$ otherwise. Figure 2.4 superimposes the distributions of propensity scores (using logistic regression) for promoted and unpromoted comics both before and after reweighting, showing substantial mitigation of differences across the propensity domain.

Figure 2.4 Distribution of Propensity Scores for WFF and Non-WFF Comics, Before and After IPW



Comparison of Correction Methods. While all approaches outlined so far – PSM, CEM, C&S, IPW – are valid and widely utilized to correct for potential selection artifacts, they are not equally suitable for our context. "Matching" methods (PSM, CEM) substantially restrict the scope of analyses to a small fraction of comics (70 & 66 of 800, respectively) as non-matched comics are dropped altogether, a 90+% data loss; that only 35 WFF comics were treated further exacerbates information loss of under such approaches. On the other hand, whereas Callaway and Sant'Anna's

(2021) estimator accommodates the staggered nature of WFF adoption, it focuses on estimating the average treatment effect (ATE), while being agnostic to *which* observable characteristics respond well to the promotion, a critical input for eventual optimization and our focal research questions. More importantly (especially for the upcoming episode-level analyses), C&S's estimator does not account for the unique and defining structure of serialized content – multiple episodes nested within each series – and cannot capture the fact that when WFF is adopted at the comic level does not necessarily coincide with when individual episodes within the series start benefiting from it.

In contrast, IPW offers distinct advantages that circumvent these pitfalls. First, unlike PSM and CEM, it enables utilization of the full set of comics to inform parameter estimates, which is critical when assessing heterogeneity in promotional lifts. Second, compared to C&S, IPW allows for more flexible model specifications, as it only entails using different weights for each data point, regardless of the overall model specification. That is, the overarching model can not only directly accommodate the nested structure of episodes within comics, but also the lift heterogeneity critical for promotional set optimization. Thus, as seen below, while all tested estimators lead to roughly the same average treatment effect of WFF, we focus mainly on those obtained via IPW so as to extend readily to the forthcoming episode-level model.

2.5.3 Estimation Results

Table 2.4 summarizes the parameter estimates for the promotional lifts of WFF ($\hat{\beta}^F$) and the other two types of non-WFF promotions ($\hat{\gamma}_D^F$ for Discount and $\hat{\gamma}_{MFE}^F$ for More Free Episodes) on the platform across different correction strategies and model specifications (i.e., without and with

controls for non-WFF promotions).[10] Given that our models take the log of free and paid views as the dependent variables, coefficients are exponentiated for interpretation as interval-scaled multipliers. For example, IPW estimates using Logistic weights indicate that the WFF-type promotion increases the number of free views for target comics by around $\exp(\hat{\beta}^F) - 1 \approx 33.61$ times and the number of paid views by around $\exp(\hat{\beta}^P) - 1 \approx 1.31$ times.

Parameter estimates are remarkably robust across various correction methods, in directionality, significance, and magnitude. Moreover, when comparing the two model specifications, not accounting for non-WFF promotions seems to lead to an underestimation of promotional lifts for WFF, most likely due to elevated viewership from other promotions being part of the "baseline" viewership. We also notice that, not surprisingly, both types of promotions – discount and more free episodes – are estimated to have a positive impact on both free and paid viewership, albeit with varying magnitudes.

2.5.4 Heterogeneous Promotional Effect Across User Groups

While we take an aggregate approach that focuses on WFF's platform-level effects, one important managerial question concerns for which types of customers WFF-type promotions are differentially effective. While the platform providing our data decided against offering WFF to only a portion of its customers, doing so might well be a valuable tactic in other contexts.

Two dimensions of users' platform activities are especially informative in addressing this question: payment incidence and visit frequency.[11] Paying (vs. Non-Paying) users are defined as those purchasing at least one episode during the one-year period prior to the

[10] Results are not available under the Callaway and Sant'Anna (2021) framework for the latter, as the **did** package in **R** does not provide coefficient estimates for control variables.

[11] Whereas we examine each of the dimensions separately here, more granular analyses overlaying the two dimensions and categorizing users into 4 groups appears in Table A11 of the Appendix.

Table 2.4 Average Promotional Effect of WFF on Free and Paid Views,
With and Without Non-WFF Promotional Controls

Correction	None		PSM		CEM		C&S (No Covariates)	C&S (6 Covariates)	IPW	
Non-WFF Controls	–	Yes	–	Yes	–	Yes	–	–	–	Yes
	Free Views									
$\hat{\beta}^F$	3.303	3.696	3.353	3.660	3.347	3.656	3.639	3.724	3.544	3.864
	(0.071)	(0.056)	(0.089)	(0.080)	(0.097)	(0.086)	(0.256)	(0.258)	(0.030)	(0.026)
$\hat{\gamma}_D^F$		0.490		0.435		0.235				0.441
		(0.023)		(0.064)		(0.077)				(0.026)
$\hat{\gamma}_{MFE}^F$		1.706		1.273		1.456				1.501
		(0.014)		(0.053)		(0.056)				(0.016)
Conditional R^2	0.7607	0.8466	0.8019	0.8446	0.6936	0.7727	–	–	0.7355	0.7997
	Paid Views									
$\hat{\beta}^P$	0.797	0.939	0.807	0.916	0.746	0.849	0.844	0.830	0.836	0.959
	(0.043)	(0.040)	(0.044)	(0.040)	(0.042)	(0.039)	(0.091)	(0.092)	(0.016)	(0.015)
$\hat{\gamma}_D^P$		0.587		0.492		0.388				0.519
		(0.016)		(0.032)		(0.035)				(0.015)
$\hat{\gamma}_{MFE}^P$		0.538		0.411		0.407				0.467
		(0.010)		(0.026)		(0.025)				(0.009)
Conditional R^2	0.8212	0.8367	0.8356	0.8561	0.7974	0.8168	–	–	0.8434	0.8563
No. Observations	27,463		2,436		2,310		27,463	27,463	27,463	

platform first adopting WFF, yielding 822,068 Paying (28.2%) and 2,088,975 Non-Paying users (71.8%). One key role of promotions is enticing users who have never paid to "convert," after which they may become less resistant to purchasing additional episodes on the platform in the future (Dhar et al. 2007). As such, platforms considering a WFF-type of promotion would wish to gauge its payment conversion capability, especially when the ratio of non-paying to paying users is high (here, 2.5).

"Frequent" users are taken as those who've viewed at least one episode on the platform on at least 10 distinct calendar days during the 1-year period before the WFF adoption,[12] yielding 914,035 (31.4%) Frequent and 1,997,008 (68.6%) Infrequent users. WFF-type promotions might be more impactful for Frequent than for Infrequent users, given that the former type might, ceteris paribus, have a higher likelihood of promotional awareness and reaction during the observation window. Given that results for the Infrequent group might be affected by some of them never returning to the platform post-WFF, we also examine a specific subset of each user group – those who visit the platform at least once post-WFF-adoption – the full results of which appear in Table A10 in the Appendix.

To assess the heterogenous promotional impact of WFF for different types of users, we construct separate datasets based on user group activity and re-estimate the model in Eqn. (1). Table 2.5 lists the estimated coefficients of β^F and β^P for each user group (analogous results using alternative endogeneity correction approaches appear in Tables A7 and A8 of the Appendix.

As per Table 2.5, WFF is slightly more effective in increasing free viewership for WFF comics among Paying vs. Non-Paying users, but the latter increase their paid viewership to a far

[12] Here we chose a cutoff that led to a similar breakdown to that of the paying vs. non-paying distinction. Table A9 of the Appendix contains a robustness check using a different cutoff from that used in our main analyses.

greater degree. Paying users generally have a higher interest in comics, evidenced by the fact that the average Paying user also views more episodes for free (252.2) than the average Non-Paying user (80.6) does. In this regard, Paying users may be less likely to pass up an opportunity to view an episode for free. On the other hand, Paying users might already be fully utilizing their budget for this platform, whereas Non-Paying users – who may be differentially reluctant to purchase platform content – might constitute an untapped revenue source. This finding is important managerially as it suggests that WFF can be an effective way of increasing the *paying* customer base, one that may prove to be accessible into the future and could very well explain why many similar platforms have adopted such promotional practices in recent years.

Table 2.5 Heterogeneous WFF Promotional Effects by User Group

Dimension	User Type	Free Views ($\hat{\beta}^F$)	Paid Views ($\hat{\beta}^P$)
Payment Incidence	Paying	3.846 (0.030)	0.793 (0.016)
	Non-Paying	3.410 (0.032)	1.916 (0.026)
Visit Frequency	Frequent	3.917 (0.031)	0.861 (0.016)
	Infrequent	2.580 (0.026)	0.466 (0.025)

Our conjecture that Frequent (vs. Infrequent) users would react more strongly to this type of promotion holds for both free (3.917 vs. 2.580) and paid (0.861 vs. 0.466) viewership. At the same time, it is also noteworthy that WFF seems to be driving an increase in both free and paid viewership for Infrequent users as well, albeit to a lesser degree than for the Frequent ones. Combined with the previous result, this finding again demonstrates the potential of WFF to be a catalyst for stimulating demand via expansion of the customer base.

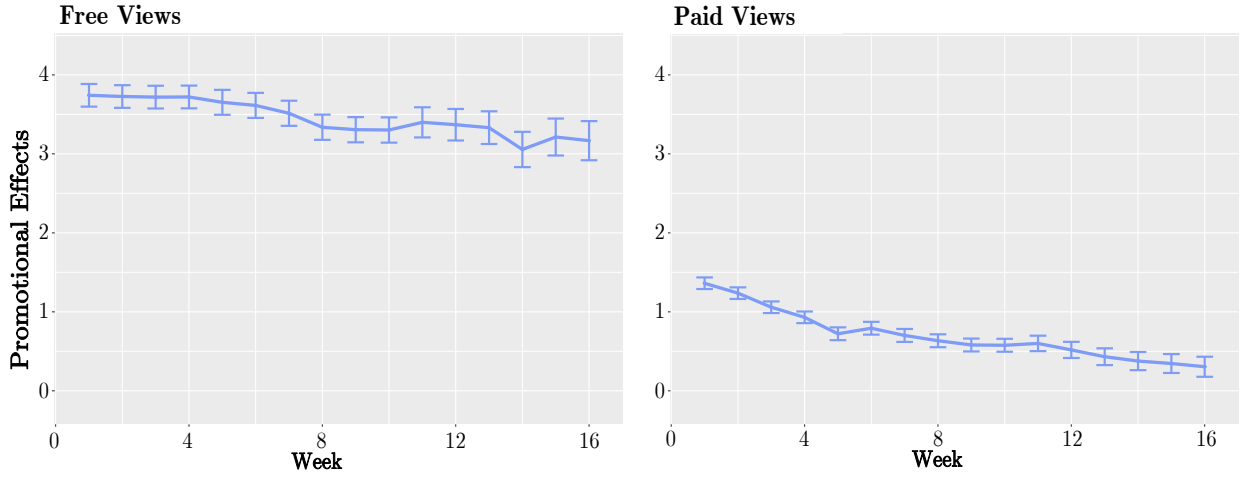
2.5.5 Time Trend of Promotional Lift

Whereas the previous analyses focused on robust causal inference of a time-invariant ATE for WFF, Figure 2.3 suggests a temporal trend, i.e., the surge in viewership dissipates over time. Specifically, we expand the WFF effect parameter (β^F) to time-varying ones $\{\beta_w^F\}$ that represent

WFF's impact on comic-level viewership on the *w*-th week post-adoption (Kumar et al. 2017). That is, the term $\beta^F \cdot 1\{c \in WFF\} \cdot 1\{t \in post\ WFF_c\}$ in Eqn. (1) is replaced with $\sum_w \beta_w^F \cdot WFF_{ctw}^{comic}$, where $WFF_{ctw}^{comic} = 1\{c \in WFF\} \cdot 1\{t = w\text{-}\ th\ week\ post\ WFF_c\}$, a flexible specification that can account for potential nonlinearities in the promotional effect over time depicted in Figure 2.3. Note that these do not perfectly correspond to the *w*-th week in calendar time since the *platform* first adopts WFF, since comics in different waves adopt WFF at different times.

Figure 2.5 depicts the point estimates and confidence intervals of $\hat{\beta}_w^F$ (left panel) and $\hat{\beta}_w^P$ (right panel), obtained using the IPW approach (comparative estimates for alternative matching approaches appear in Figure A2 of the Appendix). Intriguingly, whereas time trends in promotional lifts are nontrivial and pronounced, especially for paid viewership, those for free views are less so. This contrast appears to suggest that, while many "patient" users consistently engage with WFF comics via free consumption over long stretches of time, there are quite a few "impatient" ones who start paying for the next episode soon after starting a WFF series.

Figure 2.5 Time Trend of Promotional Lift of WFF on Free and Paid Viewership



Note: Estimated promotional effects sizes, log scale

2.6 Episode-Level Analyses

Thus far, a series of comic-level analyses provides robust demonstration that Wait For Free (WFF) is effective in boosting adopted-comic-averaged free and paid viewership. However compelling this analysis in terms of the existence of substantial treatment effect, it is nonetheless limited in an important way for policy analysis and optimization. Specifically, it does not incorporate several key features of the data-generating process specific to *serialized* content, e.g., a more granular analysis availing of the panel structure at the *episode-week* level of viewership.

As per Figure 2.1, the basic mechanism of WFF requires that a user wait a certain amount of time to unlock the following episode (for free) after having viewed one episode. This implies that most of the users taking advantage of WFF only reach episodes toward the end of the series several weeks after the comic first adopts it. Given the serialized data structure (multiple episodes nested within a single comic), estimating only comic-level promotional lifts – without considering how and when each episode might react to WFF – may result in an incomplete account of WFF effects, in turn affecting optimization of the promotion set. Moreover, such promotional lifts can also differ by episode (within comic), across different types of comics, and entail spillover effects, e.g., of promoting one comic on other (promoted or unpromoted) ones. Capturing these episode-level platform-wide consequences of WFF is crucial to combinatorial optimization for the WFF-treated set that maximizes platform-wide revenue.

2.6.1 Model Specification

Within the class of models for per-episode viewership, we start from the simplest specification, one broadly comparable to the (episode-aggregated) comic-level analysis, later expanding it by including features specifically reflective of serialized digital media consumption. More formally, the outcome variable will be the free (superscript F; P for paid) number of views

for episode e within comic c across all users on week t, or Y_{cet}^{F}.

Comic-level vs. Episode-level WFF Effect Time Trends. To account for the fact that not all episodes fully benefit from WFF the moment comic adopts it, we decompose WFF's time trend effects along two dimensions, comic-level and episode-level. Specifically, we construct two sets of dummy variables: WFF_{ctw}^{comic} indicates that week t is the w-th week that comic c has been under WFF, with an analogous definition for $WFF_{cetw'}^{episode}$. An episode is coded as subject to WFF starting from the week in which users who tightly follow the promotion schedule (i.e., who keep viewing the following episodes as soon as they become unlockable for free) can start viewing it for free. Using these two sets of time-trend variables[13] (which only take effect for episodes subject to WFF), the updated episode-level model specification is as follows:

$$\log(Y_{cet}^{F}+1) = \beta_0^F + 1\{episode\ e\ of\ comic\ c\ is\ subject\ to\ WFF\ on\ week\ t\}\cdot \{\sum_{w=1}^{11}\beta_w^{comic,F}\cdot WFF_{ctw}^{comic} + \sum_{w'=1}^{11}\beta_{w'}^{episode,F}\cdot WFF_{cetw'}^{episode}\} + \boldsymbol{X}_{cet}\boldsymbol{\gamma}^F + \alpha_c^F + \alpha_t^F + \epsilon_{ct}^F \quad \text{(Eqn. 2)}$$

Given preliminary results indicating both time trends largely flatten out after about 5 weeks, the model takes them as constant after 10 (e.g., $\beta_{11}^{comic,F} = \beta_{12}^{comic,F} = \cdots$). Moreover, in the above parameterization, $\beta_1^{episode,F} = \beta_1^{episode,P} \equiv 0$ for identification purposes.

Control Variables. We also account for other marketing variables, such as Discounts and More Free episodes, as in our comic-level model, as well as comic and episode characteristics such as price per episode (constant within each comic); number of weeks since episode first published; and the relative position of an episode within its comic. We maintain a random effects specification for the comic-specific (α_c^F) and week-specific intercepts (α_t^F), within and across the free and paid

[13] Key coefficient estimates and fit measures for the model that includes only the comic-level time trend variables (i.e., WFF_{cetw}^{comic}) are shown in the first two columns of Table A13 of the Appendix.

viewership models, while allowing for correlations between corresponding quantities in the two models,[14] e.g.,

$$\begin{pmatrix}\alpha_c^F\\ \alpha_c^P\end{pmatrix} \sim N\left(0, \begin{pmatrix}\sigma_{cF}^2 & \rho_c\sigma_{cF}\sigma_{cP}\\ \rho_c\sigma_{cF}\sigma_{cP} & \sigma_{cP}^2\end{pmatrix}\right), \quad \begin{pmatrix}\alpha_t^F\\ \alpha_t^P\end{pmatrix} \sim N\left(0, \begin{pmatrix}\sigma_{tF}^2 & \rho_t\sigma_{tF}\sigma_{tP}\\ \rho_t\sigma_{tF}\sigma_{tP} & \sigma_{tP}^2\end{pmatrix}\right).$$

While the model parameters $\{\beta_w^{comic}\}$ and $\{\beta_{w'}^{episode}\}$ capture the evolution of promotional lift from WFF over time, allowing a fuller understanding of the impact of WFF on the target comics compared to the comic-level model, these are not sufficient to address a key managerial implementation question: *which* comics should be subject to WFF to maximize platform-wide revenues. To more fully address this requires two additional features – spillovers and heterogeneity of promotional lift across comics – be accommodated. We show in Table A13 of the Appendix that their inclusion leads to a better model fit, even when considering the added model complexity.

Spillover and Cannibalization. The magnitude of the lift from the WFF promotion can involve either a synergistic or a dampening effect depending on the focal comic's similarity to other comics that adopt WFF. For instance, it could be more effective when a "diverse" set of comics is subject to it than when concentrated on a narrowly similar set (Rolls et. al 1981). At the same time, the size of the promotional effect on each of the comics may attenuate as more comics become WFF, due to users' attention being diluted among more promoted comics (Moon et al. 2007). WFF may also affect the viewership of other comics not subject to it. For example, some users, having found interest in a set of WFF comics, might venture out to non-WFF comics that they view as similar (i.e., positive spillover; Roos et al. 2020). Alternatively, there is also the possibility of cannibalization (Fong et al. 2019), e.g., users become satiated with a certain type of comic (or comics in general) and thereby become differentially less prone to consuming non-WFF comics

[14] Parameter estimates of related quantities such as the variance and correlation of the random effects are reported in Table A12 of the Appendix.

deemed similar, especially given their limited time and budget spent on the platform. Given that the platform's goal is to maximize platform-wide revenue, not the revenue just from WFF comics, accounting for these indirect effects is critical.

We construct two variables $Sim.WFF_{cet}$ and $Sim.NonWFF_{cet}$ to capture the synergy (or dampening) effect among the WFF comics and the spillover (or cannibalization) from the WFF to the non-WFF comics, respectively. As their net effects on viewership are likely more pronounced for comics with greater "behavioral overlap" to the promoted ones, vs. those that are very distinct, we base these variables on dyadic similarity measures (i.e., between comic pairs) that are directly based on users' observed viewing patterns. Specifically, we construct cosine similarity measures (Lee 1999; e.g., Hwang et al. 2010) between comic pairs, based on individual-level data for the more than 2 million users who consume at least one episode of any comic during a 1-year period before the WFF promotion was first introduced to the platform.[15] A higher value of this measure between a pair of comics indicates more overlap in viewership between them. The similarity measures used in our model are calculated as the sum of cosine similarities between the focal comic and all comics that have adopted WFF by week t. $Sim.WFF_{cet}$ takes this sum for the WFF episodes and 0 for the others, whereas $Sim.NonWFF_{cet}$ does so for non-WFF episodes and is 0 for WFF ones.

Observed Heterogeneity Across Comics. While affirming the *average* treatment effect of WFF and understanding the dynamics of spillovers are both highly relevant for managers of platforms for serialized digital media, a key implementation question they need to answer when applying a similar promotion concerns *which* types of comics tend to benefit more from the promotion. As such, we aim to offer actionable and generalizable guidelines regarding how the magnitudes of the

[15] For each comic c, consider $n_c = (n_{c1}, \cdots, n_{cN})$, where n_{ci} represents the number of unique episodes in comic c consumed by user i. Cosine distances between comics c and c' are calculated as $s_{c,c'} = \frac{n_c \cdot n_{c'}}{||n_c|| ||n_{c'}||}$.

promotional effects are related to observable comic and episode characteristics. To inform such statements, we construct a set of variables, $\boldsymbol{X}_{cet}^{WFF.char}$ that account for observed heterogeneity in promotional effects across comics in terms of their genre and other key comic characteristics utilized for endogeneity corrections in the comic-level analyses.[16] To capture potential "goal gradients" (Lu et al. 2022), we further allow the size of the promotional effects to vary based on the focal episode's relative position within the comic series.

The final model specification, including all the aforementioned elements, is as follows:

$$\log(Y_{cet}^{F}+1) = \beta_0^F + 1\{episode\ e\ of\ comic\ c\ is\ subject\ to\ WFF\ on\ week\ t\}\cdot$$
$$\left\{\sum_{w=1}^{11}\beta_w^{comic,F}\cdot WFF_{ctw}^{comic} + \sum_{w'=1}^{11}\beta_{w'}^{episode,F}\cdot WFF_{cetw'}^{episode}\right\} +$$
$$\eta_{WFF}^{F} Sim.WFF_{cet} + \eta_{NonWFF}^{F} Sim.NonWFF_{cet} + \boldsymbol{X}_{cet}^{WFF.char}\boldsymbol{\beta}_{WFF.char}^{F} +$$
$$\boldsymbol{X}_{cet}\boldsymbol{\gamma}^{F} + \alpha_c^F + \alpha_t^F + \epsilon_{ct}^F \qquad \text{(Eqn. 3)}$$

2.6.2 Estimation and Results

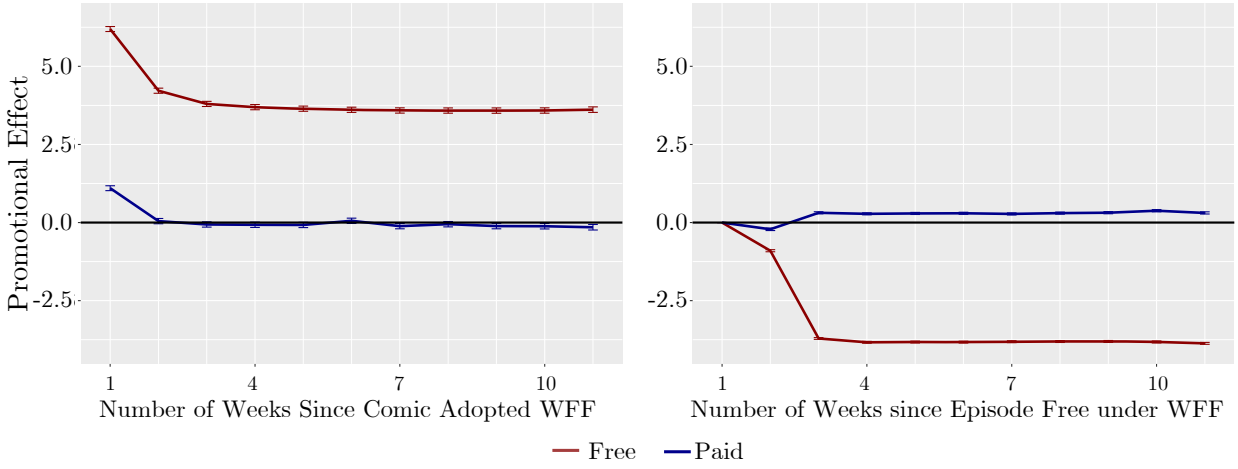
The robustness of WFF treatment effects estimates across a variety of endogeneity correction methods in our comic-level analyses strongly suggests that selection artifacts pose a limited threat to the unbiasedness of our parameter estimates. We nevertheless sustain the use of IPW with Logistic propensities to carry such robustness to our episode-level analyses; all estimation results reported in this section stem from IPW estimation (time trends for WFF effects with vs. without IPW corrections appear in Figure A4 of the Appendix).[17]

Comic-level vs. Episode-level WFF Effect Time Trends. Coefficient estimates for $\{\beta_w^{comic}\}$ and $\{\beta_{w'}^{episode}\}$ under the final model specification appear in the left and right panels in Figure 2.6, respectively, and in the last two columns of Table A13 of the Appendix. It is critical to note that

[16] We exclude the price variable due to insufficient variation in this quantity across WFF comics.
[17] Here, we use Maximum Likelihood, as opposed to REML, for model estimation purposes, as it lends itself to direct comparisons between linear mixed models with different fixed effects.

coefficients from this episode-level model should not be directly compared to those from the comic-level analyses, given that the units of analyses differ and the time trend of the promotional effect has been decomposed into two components. The "total promotional effect" – disregarding content heterogeneity and spillover effects – can be inferred from the sum of the two corresponding promotional effects in each panel. For example, for an episode in its second week of being available for free under WFF within a comic in its third week post-WFF-adoption, the promotional effect for free views on the log scale would be: $\hat{\beta}_3^{comic,F} + \hat{\beta}_2^{episode,F} = 3.797 + (-0.902) \approx 2.895$, or an overall viewership multiplier of $\exp(2.895) - 1 \approx 17.08$ on an interval scale.

Figure 2.6 Comic and Episode-Level Time Trends of Promotional Lift of WFF on Free and Paid Viewership



Delving deeper into the two dimensions of WFF's time-evolving promotional lift, one striking feature of the comic-level time trend (left panel) is that the promotional effect is far more pronounced for free than for paid views. This aligns with the model-free evidence of Figure 2.3, as does the fact that lifts for both free and paid views wane quickly after only few weeks. The *episode*-level counterpart (right panel) shows substantial divergence: while the number of free views suffers a steep decline once the episode has been available for free for more than two weeks, the lift for paid views sustains itself for many weeks post WFF availability, albeit with a smaller magnitude.

Spillover and Cannibalization. Our results imply moderate dampening of effects for both free and paid views on *other* WFF comics ($\hat{\eta}^{F}_{WFF} = -0.021, p < 0.001; \hat{\eta}^{P}_{WFF} = -0.016, p < 0.001$). Given that $Sim.WFF$ is a function of both the number of WFF comics and their similarity with other promoted comics, there are two ways to interpret these results. First, it is intuitively plausible that, once more comics enter the WFF portfolio, the promotional effect gets diluted, implying a diminishing return for each (Moon et al. 2007).[18] Second, the results also suggest that promoting comics very similar to one another may be detrimental: if only comics "adjacent in users' perceptions" adopted WFF, the promotion would appeal to just a small set of users, who might be easily attribute-satiated (McAlister 1982) from the set of WFF comics, leading to a smaller lift. This is also consistent with prior literature suggesting that portfolio or "bundle" variety tends to increase overall consumption (Ratner et al. 1999). In terms of practical platform management, this finding suggests that, in deploying WFF, there is a trade-off between promoting comics that might individually benefit most from WFF and maintaining suitable variety.

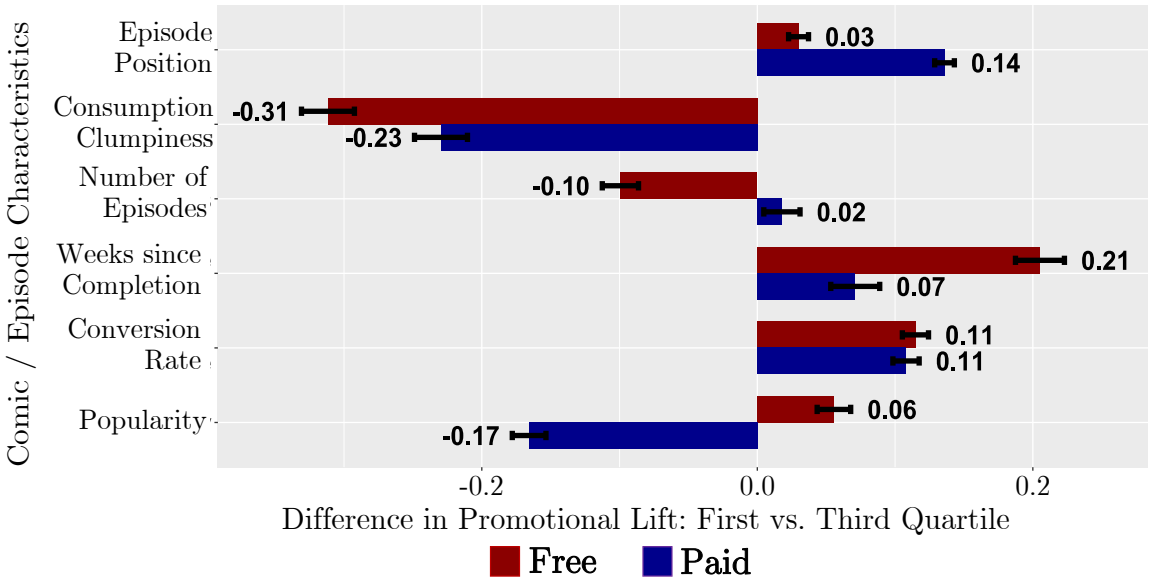
Beyond the interactions among promoted comics, WFF can also affect *non*-WFF comics. Model results suggest a nontrivial net positive spillover for free views and relatively modest net cannibalization for paid views. Given the median value of the similarity scores that enter the model ($\approx$ 0.926), results ($\hat{\eta}^{F}_{NonWFF} = 0.024,\ p < 0.001; \hat{\eta}^{P}_{NonWFF} = -0.008,\ p < 0.001$) suggest a median spillover of a 2.2% lift for free views, whereas estimated cannibalization corresponds to a 0.8% decrease in paid views. Whereas such trade-offs are strongly significant in a parametric sense, their practical substantive implications manifest themselves more clearly within the contexts of answering the key question for managers – *which specific units to promote (treat)*? – which will be addressed by the forthcoming combinatorial optimization.

[18] Although this is also consistent with the possibility of habituation (users becoming less sensitive to WFF over time) as opposed to dilution (the availability of additional WFF comics), supplemental wave-specific analyses suggest the former is not the sole contributory factor.

Observed Heterogeneity Across Comics. Table A14 of the Appendix illustrates promotional effects broken out across genres relative to a baseline of "Others." There is a notable, but not unreasonably large, degree of heterogeneity in the estimated promotional lift for comics belonging to different genres. For example, the "School" and "Mystery" genres appear to obtain the greatest boost in the number of paid views from WFF, whereas the "Fantasy" and "Romance" genres benefit less.

A variety of other pertinent comic and episode characteristics can also alter the magnitude of the promotional lift from WFF, which will not only be critical inputs for our optimizer but may also be generalizable to other platforms considering applying it to their content portfolio, and assist with choosing the optimal promotion set. As these variables operate on different scales, to enhance comparability we illustrate changes in the magnitude of WFF effects in log viewership when each characteristic increases from its first to its third empirical quartile, holding all other covariates constant, in Figure 2.7.

Figure 2.7 Promotional Effect, 3rd vs. 1st Quartile, for Comic and Episode Characteristics



Note: The bars represent the change in magnitude of WFF effects when each comic characteristic changes from the respective first to third empirical quantile while all else remains unchanged. The bold numbers in the plot represent statistically significant effects at the 5% level.

The *serial* nature of the focal content may impact the effect of the promotion in two distinct ways. First, considering the natural drop-off of consumption as users progress through episodes within a series (Lu et al., 2022), the effect of WFF may differ according to the *relative* position of each episode within the series; our results suggest greater impact for those near the end, potentially due to the fact that (mostly) daily engagement with each comic induced by WFF helps more users continue through the series without lapsing, compared to pre-WFF (Shu and Gneezy 2010). Second, the promotional impact of WFF can vary by consumption "clumpiness," or the average user's within-series progress in a single sitting. Model results suggest that WFF is less effective for comics that tend to be consumed in a concentrated manner for both free and paid views. This is in line with expectations, especially for the number of free views, given that the promotion by design works to "decelerate" free consumption.

Other operational characteristics of comics also influence the WFF effect. For example, comics with more episodes are estimated to experience a greater positive effect on the number of paid views due to WFF, while the opposite occurs for free views. Similarly, "number of weeks since completion" impacts promotional effect magnitude: comics that have had no new episodes uploaded for a longer time on average benefit more from WFF in terms of both the free and paid views. These findings might be anticipated given that the adoption of WFF may refresh users' interest in target comics by making them more noticeable, the effects of which would be more prominent for older comics than the new ones.

Estimates for the two variables capturing the degree to which consumers prefer the focal comic (i.e., conversion rate and popularity) show an intriguing pattern with respect to promotional effect size. As can be deduced from our previous discussion of the mechanism through which WFF operates – getting users hooked by offering free episodes and eventually converting them into

paying ones via consumption inertia – comics that easily get users "hooked" into paying for them (captured by a high conversion rate) benefit more from WFF both in terms of their free and paid views. While we are unable to directly incorporate the quality of each comic into the analyses due the lack of such data – the platform explicitly disallowed such forms of user feedback as ratings and comments – we posit that this measure serves as a reasonable proxy assuming that users would be more willing to pay for comics of higher quality. In that regard, our results show that managers would benefit from applying WFF to higher-quality comics than their lower-quality counterparts. Finally, the popularity-related coefficients indicate that more popular comics tend to experience a greater lift for free views but a smaller one for paid views. However, it is critical to note that this does not mean that popular comics will benefit less than those with lower popularity in the absolute number of paid views. Since our model is a log-linear one, this simply indicates that the model estimates that more popular comics will see a smaller lift in paid views percentage-wise, which could be substantially greater than the lift experienced by their less popular counterparts in absolute terms. This issue will figure empirically into the forthcoming combinatorial search for the "optimal set" of treated comics. These findings are particularly meaningful for managers, as popularity and conversion rates may be among the key determinants of content set composition for WFF.

2.6.3 Robustness Checks

In the Appendix, we present various alternative analyses to assess the substantive robustness of our key results. We demonstrate robust results with regards to separate (vs. joint) estimation of the free and paid viewership models (Figure A5 of the Appendix), implicit assumptions regarding the correlation of error terms between the two models (Figure A6), the number of comics included for analysis (Figure A7), and handling of outliers (Figure A8).

2.7 Optimization

2.7.1 Optimization Settings and Objective Function

Here we aim to identify the "optimal" set of comics to be subject to WFF for maximal platform-wide revenues[19] based on the episode-level model results. We again highlight that the only source of revenue for this platform is through the sales of its content (i.e., paid views of each episode), as it does not monetize its free views through advertising and does not offer an all-you-can-eat type subscription. This involves a delicate balancing act. At first blush, one might surmise that simply promoting those comics estimated to directly benefit most from WFF (i.e., with characteristics positively associated with the promotional lift) would be the platform's most judicious policy. However, this disregards a key dynamic at the platform level: the effect of assortment variety within the promotion set as well as potential cannibalization from the non-WFF comics. For example, there might well be a "small but cooperative" set of comics, i.e., one that experiences a modest direct lift from the WFF promotion but also cannibalizes less from others and among themselves, resulting in a greater overall platform-level viewership. As such, examining the significance of specific parameters is suggestive, but does not constitute a fully appropriate counterfactual.

To that end, we investigate whether the platform could have applied the promotion to a different set (of the same number, 12) of comics in the first wave to further improve its overall platform-wide revenue[20] over a 16-week time horizon, which corresponds to the length of window of post-WFF-adoption in our dataset. Note that each WFF comic would have a number of free

[19] We also perform an optimization of the expected number of paid views, the results of which are presented in Table A17 of the Appendix.

[20] Determining the optimal distribution of comics across multiple waves (i.e., which comics should be included in how many waves) is a very high-dimensional problem, one largely infeasible except for heuristic approximation, and one we leave to future, algorithmically-oriented research.

episodes at the beginning, a number of paid-only episodes at the end, and the wait time needed for users to unlock the next episode. Given little empirical variation in these breakpoints in our dataset, we apply dataset-guided rules for each comic, including the WFF ones, as laid out in Table A15 of the Appendix.

Based on model estimates in the previous section and these configurations, we determine the "best" combination of 12 comics out of the 571 candidates[21] that is expected to maximize platform-wide revenues over a 16-week time horizon *while keeping all else, such as comic characteristics and prices, equal*. To this end, we used the **`R`** function **`kofnGA`** in the **`kofnGA`** package (v1.3), which performs efficient combinatorial optimization using a genetic algorithm (GA; e.g., Raut et al. 2008). Specifically, the algorithm ran with 500 generations, each with a population size of 200. We use 10 different starting points, finding that the GAs largely converged to the same "optimal" solution with minimal differences in objective when they didn't.[22] Convergence plot for the GA optimization (Figure A9 in the Appendix) also confirms that the chosen settings (generations, population sizes, runs) were sufficient.

2.7.2 Optimization Results

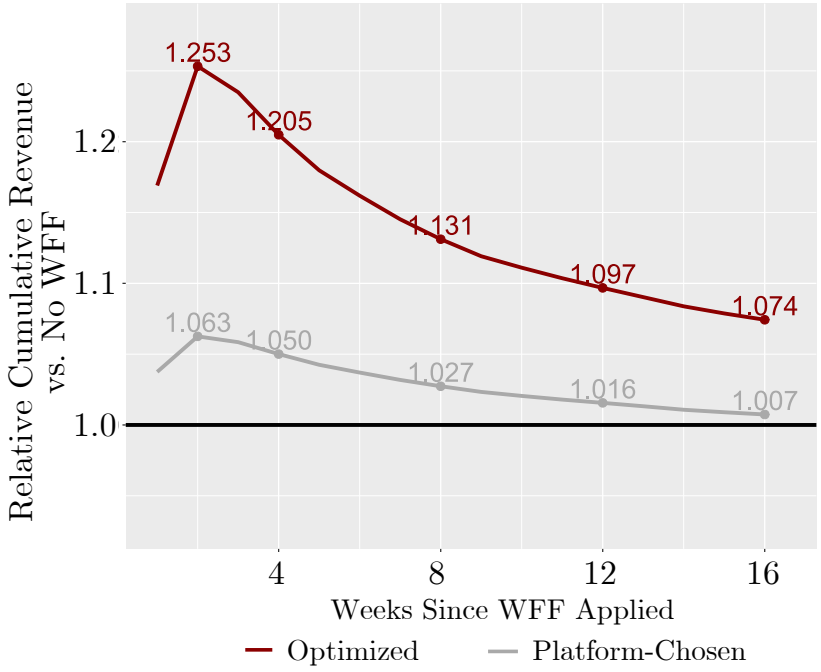
We calculate weekly predicted cumulative performance for the solution and compare it with two benchmarks: (1) those when we assume *only* the 12 actually-promoted comics were subject to WFF, and (2) those when we assume no comics were promoted (i.e., the platform never adopted WFF). Figure 2.8 depicts the evolution of the relative cumulative performance for the other two scenarios over time to the "no promotion" one (represented by the the horizontal baseline

[21] Only "completed" comics, defined as those with no new episodes published during the 14-week period after WFF was first introduced, were included due to the existence of paid-only episodes near the end of the series as a component of WFF.

[22] Each round of optimization took between 6-7 hours while using 20 server-capacity CPUs for parallelized computation of the objective function, requiring a total of around 65 hours for 10 starting points.

of 1). A comparison of the red and black lines highlights the time-varying nature of the magnitude of WFF's promotional lift: it is most pronounced during the first few weeks after adoption, consistent with the model-free evidence as well as comic and episode-level model coefficients. For example, on the second week, our optimized set of WFF comics is predicted to increase the cumulative revenue by nearly 25.3% compared to the no promotion scenario. However, after 16 weeks, it is estimated to have a lift of roughly 7.4%, substantially positive but smaller in magnitude compared to its peak. A comparison of the gray line to red and black ones indicates that while the platform-chosen set of comics[23] are also expected to generate more platform-wide revenue than under the no-WFF scenario, it does noticeably worse than our optimal set, where the gap between the two can be as wide as 17.9%.

Figure 2.8 Relative Cumulative Revenue of Optimized and Platform-Chosen Sets vs. No WFF



[23] We compare the characteristics of the promotion set found using optimization and the firm-enacted one in Table A16 of the Appendix.

2.8 Individual-Level Analyses

While the comic-level and episode-level models collectively afford insight into how different types of users are differentially affected by WFF and observed heterogeneity in WFF's promotional effects across comic types as well as the interactions among the comics at the platform level, it provides limited insight into how it alters *individual* users' consumption inertia typical for serialized digital content and their switching behavior across comics. For example, if WFF promotes comic viewership for one group of users while slowing it down for another, the episode-level model might see this as a blip, if at all. Site activity data affords a "street-level" view of how WFF alters individuals' usage trajectories, specifically, whether it actually decelerates consumption and what sorts of switching behavior between free and paid viewership it differentially generates.

The two defining aspects of WFF are "wait" and "free" and – as alluded to in the rationale for adopting two sets of time-varying promotional effects in the aggregate model – time is a critical element in how it operates. A more subtle advantage of modeling user-level behaviors (relative to aggregate platform-level activities weekly) is that the lack of temporal aggregation allows it to account not only for *what* users are doing, but also *when* they are doing it. Yet this advantage comes with a veritable explosion in "data points," because any target user views only a tiny fraction of the platform's available episodes, which we address below.

2.8.1 Model Specification

The Cox proportional hazards framework has been widely-applied in quantitative marketing to model the relationship between the time it takes for an event to occur and a set of predictor variables. Although a full account is beyond our scope, prominent examples include Bolton (1998), who studied the relationship between customer satisfaction and the duration of

the relationship between consumers and a continuous service provider and, more recently, Fradkin et al. (2021), who used the Cox framework to model the time it takes hosts and guests to leave a review after the other has reviewed them. Conditional on its not having occurred yet, an event's hazard rate diverges from its "baseline" depending on covariates in a multiplicative manner. A strength of the Cox framework is its semiparametric nature: the baseline hazard needn't be directly specified (parametrically), and right-censoring (i.e., events that do not occur during the observation window) is parsimoniously accommodated. As such, the model takes a tuple of two quantities as its outcome: (1) a binary indicator of whether the event has occurred during the observation window; and (2) if so, at what time.

We adopt similar Cox proportional hazards formulations for the free (*F*) and paid (*P*) views to investigate the effect of the WFF promotion on individuals' viewership propensity, or the "hazard" of each user (i) viewing each episode (e) within each comic (c) at time t in days:

$$\lambda_{icteF}(\mathrm{t}) = \lambda_{0F}(t) \cdot \exp\left(\alpha_{tF} + \alpha_{cF} + \boldsymbol{X}_{icetF}\boldsymbol{\beta_F}\right) \quad (2)$$

$$\alpha_{cF} \sim N(0, \sigma_{cF}^2),$$

where $\lambda_{0F}(t)$ is the (latent) baseline hazard. Along with week-specific intercepts (α_{tF}) and Gaussian-kernel regularized comic intercepts (α_{cF}), time-varying covariates ($\boldsymbol{X}_{icetF}$), taken in tandem, allow us to assess WFF's impact on users' momentum in continuing through the series and their tendency to switch between comics, specifically:

- Previous_Viewed_Free/Paid (PVF/PVP): To explore the degree to which users demonstrate inertia, or "stickiness," in their content consumption, we construct two dummy variables indicating whether the user's previous view of a *different* episode within the same comic was for free/paid. A positive coefficient indicates that users do exhibit some momentum in continuing through the comic after having seen an episode for free/paid, respectively.

- Promoted (P): To relate the propensity of content consumption to whether the episode was promoted, we include a dummy variable for whether the episode has become available for free under WFF by time t. A positive coefficient would indicate that WFF is indeed effective in boosting users' propensity to consume the promoted comic in the near future.
- Previous_Viewed_WFF (PVW): To investigate whether the momentum of continued content consumption is significantly different for the WFF vs. the non-WFF comics, we construct a dummy variable for whether the user has seen another promoted (at the time of the view) episode of the same comic before. A positive coefficient indicates that users show a stronger sign of momentum for WFF (vs. non-WFF) comics. We also interact this variable with Previous_Viewed_Free (PVW×PVF) to measure potential differences in the boosts in inertia/stickiness arising from a free vs. paid view of WFF comics.
- Similarity_Score_Non_WFF (SSN): The episode-level model allowed us to determine how comics being promoted affect other WFF or non-WFF comics; with the individual-level model, we can estimate the degree to which individual users tend to stick to comics similar to the ones *they* have seen before, as opposed to seeking variety. To this end, we calculate the sum of the dyadic similarity scores (those also used in the episode-level model) between the focal comic and all other non-WFF comics of which the user has seen at least one episode before. A positive coefficient implies that users tend to "stick with" comics similar to those they have viewed before. We also interact this with Promoted (SSN×P) in order to compare the degree to which users demonstrate similar tendency for WFF comics to the non-WFF ones. A negative coefficient for this interaction suggests users are more willing to try out promoted comics that are dissimilar to the ones they have seen before, in line with a trial effect being induced by WFF.

- Similarity_Score_WFF (SSW): We can also measure whether users who view promoted comics choose to view comics that are similar to them in the future. A positive coefficient suggests that users venture out to try out other comics similar to the ones they have been introduced to due to WFF, whereas a negative coefficient would suggest potential satiation.
- Num_Free/Paid_Views (NFV/NPV): To account for users' heterogeneous levels of engagement and their willingness to pay for content in general that may vary substantially over time within each user, we control for log(number of free/paid views) across all comics the previous week (as opposed to using time-invariant user-intercepts). A higher number of free views coupled with a low number of paid views would imply a generally low willingness to pay for the content on the platform at the time, and vice versa.
- More_Free_Content/Discount (MFC/D): To control for the impacts of each type of promotion on the platform other than WFF, we include dummy variables indicating whether each was enacted at that time for the episode.

2.8.2 Data Sampling and Reweighting

The sheer volume of individual-level data is noteworthy and prohibitive. Even with only the top 800 comics (those used in the episode-level model) and a modest sample size of 2,500 users (randomly selected from over 2 million), the enormous number of entries in the dataset would make even reasonably simple models virtually impossible to estimate. This is further complicated by the fact that the average user only viewed 133 episodes (out of 46,687) across 14.8 comics during our focal window of analysis, resulting in a preponderance of zero outcomes that provide any model with extremely sparse signals to inform parameter estimation.

Given the enormity and sparsity of the data, we retain information content while reducing volume through sampling and reweighting. Specifically, only episodes considered "free" or

"paid"[24] were included in the respective models. Out of 46,687 episodes included in our analysis, 7,993 (17.1%) were considered free. As for the promoted episodes that could be accessed by both paying for them and waiting to view them for free, all episodes beyond the first few in the beginning offered for free were included in the paid model, whereas they were included in the free model only if the user had viewed the previous episode after becoming available for free.

Likewise, we randomly under-sampled entries with zero outcomes (i.e., those where the user never viewed an episode), assigning weights reciprocal to the sampling probabilities, while keeping all positive outcomes intact. Specifically, for the paid model, 154,163 (0.16%) of zero outcomes were retained out of the 96,352,236, with a sampling weight of 625 (1/0.0016), whereas for the free model, 364,154 (1.6%) of them were out of the 22,759,660, with a weight of 62.5 (1/0.016). Simulation studies, summarized in Table A18 of the Appendix, affirm the validity of this approach for the survival analysis framework employed here.

2.8.3 Results

Table 2.6 summarizes estimation results for the individual-level Cox proportional hazards model for both free and paid views. Given that the coefficients relate each covariate to the "hazards," or propensity of each user viewing an episode conditional on that user not having previously done so, in a multiplicative manner, a positive coefficient suggests its variable "accelerates" consumption of the episode (with analogous interpretation for negative ones).

Before turning to the results, it is critical to note that implications stemming from the individual- and episode-level models might not perfectly coincide, for at least two reasons. First, given the differences in data granularity between the two models, some discrepancies in coefficient magnitude and even direction are to be expected. Second, and more centrally, the two

[24] An episode was considered "free" for any given week if at least 80% of its views were free that week. An episode was considered "free" in general if it was (considered) "free" on its median week.

models may differ, even substantively, because they account for different *outcomes*: the episode-level model considers only the number of users viewing the episode each week, whereas the individual model additionally accounts for "when" a user views an episode, over and above whether or not s/he does end up viewing it.

Table 2.6 Coefficients for Individual-Level Cox Proportional Hazards Model

	Free Model		**Paid Model**	
Coefficient	Estimate	Std. Error	Estimate	Std. Error
Previous_Viewed_Free (PVF)	2.532***	(0.082)	0.428	(0.890)
Previous_Viewed_Paid (PVP)	1.498***	(0.151)	4.607***	(0.941)
Promoted (P)	1.639***	(0.049)	2.490***	(0.722)
Previous_Viewed_WFF (PVW)	1.500***	(0.182)	7.908***	(0.609)
PVW×PVF	4.322***	(0.198)	-4.116***	(0.815)
Similarity_Score_Non_WFF (SSN)	0.261***	(0.001)	0.245***	(0.001)
SSN×P	-0.292***	(0.003)	-0.359***	(0.014)
Similarity_Score_WFF (SSW)	-1.043***	(0.035)	-0.657***	(0.032)
Num_Free_Views (NFV)	0.519***	(0.002)	-0.190***	(0.005)
Num_Paid_Views (NPV)	-0.073***	(0.003)	0.906***	(0.004)
More_Free_Content (MFC)	1.559***	(0.016)	0.832***	(0.015)
Discount (D)	0.405***	(0.022)	0.919***	(0.029)

The first two sets of coefficients (PFV and PFP) in Table 2.6 reveal the degree of inertia embedded in consumption of comics as a form of serialized digital content. Specifically, users who have viewed an episode of the same comic are more likely to continue viewing another episode in the near future than those who have not, with a previous free view somewhat more indicative of a future free view than a previous paid view ($\hat{\beta}_F^{PVF} = 2.53,\ p < 0.001$; $\hat{\beta}_F^{PVP} = 1.50, p < 0.001$). However, only a previous paid view seems indicative of an *accelerated* future paid view of another episode of the same comic ($\hat{\beta}_P^{PVF} = 0.43,\ p > 0.60$; $\hat{\beta}_P^{PVP} = 4.61, p < 0.001$). These results are consistent with a process model wherein users "just trying out" the first few episodes (e.g., for exploration, judging quality, etc.) continue to view other free episodes, but only rarely end up converting to paid viewers of a particular comic. On the other hand, users

already "invested" enough in a comic to have paid for an episode seem to maintain suitable momentum carrying them through the series (Zhang et al. 2021).

We next examine the various ways in which the WFF promotion affects users' consumption patterns. First, the positive and strongly significant coefficients for "Promoted" ($\hat{\beta}_F^P = 1.64,\ p < 0.001$; $\hat{\beta}_P^P = 2.49, p < 0.001$) suggest promoted comics are more prone to be consumed by a user even when that user has never seen any episodes of the same comic. Two potential factors contributing to this lift are the promoted comics being more noticeable on the platform, and that users may be more willing to try them out for the first time, knowing that WFF offers an opportunity to traverse through many of the comic's episodes without paying, corroborating our free trial account of the effectiveness of WFF in the episode-level model.

Previous_Viewed_WFF and its interaction with Previous_Viewed_Free both suggest that the promotion significantly boosts the inertia of comic consumption as well. Specifically, users who view episodes that have already been promoted accelerate (on average) their subsequent consumption of the same comic (vs. users who viewed an "ordinary" episode; $\hat{\beta}_F^{PVW} = 1.50$, $p < 0.001$; $\hat{\beta}_P^{PVW} = 7.91, p < 0.001$). Moreover, this acceleration effect is more pronounced for free views when the previous view of the promoted episode was also a free view ($\hat{\beta}_F^{PVW \times PVF} = 4.32,\ p < 0.001$), whereas the opposite was true for paid views ($\hat{\beta}_P^{PVW \times PVF} = -4.12,\ p < 0.001$).

These results are revealing for the platform, as they suggest that WFF, which on the surface encourages delayed content consumption, appears to be doing something close to the opposite in reality. One possible explanation for this phenomenon, especially for those taking advantage of the free access offered by WFF, is that the waiting time periods act as a perpetual "nudge" for users to return to the platform; that is, the ability to unlock subsequent episodes in a

structured manner leads to quick consumption, whereas a complete lack of time constraints might lead to indefinite procrastination (Shu and Gneezy 2010). The dramatic acceleration of consumption for users whose previous views of the same comic were a paid view of a WFF episode reflect these users' having become impatient or emotionally invested in the comic, considering their choice to pay for an episode even when given the option of viewing it for free had they been willing to wait.

One critical implication combining these results is that even users who have viewed the promoted episodes *for free* under WFF are estimated to have a significantly higher propensity to *pay* for another episode of the same comic ($\hat{\beta}_P^{PVF} + \hat{\beta}_P^{P} + \hat{\beta}_P^{PVW} + \hat{\beta}_P^{PVW \times PVF} = 0.43 + 2.49 + 7.91 - 4.12 = 6.71,\ p < 0.001$) compared to users who have never engaged with the comic. This result supports an account of WFF as free trials, where a substantial number of users who interact with a comic for free due to WFF may get hooked (to that comic) and end up converting to paying for it eventually; this in turn corroborates a key episode-level model result: that the promotion is counterintuitively effective in boosting not only free but also paid viewership of the promoted comics.

Next, we examine the general substitution or switching patterns of comics within each user for *all* comics, not only those related to the WFF promotion (as we did in our episode-level model). The positive coefficient estimates for Similarity_Score_Non_WFF in both models ($\hat{\beta}_F^{SSN} = 0.26,\ p < 0.001$; $\hat{\beta}_P^{SSN} = 0.25, p < 0.001$) are consistent with an anticipated phenomenon where users tend to stick to comics similar to those they have viewed previously. In the present case, these variables are measuring the tendency of users to stick to content that appears to be in their "comfort zone.". Interactions between Similarity_Score_Non_WFF and Promoted suggest that the promotion may indeed nudge users into trying out comics somewhat

different from their usual practice ($\hat{\beta}_F^{SSN} + \hat{\beta}_F^{SSN\times P} = -0.03,\ p < 0.001$; $\hat{\beta}_P^{SSN} + \hat{\beta}_P^{SSN\times P} = -0.11,\ p < 0.001$), indicating that WFF is being successfully utilized as a trial mechanism by end users.

However, at the same time, users appear to have a lower propensity to view comics similar to the promoted ones of which they have seen at least one episode ($\hat{\beta}_F^{SSW} = -1.04,\ p < 0.001$; $\hat{\beta}_P^{SSW} = -0.66,\ \mathrm{p} < 0.001$), indicating little to no positive spillover effect of users viewing the WFF comics, and possibly supporting the cannibalization indicated in the aggregate episode-level model. Combined with the previous negative values of $\hat{\beta}_F^{SSN}$ and $\hat{\beta}_P^{SSN}$, these estimates might also suggest that users who have tried out WFF comics that diverge from their usual repertoire simply revert back to those similar to those they've consumed before.

We observe substantial observed heterogeneity in users' comic consumption propensity according to their recent engagement with free and paid content. Coefficient estimates for Num_Free_Views ($\hat{\beta}_F^{NFV} = 0.52,\ p < 0.001$; $\hat{\beta}_P^{NFV} = -0.19, p < 0.001$) suggest that users who have viewed more free episodes in the past week have a higher propensity to view another for free, but are less prone to view another paid episode. In contrast, users who have more paid views recorded in the previous week are more likely to pay for another episode, but are less likely to view another free episode ($\hat{\beta}_F^{NPV} = -0.07,\ p < 0.001$; $\hat{\beta}_P^{NPV} = 0.91, p < 0.001$). These results, while intuitive, seem to suggest that users' behaviors may not be fully captured by their *overall* platform activity level alone; rather, there are some casual users who would very rarely pay for anything while mostly limiting their consumption to the free trial episodes, alongside some heavy users who pay for most of their content consumption.

Finally, coefficients for More_Free_Content and Discount suggest that the two types of promotions other than WFF are, all else equal, effective overall: both boost the propensity of

users' free ($\hat{\beta}_F^{MFC} = 1.56,\ p < 0.001$; $\hat{\beta}_F^{D} = 0.41, p < 0.001$) and paid views ($\hat{\beta}_P^{MFC} = 0.83,$ $p < 0.001$; $\hat{\beta}_P^{D} = 0.92, p < 0.001$). Note that this partly diverges from findings of the aggregate model, which suggested that discounts were *not* effective in boosting number of free views; this likely owes, as mentioned earlier, to distinct data granularity and outcome variables in the two models. All other coefficient estimates align directionally and in significance across the two formulations.

2.9 Discussion

In this paper, we examine the impact of a novel promotion, "Wait For Free" (WFF), applied to serialized digital content that is monetized through episode-level paywalls, using models that vary by aggregational granularity. Specifically, we first estimate the direct impact of WFF on the treated comics via a comic-level Difference-in-Differences specification using a wide range of endogeneity corrections. We also examine managerially relevant aspects of WFF's promotional lift, such as heterogeneity across different groups of users and its time trend. Then, in order to assist in the platform's ultimate goal of optimizing its WFF promotion – here, the set of comics to apply it to – we take a more granular approach of modeling episode-level viewership while dissecting the time trend of promotional lift at the comic- and episode-level and incorporating WFF's spillover effects (on both other WFF and non-WFF comics), as well as observed heterogeneity of WFF's lift across different types of comics. We then perform a combinatorial optimization informed by these results to determine the set of WFF comics that maximizes platform-wide revenue. Results indicate that, despite its seemingly consumption deceleration-incenting nature and overall cannibalization potential, WFF applied to our optimized promotion set leads to a substantial boost in platform-wide revenues. The aggregate

models are deepened via recourse to individual-level site usage trajectories, which are analyzed in a survival (Cox proportional hazards) framework that assesses how viewing behavior and stickiness, both within and across comics, are impacted by WFF. Individual-level model results strongly corroborate the key aggregate-level finding: WFF "paradoxically" accelerates, not decelerates, users' consumption of future episodes within a comic; and, moreover, it operates as a sort of free sampling device for comics (potentially) dissimilar to the ones each user has viewed previously, "stochastically converting" free users into paying ones in some cases.

Our analyses have several distinct lines of implications for marketing scholars and practitioners. Perhaps foremost among these is, to the best of our knowledge, introducing to the marketing literature a novel, yet increasingly widely-adopted, type of promotion for firms monetizing serialized digital content through episode-level paywalls, as well as a series of complementary analyses of its efficacy. While this uptick in adoption suggests practitioners find WFF profitable, there has thus far been no guidance regarding its potential or ability to enhance it, nor an attempt to scrutinize how individual users' behaviors are impacted by it. We stress here that our analyses do not suggest WFF is readily applicable to all platforms serving digital content; rather, it is the *serial* nature of content being monetized – for example, a tight storyline woven through multiple episodes – that seems critical to sustaining inertia in consuming additional content, and episode-level paywalls where increased content consumption translates to a lift in revenues.

Our analyses extend existing research on digital media promotion – for example, how much free content (if any at all) should be provided to users – by directly addressing *which* specific content types should be subject to WFF. Specifically, several actionable content characteristics emerge in terms of superior response to WFF (e.g., the popularity of the series and

conversion rates), providing generalizable practical insights for managers intending to implement similar type of promotion. In line with extant literature, we reaffirm the importance of accounting for the complex dynamic interplay in consumption of content with varying "dyadic behavioral distance," both across the promoted and unpromoted but and among the promotion set itself. These dyadic distances between series were operationalized via a straightforward yet theory-grounded similarity measure using individual-level consumption behavior, one that can be readily adopted on similar platforms hosting digital content.

While our research offers several lines of insight to serialized digital content platforms considering implementing WFF, some limitations must be acknowledged. A primary one involves the curse of dimensionality: due to sheer combinatorial complexity, our optimization necessarily focuses on discovering the set of series to be included in a single wave of promotion rather the number and composition of a sequence of promotion waves over time; because this is a purely algorithmic issue, we relegate it to future research, while acknowledging that approximate or heuristic methods are likely required. Another is the inability to capture heterogeneous promotional effects for comics with different per-episode prices, a potentially key question for managers, owing to the lack of sufficient price variation in our data. Lastly, coping with the full complexity of *potential* individual-level behavior – which can encompass switching noncontiguously from any episode of any comic to any other, leading to a gargantuan state-space – coupled with potentially millions of customer records and massive missingness, awaits innovations in scalable estimation technology.

Despite these caveats, Wait For Free is, as we document empirically, an impactful promotional vehicle for serialized digital media. Because managers typically track consumers by segment and take account of spillovers – especially synergistic ones – they can deploy it

selectively and refine its use according to the user and content characteristics particular to specific digital marketing contexts.

CHAPTER 3

Comparing the Effectiveness of Retargeting and Acquisition Online Banner Ads: A Flexible Approach to Estimating Ad Stock

3.1 Abstract

One of the earliest and most extensive literatures in quantitative marketing concerns the measurement of ad effectiveness. Because ads do not "work" immediately, and often require multiple exposures, econometric approaches to measuring the cumulative impact of advertising typically rely on the concept of a (latent) "ad stock" or "goodwill." Owing to the influential work of Nerlove and Arrow (1962), it has often been assumed that the contribution of each ad to the ad stock decays exponentially over time at a constant rate, captured by a single parameter common across all users and ad types. Here we examine how two fundamentally distinct types of online advertising campaigns – for acquisition and retargeting – differentially affect online users' behaviors by proposing flexible semiparametric and nonparametric Bayesian approaches for regularizing over past ad weights. Specifically, we make use of an online panel of individual-level ad impression data for a French financial services firm and online activity data of internet users who were shown its banner ads to show that relaxing the restrictive "single parameter decay" assumption allows us to capture the differential impacts of acquisition and retargeting ads on website visits, both in terms of their initial impact on the day of exposure and their lingering effects over time. Our results also suggest that a constant decay rate over time may be an over-

regularization, entailing substantive artifacts dictated by the common Nerlove-Arrow assumption. We then demonstrate the potential of using Gaussian Process (GP) Priors for capturing the potentially heterogeneous decaying patterns of the impact of past marketing interventions over time through a simulation study.

Keywords: online banner ads, retargeting ads, ad stock, Nerlove-Arrow, Gaussian processes

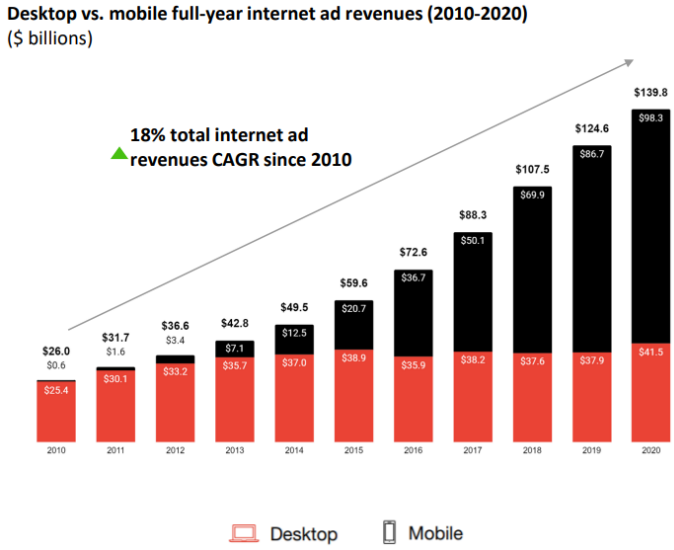
3.2 Introduction

As more and more of the global population gets connected to the internet, many firms' advertising efforts have shifted rapidly from traditional channels such as print, radio and television to the online ones such as Google search results, video ads on YouTube or social media such as TikTok. For example, according to the 2020 IAB/PwC Internet Ad Revenue Report, the online advertising business as a whole has witnessed a compounded annual growth rate (CAGR) of 18% from 2010 to 2020 to become one with $140 billion in revenue in 2020 (Interactive Advertising Bureau and PricewaterhouseCoopers, 2021). The rapid increase in popularity of online advertising has been attributed to many of its advantages over the traditional means of advertising. For example, online advertisements make it easier for marketers to evaluate the effectiveness of their various campaigns through the use of cookies, and to adjust budget allocations based on real-time feedback (e.g., Schwartz et al. 2017).

Despite the recent surge in the popularity of digital video advertising, banner advertising – where promotional messages or images related to the product or service being advertised is shown alongside the main content of a webpage – still maintains its position as one of the most common forms of online advertising, estimated to have accounted for 31.5% of digital advertising revenue in 2020 (Interactive Advertising Bureau and PricewaterhouseCoopers,

2021). Naturally, there have been extensive studies by marketing academics regarding whether they have economically significant effects on sales, and in which contexts they are more effective (e.g., Sherman and Deighton 2001, Chatterjee et al. 2003, Manchanda et al. 2006, Bruce et al. 2017). It is in this particular context, investigating effectiveness of online banner advertising, that our empirical analysis lies yet it attempts to specifically examine the differential impact of the two types of commonly used online banner ads: acquisition and retargeting ads.

Figure 3.1 Growth of Internet Ad Revenues from 2010 to 2020
(Interactive Advertising Bureau and PricewaterhouseCoopers 2021)

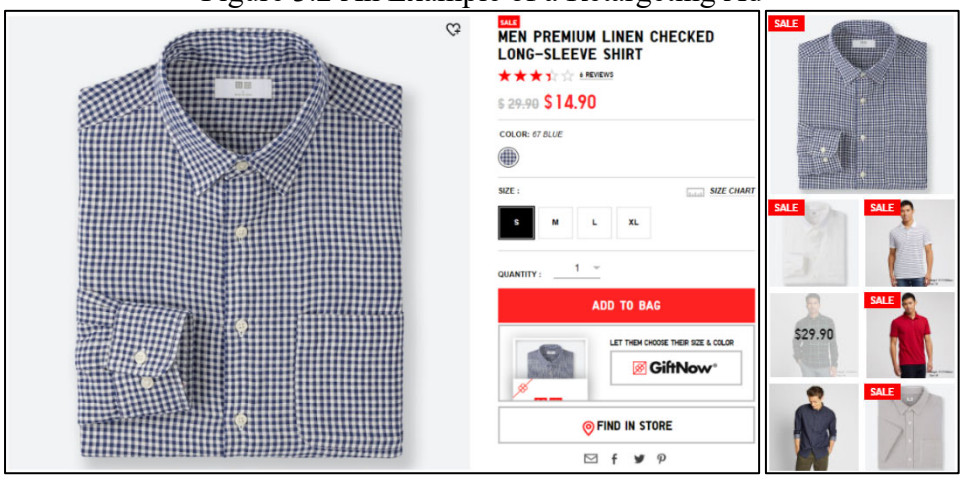


Acquisition ads are those that focus on informing potential customers of the general appeal of the products or services the firms provide in order to generate interest in the product or service being provided and acquire new customers. In the early days of advertising, they tended to be distributed without messages or content specifically geared towards a particular segment of audience or individual users, much in line with contemporaneous ads in traditional channels, like radio and television. Even in an online context, many of them used to be distributed rather indiscriminately to any users who visited a specific site without explicitly taking into account the demographic or behavioral characteristics of the users. However, as technologies to identify

individual users kept improving and online ads became extensively prevalent – making it more challenging to stand out from the myriad ads displayed and to grab internet users' attention – firms have been interested in improving the effectiveness of their ads and user responsiveness by optimizing the content of, time intervals between, and the sequencing of ads (Brown et al. 1998, Naik et al. 1998, Rafieian 2022).

Among the most common strategies currently being used by advertisers is retargeting: serving ads to users based on prior exposure or user-generated action. Retargeted ads often rely on internet cookies, which allow ad publishers to keep track of each user's internet activity over time and enable them to recognize specific users when they visit partner websites, which in turn allow them to display messages specifically customized to a target user's past browsing behaviors. As such, retargeted ads often include information relevant to items or products that the user has previously browsed. For example, if a user has visited a product page on a clothing manufacturer's website (left panel in Figure 3.2), then the manufacturer would display that product in one of its banner ads (right panel in Figure 3.2) alongside other products of its own. Similarly, retargeting mechanisms allow advertisers to include messages that may be particularly persuasive to consumers who already have developed some interest in the firms' products, ranging from a simple reminder to complete an unfinished transaction to more detailed information about the product or its marketing, such as price discounts. The reasoning behind this approach is most likely that the marketer is already aware that the user is at least interested in one of its products, and by including the product the user is interested in and many similar ones, it can increase the likelihood of the user noticing the banner due to familiarity, which hopefully leads to a higher probability of click or a visit to the website, which could ultimately lead to conversion.

Figure 3.2 An Example of a Retargeting Ad



While such retargeting ads have remained very popular and its motivation appears logically appealing, there has been little work in rigorously demonstrating the effectiveness of retargeted ads compared to generic ones that are not retargeted, with a few exceptions such as Lambrecht and Tucker (2013) and Sahni et al. (2019), potentially due to the lack of data readily lending to such analyses or methodological limitations. One of the main goals of the current research is to fill this gap in the literature by developing a model that allows marketers to more flexibly gauge the relative effectiveness of retargeted ads compared to their acquisition counterparts by allowing for differential initial and lasting impacts of each. In doing so, we adopt a Bayesian nonparametric approach using Gaussian Process (GP) priors to estimate the latent ad stock variable for each individual customer, allowing researchers to better account for the impact of previous ad exposures of each type. We hope to, by doing so, be able to eventually offer a prescriptive recommendation on optimizing the firm's advertising campaigns in terms of the types of ads (acquisition vs. retargeting; or which ad creative) and the right timing according to the objective function or marketing objectives of each firm.

The structure of this essay is as follows: we review the current literature studying the efficacy of online advertising including retargeting ads, traditional approaches for

operationalizing the effectiveness marketing efforts that persist over time using Ad Stock and the Nerlove-Arrow framework, and the use of nonparametric function estimation via Gaussian Process priors in marketing. Next, we highlight the important aspects of the data used in the project, and then present results from our empirical analysis suggesting that the traditional Nerlove-Arrow measure may fail to capture substantively important elements in the dynamics of the (latent) impact of ad exposures. Finally, we conclude by summarizing our findings and laying out plans for further analysis.

3.3 Literature Review

3.3.1 Advertising Effectiveness: From Offline to Online and Retargeting

Being among firms' primary and most venerated stimuli, advertising has long been the topic of interest for many marketing researchers. There have been numerous attempts to better understand how consumers process and react to the messages contained in advertisements (Wright 1973, Danaher and Dagger 2013) and to verify whether the ads have any meaningful downstream consequences that justify spending by connecting ad expenditures to sales at both the individual consumer level (Manchanda et al. 2006) and at the aggregate level (Rao 1986). There have also been efforts to improve returns to such advertising expenditure as much as possible by optimizing the affective nature of ad content (Brown et al. 1998), time intervals between impressions (Naik et al. 1998), the sequencing of ads across time within each ad session for a user (Rafieian 2022), or the context in which they are presented (Kupor and Tormala 2015).

Recently, due to the rapid growth of online advertising, the main focus of marketing researchers with regards to advertising has shifted away from those in traditional channels such as radio and television towards those in the online ones (Sherman and Deighton 2001). One of

the key benefits of studying data from online ads is that individual-level data are more readily available (Lambrecht and Tucker 2013). Studies investigating traditional offline advertising channels at the individual level have often necessitated managing a panel of representative households, the data from which had to undergo a complicated merging process with ad exposures from multiple sources and channels that often turns out to be not only obtrusive, potentially leading to inaccurate or biased data collection, but also cost-ineffective for a large-scale study (Danaher and Dagger 2013). Such limitations of traditional ads have often hindered researchers from studying the returns to advertising expenditures in a more granular fashion, beyond the aggregate or market level (Rao 1986).

Online cookies, extensively used by modern online advertisers, have made it far easier and less intrusive to keep track of individual consumers' activities online, by removing the need for users to manually record their browsing history and any ad exposures they had along the way. Naturally, the most relevant information pertains to the ad impression history for the brand and the browsing history within the brand's website. Such information allows the analyst to infer the efficacy of advertising activities in real time, by comparing arrival rates to a target webpage across users with differing ad exposures. However, cookies also generate data regarding each user's clickstreams outside the advertising firm's websites, which help researchers characterize the overall web browsing behaviors of each individual in terms of their overall engagements online and their interests in the relevant product category, allowing researchers to include necessary controls in their analyses.

There have been numerous works that take advantage of such granularity of the data available and attempt individual-level or segment-level analysis and optimization, all of which would have been daunting in a traditional media environment. For example, Hoban and Bucklin

(2015) use internet display advertising data and consumers' clickstream data to find that the effectiveness of display advertising differs across customers in different stages in the purchase funnel. Braun and Moe (2013) utilize the flexibility of online ads by modeling the effectiveness of multiple ad creatives so that the advertising firm might be able to not only optimize ad timing but also the creatives shown. Relatedly, Schwartz, Bradlow, and Fader (2017) use a multi-armed bandit framework to balance the "learning and earning" regarding the most effective ad creative for each website. Chae et al. (2019) study whether high ad exposures may lead not only to a wearout of ad effectiveness but also a potentially negative response, using real-world exposure and site visit data.

Due to the heightened availability of individual-level arising from cookies, advertisers have demonstrated growing interest in targeted placements of ads for each individual, which have shown to be promising (e.g., Hoban and Bucklin 2015, Schwartz et al. 2017). One prominent form of targeted ad placement is retargeting ads that exclusively pursue consumers who have visited the website before, often customizing the ad content according to the pages they have visited on the advertising company's website (Goldfarb and Tucker 2011, Lambrecht and Tucker 2013). While there has been some traction in researching the effectiveness of retargeting ads that are becoming ever more prevalent in the online world (e.g., Sahni et al. 2019), there has been a dearth of work on a model-based comparison between the effectiveness of retargeting ads and acquisition ads or on studying how retargeting ads interact with the generic or acquisition ads shown to the same customer. For example, a practitioner might then ask whether the retargeting ads would be more effective to customers who have been exposed to many acquisition ads relative to those who have been shown only few of them. Similarly, if there is truly a wearout of ad effectiveness or weariness to excessive advertising (Chae et al., 2019),

and each consumer has different optimal levels of exposures to ad messages of each kind, the advertising publisher would most likely benefit from optimizing the type of ad to show to each customer and when, with ad exposures in the future in mind (Rafieian, 2022). In the current project, we aim to answer such questions related to the interplay between retargeted ads and acquisition ads so that firms will have access to a tool that could help in determining whether the retargeting campaigns should be deployed vis-a-vis acquisition ones and how frequently.

However, certain challenges have also been reported in measuring the effectiveness of online ads. One of the prominent issues often discussed is the so-called "activity bias" (Lewis et al. 2011). That is, while it might appear on the surface that consumers who have been exposed to more ads have a higher chance of visiting the advertising firm's website in the near future, which is often associated with more conversions in observational data, such positive correlation stems partly or largely from a third factor: the varying degree of online engagement across users and time. That is, it is not the ad exposures that are causing consumers to visit the websites more and buy the advertised product, but the increased activity online on that particular day, that is driving both the increased exposure to the ads and the site visits. Lewis et al. (2011) also demonstrate that while the overall activity level of individuals or activities in the past are often used as control variables to account for this fact, they are not particularly good predictors of online activity of that user on any particular day. As such, they call for field experiments in order to directly uncover the true causal effects of online advertising through random assignments of levels of advertising across users. In our analyses, while not directly engaging in such field experiments, we use direct proxies for the degree of online activity of the users at the daily level, such as the number of webpages of the ad publisher's 200+ partner websites visited by the user and their variety that day in an attempt to control for such bias.

Another empirical issue that often arises while studying the effectiveness of online banner ads is that their site visit rates are typically extremely low. For example, Sherman and Deighton (2001) found a clickthrough rate around 0.3%, which is typically even lower than those for direct mail advertising. In a more extreme case, Lambrecht and Tucker (2013) found it to be lower than 0.01%; users clicked on 1 ad out of more than 10,000. The site visit rate in our data is also relatively low. While not surprising and theoretically benign, such a low site visit rate could in practice threaten stability in estimating any model with the site visit as an outcome variable. Such extreme imbalance of outcomes often warrants a very large sample size to make sound inferences, which in turn inevitably leads to excessive computational time. This problem is further exacerbated by the fact that conversion, the ultimate desired outcome, is again conditional on already sparse visits to the website. One potential solution for this issue is assigning sampling weights based on the impression counts (Malthouse et al. 2007), which would lead to a sample with a larger number of site visit counts than that of the general population, reducing the computational burden necessary for reasonable inference. Ideally, we would not perform a weighted sampling based on the outcome variables such as the visits to the advertising firm's websites or conversions as doing so may bias our parameter estimates.

3.3.2 Ad Stock

Central to much work investigating advertising effectiveness, whether offline or online, is the concept of "Ad Stock." Since the early influential work of Nerlove and Arrow (1962), most studies in the ad effectiveness literature, even the most recent ones, have used the notion that while a single exposure to the ad may not directly lead to an immediate response (e.g., a visit to the store or a purchase from consumers), it does create "goodwill" towards the brand or the product being advertised that might affect customers' latent purchase intentions (Rutz and

Bucklin 2011, Braun and Moe 2013, Bruce et al. 2017, Chae et al. 2019). Such goodwill might lead to future actions from customers if properly retained and managed; in fact, in some cases, long-term effects of advertising are estimated to be greater than short-term ones (Tellis 2009). Naturally, given the unknown delay between the ad expenditures and a significant portion of their returns in sales that occur over time, it is extremely challenging to precisely attribute any changes in sales to any particular ad campaign in the past, posing a profound challenge in the measurement of ad effectiveness.

Another related complication regarding the ad stock variable that may be obvious, yet is central to any discussion of its use, is that it is latent. That is, there is no direct and unobtrusive way of testing whether the ad stock variable constructed truly reflects consumers' goodwill obtained from advertising. Naturally, researchers have imposed a few assumptions regarding how ad stocks behave. One of the central assumptions made about the ad stock is that it decays over time in the absence of further ad exposure. That is, without any interventions from a target firm, consumers tend to "forget" the messages the brand or product tries to convey through their ads as time goes by. What further contributes to such decay is the fact that consumers tend to be exposed to numerous competing marketing messages from many other sources, vying for the limited cognitive resources of consumers. One particularly common assumption often made without a full justification in order to construct the ad stock is that of geometric decay (Nerlove and Arrow 1962, Naik et al. 1998, Rutz and Bucklin 2011, Braun and Moe 2013, Chae et al, 2019), with very rare exceptions (Sahni 2015). That is, at each of the (often discrete) time periods, a certain proportion of the goodwill, or ad stock, from the previous period is retained while the ad exposures in the current period contribute to the current ad stock: $Ad_Stock_{it} = \delta \times Ad_Stock_{i,t-1} + Ad_Impression_{it}$, which is formally equivalent to $Ad_Stock_{it} =$

$\sum_{t'=0}^{t} \delta^{t-t'} Ad_Impression_{it'}$. This recursive evolution in effect means that the contribution of ads in each previous period to the current ad stock decays exponentially over time at a constant rate ($0 < \delta < 1$ is usually assumed).

While such an exponentially-decaying setup makes intuitive sense and allows researchers to conveniently construct the latent ad stock variable using ad impression data directly, there is a critical limitation to this approach. Foremost, it assumes that the ad stock variable at the current period can be expressed as a weighted sum of previous ad exposures, where the weights are a geometric sequence with the ratio between terms held constant over time. However, it is perfectly possible that the pattern of decay of the goodwill created from past ad impressions deviates from such stringent formulations of a geometric decay. For example, it might be that the effect of ad exposures stays relatively robust for a limited time and then drastically fades away over the following days or weeks. That is, the rate of decay can be time-varying within each individual. It is also possible that the decays occur mostly right after the ad impressions and very slowly after a certain amount of time. While not directly related to consumers' processing of advertisements, a similar hyperbolic pattern of discounting future utilities has been documented (Laibson 1997). The functional form of geometric decay cannot accommodate these patterns.

Moreover, most previous works have also implicitly assumed the same rate of decay for all individuals and for all content. This in practice would mean that every customer who was shown the same ad at the same time would possess the same goodwill toward the brand, a strong assumption that may not be borne out empirically, given the different responses to similar ad exposures and heterogeneous cognitive capacity and category involvement of different consumers (Wright, 1973). At the same time, it may be unwarranted to assume that the ad stock accumulated from the different types of ads are to be "forgotten" at the same rate, given that

some content may prove to have a more lasting impact on their goodwill. While whether the model can successfully recover such heterogeneity in the decay rate across users and types of content is an empirical question, it seems theoretically and practically pertinent to search for empirical evidence of such heterogeneity.

Restrictions imposed by the commonly used parametric approaches in constructing the latent ad stock variables naturally lead researchers to consider the possibility of utilizing nonparametric approaches as an alternative. That is, instead of allowing only the ad impressions determine the latent ad stock variables assuming a parametric relationship, we could estimate them nonparametrically utilizing information from both the ad impressions and the outcome variables, jointly with other parameters in the ad response model.

3.3.3 Nonparametric Estimation in Marketing

Nonparametric approaches to estimating latent quantities have been extensively applied in the field of marketing. Nonparametric estimation has an advantage over its parametric counterpart in that it does not impose a restrictive functional form to the relationship between any quantities that are often implicitly made, such as the exponential decay of the ad stock variable, or bell-shaped, symmetric distribution of price sensitivities of consumers with very thin tails implicitly assumed by a Gaussian heterogeneity distribution. While parametric assumptions can improve efficiency of the estimators when correctly specified, misspecification of such distributional or relational assumptions could potentially lead to biased or inefficient estimators (Shively et al. 2000, Li and Ansari, 2014). Nonparametric estimation techniques allow researchers to be more robust to these potentially problematic assumptions at the potential cost of being able to make precise statements relative to the “correct” parametric model, were one able to find it.

Nonparametric estimation techniques have often been coupled with Bayesian methodologies for a few reasons. Foremost, imposing a prior on the quantities to be estimated usually acts as a natural form of regularization; it prevents the estimates from getting very extreme based on only a few data points (Dew and Ansari 2018) which could severely hurt predictive performance. For example, when estimating a household's purchase probability nonparametrically by the simple shares in the historical panel data, a household that is never observed to purchase a particular brand would be estimated to have an exactly 0 probability. However, with a prior distribution, the estimate would end up being a small yet positive number, a much more reasonable and realistic estimate of their purchase propensity. Relatedly, a hierarchical structure on priors allows the pooling of information across estimation; when estimating quantities with a limited source of information, such as a time-specific constant with only a few observations, the hierarchical structure allows information used to estimate other similar quantities, such as outcomes from nearby time periods, to be propagated to estimating it. A Bayesian setting is also useful when seamlessly incorporating heterogeneity of customers or in our application, heterogeneity across ad creatives, as a Bayesian estimation framework can readily incorporate a heterogeneity distribution in it.

For example, Shively et al. (2000) estimate the relationship between user characteristics and the partworths in a conjoint study using nonparametric Bayesian methods. They show that a typical linear association between the two would have resulted in a mischaracterization of such relationship due to the existence of a nonlinear relationship in certain areas of the user characteristics, leading managers to make misinformed decisions regarding new product developments. Li and Ansari (2014) also find that imposing a parametric Gaussian structure on the error terms or the heterogeneity distribution can seriously bias the parameter estimates in

choice models. They argue for a semiparametric approach using a Dirichlet process mixture that allows for a discontinuous, multimodal, or asymmetric distributions of such latent quantities (Ansari and Mela 2003).

One nonparametric Bayesian estimation approach recently gaining traction among marketers is the Gaussian Process (GP) prior (Dew and Ansari 2018, Dew et al. 2020, Kim et al. 2021). Gaussian Process priors are often used to nonparametrically estimate functions by extending the typical normal prior setup. Instead of assuming a Gaussian distribution over a finite set of values (e.g., functional values at a pre-specified set of points), inferences utilizing Gaussian Process priors assume that the values of a function for any finite draws of points in the domain follow a multivariate Gaussian distribution with certain mean function and covariance structure. Gaussian Process priors readily lend themselves to situations where there are many time-varying quantities that are correlated without necessitating that the analyst explicitly impose a complex interrelationship among them. Dew and Ansari (2018) use such a parsimonious setup when estimating the latent time trend of sales by decomposing it into various layers. Dew et al. (2020) further generalizes the setting where they introduce individual-specific evolution of consumers' price sensitivities where the dispersions of one's coefficients from the mean trend are allowed to be time-varying, using a hierarchical structure where the mean functions of the individuals' time specific coefficients are themselves a Gaussian Process.

In this essay, along with a fully Bayesian parametric approach of which the estimation results are presented, we plan to adopt in the future a Bayesian semiparametric approach that incorporates GP priors while estimating the latent ad decay functions over time across multiple ad creatives. As has been shown in previous works utilizing GPs, we expect this approach to have many advantages compared to the current practice of enforcing a geometric decay of past

ad impressions. First, we can allow for a more flexible relationship between the historical ad impressions for different ad creatives and the goodwill each customer has towards the advertiser's brand and services as time progresses. The mean functions in the Gaussian Process play a much less pronounced role in determining the estimated values of the ad stocks compared to in a traditional approach, mostly freeing the ad stock formulations from any restrictions imposed by the parametric assumptions. At the same time, the covariance structure, which has more influence on the estimated ad decay functions, would allow us to correlate the ad decay function values across time; it is likely that an ad's contribution to consumers' ad stock is more similar at times nearby than those in times far apart. We believe such formulations can allow us to get the best of both worlds: an easily specified general relationship between ad impressions and ad stocks while retaining the flexibility such as heterogeneous evolution of ad stocks necessary to fit the data closely.

3.4 Data and Descriptive Analyses

3.4.1 Data Description

The datasets we make use of in this essay are provided by an online advertising service firm named 1000mercis group, operating on behalf of a financial services firm who chooses to remain anonymous ("target firm" hereafter), both of which are located in France. The target firm provides financial credits and loans to its customers and their goal in their advertising campaigns recorded in the data is to maximize the number of "quotes," defined by a customer approaching the target firm and making inquiries about its services by inputting their personal information, as opposed to finalizing their purchases of the services. As such, we refer to customers requesting

the quotes as “conversions” in this essay. Within Figure 3.3 are a couple of examples of the ads the target company used.

Figure 3.3 Examples of Online Banner Ads Used by the Target Firm



1000mercis, the online advertising firm, posts advertisements of varying contents and sizes on its partner websites’ pages on behalf of the target financial company. There are 2 distinct types of campaigns they run: acquisition and retargeting. Acquisition campaigns are, as the name suggests, aimed at generating interest from new customers, who have no prior recorded visits to its website, and thereby attracting them to visit its website for the first time. On the other hand, retargeting ads are directed towards users who have already signaled that they most likely possess a higher interest in the target firm’s services than those who have never done so. Specifically, they are distributed to users whose online cookies indicate either a previous visit to the target firm’s website but no conversions, or a previous conversion (due to the wide range of services the target firm provides, it is possible for a single customer to ask for a quote multiple times). Retargeting ads are placed on the partner websites of the advertising firm when the users identified as targets of such ads visit them. Due to the fact that these users have some form of prior interaction with the target firm, albeit at various stages of the purchase funnel, retargeting ads are often assumed to be more effective in inducing further visits to the websites and ultimately conversions compared to the acquisition ads (Sahni et al. 2019).

One of the most distinctive features of our data is that, according to 1000mercis, ad impressions and the creatives used for the acquisition campaigns were assigned *randomly*: it did not systematically select users who will be shown the acquisition ads and messages used in them based on the past browsing history of the users or their activities on the target firm's website, with specific evidence of lack of such selectivity demonstrated by Chae et al. (2019). (Nevertheless, we do observe that not all creatives are used for the entirety of our observation window; there exists some apparent nonrandomness in the creative assignments. In fact, the 2 most often used ones are used between May 21st to June 13th and between June 13th and July 8th. However, we have no a priori reason to believe that users who are exposed to ads at different times are systematically different in a way that will affect our future analyses.) Such randomness in the impressions of acquisition ads helps us greatly in breaking free from the commonly encountered endogeneity concerns when estimating the causal effects of online ads (Rossi, 2014).

With regards to the retargeting ads, it is indeed true that they are not distributed randomly across all users: they are only shown to a group of users who have displayed interest by visiting the target firm's websites. More specifically, the data provider explained that visiting specific pages on the target firm's website makes a user eligible for the retargeting campaign (not all users who visit those pages are necessarily retargeted). Unlike for the data pertaining to ad impressions of the acquisition ads, it seems that there could well be non-negligible endogeneity issues in those of retargeting ads: there is a specific rule the ad publisher used in selecting those who it sees fit for retargeting campaigns, unknown to us as researchers. Turning a blind eye to this feature of the data generating process may lead to biased estimates of the causal effects of retargeting ads, thereby distorting the target firm's optimal policies in its retargeting campaign.

While the most complete way of addressing this issue would be to directly model the ad retargeting process for each user jointly with their response to these ads (Rutz and Watson 2019), we opt for simpler approaches for now, mainly by incorporating heterogeneity in the user-level baseline visit propensities and including control variables that could potentially assuage these issues.

The 3 main datasets that we analyze each contain the necessary information regarding three important components of the users' customer journey: 1) Ad Impressions, 2) Visits, and 3) Conversions. We briefly touch upon some of the important components of each and then summary statistics.

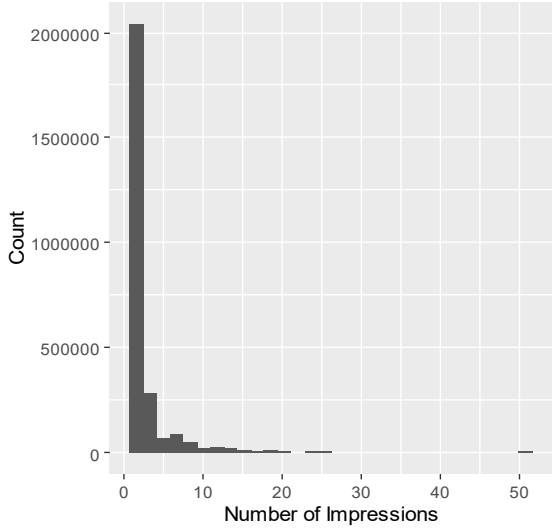
3.4.2 Ad Impressions

The ad impressions data contain information pertinent to each of the 7.7 million impressions of the online banner ads of the target firm posted by 1000mercis over a window running from May 21, 2013 to July 13, 2013. This dataset, derived from internet cookies, contains information such as the timestamps of the impression at the seconds level, the campaign ID (there are numerous campaigns within each of the Acquisition and Retargeting campaign groups) associated with each impression, ad creative ID (the type of ad content used; there were 7 mainly used for acquisition campaigns and 5 for retargeting campaigns), the web page domain the ad was posted on (there are 11,510 unique domains), the size of the ad and the ID of ad creative used, whether the ad was considered "visible" (defined by whether at least half of the ad was visible on the screen for at least 1 second), and the ID of the user to whom the ad was shown (there are about 2.7 million unique users in this dataset). Unfortunately, it does not have any information on competing ads that are shown to be important (Park and Hahn 1991; Villas-Boas 1993) and about the bidding process (and failed attempts) to show the target firm's ad over the

competing ads, which would have shed detailed insight on the underlying data generating process.

The impression dataset indicates that the average user was exposed to 2.86 ad impressions for the target firm. However, the distribution of the number of impressions is heavily skewed, as shown in Figure 3.4. At one extreme, there are a few users who have unrealistically large number of impressions. For example, 17 users were shown the target firm's ads at least 1,000 times. These "users" are most likely web crawlers that are irrelevant to the objective of the current study as they most likely do not respond to ad impressions. On the other hand, a majority (61.3%) of the users were exposed to the target firm's ad only once. Such a skewed distribution might make it challenging to make any meaningful inferences regarding the ad's impact on consumers' behaviors following the impressions, as most users would have ad stocks that are close to 0 during a substantial portion of the window of our analyses.

Figure 3.4 Histogram of Number of Ad Impression Across Users



About 8% of the impressions belong to the retargeting campaign while the remaining 92% of the impressions are those from an acquisition campaign. There might be several reasons for such a low rate of retargeting ads: first, the data are collected when retargeting techniques

using cookies were not as popular and sophisticated as at present. Moreover, not many people have visited the target firm's website prior to seeing the ads in the first place, which severely restricts the pool of users to which retargeting ads may be displayed.

3.4.3 Visits

Visits data contain the records of users' internet usage, or more specifically, their visits to one of 1000merci's 212 partner websites that include the target firm's website, from May 1, 2013 to July 31, 2013. Information in this dataset includes the timestamp of the visits, ID and category of the partner website, and the ID of the user that made the visits that allows us to connect this dataset with the ad impressions dataset. This dataset, again collected by online cookies, contains 38.8 million page visits by 817,001 users, who are a proper subset of the 2.7 million users in the ad impressions dataset.

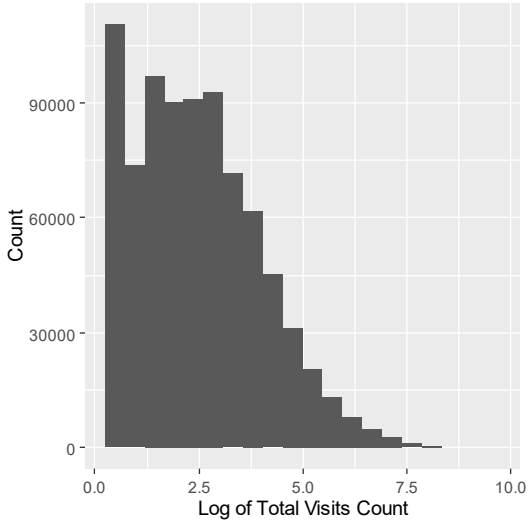
The visits dataset is important to our analyses in that records pertaining to visits to the target firm's website allow us to correlate ad exposures to each user with the intermediate outcome for the advertising campaign. Table 3.1 shows the frequency distribution of the number of visits to the target website across users. It is easily noticeable that the distribution is heavily skewed, with the vast majority of users never visiting the target firm's website, even after ad exposures. Another important measure of the success of online ad campaigns related to the number of visits is the clickthrough rate. We approximate it by calculating the proportion of days in which a customer visited it among those he/she were shown an ad of the target firm. Using this definition, the clickthrough rate is about 0.39%, which does not diverge drastically from previous reports of sparse clickthrough rates of online ads (Lambrecht and Tucker, 2013; Sherman and Deighton, 2001). To accommodate such sparsity, we opt for a count model of visits to the target firm's website that inflates the probability of zero counts as is typical when

analyzing similar dataset in our analyses below (Bruce et al., 2017). On the other side of the distribution, we see a somewhat long tail to the right; there is a very small proportion of users who visit the target firm's website more than 5 times. As these right tails might have meaningful information, we decide not to collapse the site visits into a simple binary variable of whether a user has ever visited the target firm's website.

Table 3.1 Distribution of Number of Visits to the Target Firm's Website Across Users

Number of Visits	Number of Users	Proportion (%)
0	2,646,890	98.76
1	11,348	0.42
2	4,031	0.15
3	2,680	0.10
4	1,888	0.07
5	1,611	0.06
>5	11,786	0.44

Figure 3.5 Histogram of Log Number of Recorded Online Activities Across Users



Another major use of the visits data is to control for the so-called "activity bias" that has been shown to lead to significant over- or under-estimation of the effectiveness of banner ads on prompting the web users to visit the target firm's website, and ultimately on their converting due to the spurious correlation caused by a user being active on a particular day driving up both the ad exposures and visits to the target firm's website (Lewis et al. 2011). As we can see in Figure

3.5, there is a substantial variation in the level of online activities across users, which increases the chance of such spurious correlation being picked up by the model. We can use representative sample of the internet activities in the visits dataset to control for their overall level of online activities each day, which would allow us to reasonably and robustly estimate the true effectiveness of banner ads without running a full-fledged field experiment.

3.4.4 Conversions

Finally, the conversions dataset includes the information regarding which users converted, or in other words, requested "quotes" for the target firm's financial services, and when. Due to the nature of the services it provides, it is not impossible for a user to ask for multiple quotes from the target firm. The dataset contains information regarding 53,482 conversions by 41,444 unique users during the timespan of between May 6, 2013 and July 31, 2013. As can be seen in Table 3.2, most users convert only once during our observation window. However, we have ad impression and internet usage data of only 4,055 users out of the 41,444 users who have ever converted, which leads to extremely sparse positive outcomes for the users with ad impression and visits information, which is a reason why we mainly focus on the first two datasets in our current analyses.

Table 3.2 Distribution of Number of Conversions Across Users

Number of Conversions	Number of Users	Proportion (%)
1	35,176	84.88
2	4,827	11.65
3	1,023	2.47
4	243	0.59
5	88	0.21
6	32	0.08
7	17	0.04
>7	38	0.09

3.5 Empirical Analyses Using a Bayesian Parametric Approach

In this section, we search for empirical evidence against the single parameter decay assumption. We specify a 'holistic' model for the relationship between ad impressions and visits to the target firm's website that parametrically nests the traditional geometrically decaying ad stock assumption. Specifically, we relate historic ad impressions for each user to the probability he/she visits the target firm's website at the daily level. We would expect that all else equal, more ad exposures to each user would tend to build more goodwill and result in a higher probability to visit the website, where the contribution of each ad exposure to the goodwill diminishing over time.

3.5.1 Model Specification

As mentioned in the literature review, the traditional formulation for ad stocks can be considered a weighted sum of previous ad exposures, where the weights are a geometric function of time elapsed since the exposures, with a constant decay rate. In other words, the log of the weights for previous ad exposures are a linear function of the time elapsed. The advantage of reformulating these weights is that we are now able to readily generalize the functional form of the weights and incorporate other important factors to affect the decay rate over time. For example, we may allow for different ad creatives to have their own decay rates as some ad content may be more memorable than others. In this case, more memorable and noticeable ads would have a slower decay rate whereas those that are less "sticky" in users' minds would have a faster decay rate. We could also incorporate user heterogeneity in decay rates to reflect the fact that some users may retain their goodwill towards a brand for a much longer time than others. The former group of users would be assumed to have a slower decay rate overall compared to the forgetful ones.

For example, we may specify a model where all the aforementioned features affect the decay rates:

$$\log(w_{ict}) = \log(c_0) + \sum_{d<t} \gamma_{icd},\ where\ \ \gamma_{icd} = \gamma_0 + \gamma_i + \gamma_c + \gamma_d, \quad (1)$$

The weights w_{ict} directly relate the previous ad exposure of ad creative c from t time periods ago for user i. It is mainly composed of two parts: 1) c_0 that accommodate the heterogeneous initial efficacies of different ad creatives, and 2) γ_{icd}, or the log of the individual-creative-time-specific decay rate, which is expected to be negative in most cases to reflect diminishing impact of previous ad exposures over time. When we exponentiate the first equation, it is trivial to see that the weights are a product of the initial impact parameter c_0 and all the exponents of subsequent decay rate parameters leading up to relevant time period, or $\exp(\gamma_{icd})$. γ_{icd} is assumed to take an additive form of 1) γ_0, the baseline log decay rate for all users and creative for all time periods, 2) γ_i, the user-heterogeneity parameter that dictates the user-specific decay rate that deviates from the overall mean where a positive value would imply a slower decay rate for that individual, 3) γ_c, creative-specific decay rate for which a positive value would be assigned for creatives that are more long-lasting in the average user's memory, and finally 4) γ_d that specify the deviation from the geometric decay where a positive value would indicate a slower decay than what the geometric decay would imply.

It is also relatively straightforward to notice that when we "turn off" most features of these specification, namely if we set $\gamma_i \equiv 0,\ \gamma_c \equiv 0$, and $\gamma_d \equiv 0$, the current model reduces to $\gamma_{icd} = \gamma_0$, and $w_{ict} = c_0 \cdot \exp(\gamma_0)^{t-1}$, which is exactly the traditional geometric decay function with common time-invariant decay rate across users and creatives, but with heterogeneous initial impact across various creatives. In this sense, the deviations of parameters γ_i, γ_c, and γ_d away

from 0 represent the degree to which the empirical data violates the constant-rate geometric decay assumption in their respective dimensions.

In the current model, due to the lack of sufficient information in the empirical data and realistic limitations in terms of the computation time to estimate the model, we assume a homogeneous decay rate across the users, or $\gamma_i \equiv 0$ for all i. Moreover, to bring the model further into the realm of reasonable estimation, we replace decay rate parameters specific to each of the creatives, γ_c, with γ_r which denotes the difference in log decay rates for all 5 ad creatives used for retargeting campaigns compared to the remaining 7 used for acquisitions. Under this specification, γ_0 would correspond to the log of the "baseline" decay rate for the acquisition ad creatives, whereas $\gamma_0 + \gamma_r$ is that for the retargeting creatives. We keep γ_{d}, parameters for time-varying deviations, as in the previous more general settings.

Given the empirical evidence provided from preliminary analyses, we further assume that for all creatives, the impact of ad creatives from more than 30 days are 0; that is, any goodwill created by any of the banner ads dissipates completely in this time period. Our results find this assumption to be reasonable. As for the γ_ds, or the parameters dictating time-specific deviances from the geometric decay, we allow for independent priors for the first 5, while imposing some regularization for the others, to accommodate for relative lack of information to inform each for the latter. For the initial impact of each creative, c_c, we assign a Beta prior ($Beta(6,6)$ to be specific; we have tested other values and the results are robust to such changes) with mean 1, so that the values of c_c represent the relative impact of each creative compared to the overall average *a priori*. Our ad stock formulation for all users at any time adheres to the basic principle of weighted average of previous ad exposures, where the weights depend on the type of ad creative (c), and the time elapsed since each exposure ($t - t'$). Then, we relate the constructed ad

stock to the final outcome variable, which we define as whether each user has visited the target firm's website on each day. We use a logistic regression model that incorporates heterogeneity in the baseline visit propensity (μ_i) and in ad responsiveness ($\beta_{Ad,i}$), both with a Gaussian mixing distribution, where the log-transformed ad stock variable is included as one of the covariates. Incorporating heterogeneity in the baseline visit propensity (μ_i) seems especially pertinent for the current analysis as there may be substantial differences in users' prospects of reacting to the ads between users that are the targets of acquisition ads who have never shown interests in the target firm's products and users who are being retargeted, that are not completely captured by some of the control variables that we list below. Not fully resolving these issues may bias some of the parameter estimates that dictate the differential initial impacts and decay rates for retargeting ads compared to the acquisition ones.

$$\gamma_{cd} = \gamma_0 + \gamma_r + \gamma_d \tag{2}$$

$$\gamma_{d=1}, \cdots, \gamma_{d=5} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1^2),\ \gamma_{d=6}, \cdots, \gamma_{d=30} \sim \mathrm{N}\left(0, \sigma_\gamma^2\right) \tag{3}$$

$$\log(w_{ct}) = \begin{cases} \log(c_c) + \sum_{d \le t} \gamma_{cd}, & \mathrm{t} \le 30 \\ -\infty, & \mathrm{t} > 30 \end{cases},\ c_c \sim Beta(6,6) \tag{4}$$

$$AdStock_{it} = \sum_{t' \le t} \sum_{c=1}^{C} w_{c,(t-t')} \times c_c \times AdExposure_{ict'} \tag{5}$$

$$P(Y_{it} = 1) = inv - logit(\mu_i + \beta_{Ad,i} \times \log(1 + AdStock_{it}) + \boldsymbol{x}_{it}'\boldsymbol{\beta}) \tag{6}$$

$$\mu_i \sim N\left(\bar{\mu}, \sigma_\mu^2\right), \quad \beta_{Ad,i} \sim N(\bar{\beta}_{Ad}, \sigma_{Ad}^2) \tag{7}$$

The covariates that enter the linear predictor as control variables in $\boldsymbol{x}_{it}$ are:

- Number of webpages visited and the Number of unique websites visited: As we detailed in our literature review, one of the spurious correlations that are often captured by the model when studying ad effectiveness using observational data is a tendency where users who are more active on the internet at certain points of time naturally having a

higher propensity to both be exposed to ads and the website of firms being promoted in the ads than at other times. Positive correlation in these instances may not always originate from a direct causal effect of the ad exposures leading to more frequent visits to the target firm's website. As such, we control for this by constructing two variables using our Visits dataset. For each user, we construct the logs of the number of webpages visited and the number of unique websites visited each day, both *excluding* the target firm's website. These are intended to capture two dimensions of users' online activity: amount and breadth.

- Week-lags of Number of webpages visited and the Number of unique websites visited: Preliminary analyses have provided some evidence of potential cyclicality of user activities. To correct for it, we also include the week (7 day) lags of each of the two variables mentioned above.
- Have Visited, Cumulative Visits, Cumulative Days of Visits: We expect users who have already visited the target firm's website to exhibit different levels of propensity to visit the target firm's website after having done so. On the one hand, users who have visited the website would arguably be more interested in the target firm's products, signaling an overall tendency for such users to exhibit higher likelihood to return later. On the other hand, one may also argue that users who have visited it many times already may not achieve much in terms of informational value in their future ones compared to those who has never visited it. To control for these factors that might be of empirical interest, we include three variables: 1) an indicator variable whether has already visited the target firm's website at least once, 2) the number of cumulative visits to the target firm's website, and 3) the number of unique days on which they have visited it.

- Recency of Visit, week- and 2-week-lags of visits: We also expect there may be some information the timing of users' most recent visit to the target firm's website might contain with regards to their future visits. As such, we construct the log of the number of days since that user's most recent one. In order to account for the cyclicality of the visit to the target firm's website, we also include indicator variables of whether the user has visited the target firm's website a week, or two weeks ago, respectively.
- Whether first ad impression was retargeting, Number of unique retargeting ad creatives: Since the data we use for analyses are left-censored, we may need to account for important user-firm interactions that took place before the starting date of our data window. For one, we include a dummy variable of whether the first ad impression we observe in our data window is a retargeting ad. A value of 1 for this variable signals that these users have already visited the target firm's website, indicating a deviation in propensity to visit it in the future. Furthermore, in order to control for the intensity to which each user is being retargeted, we also control for the number of unique retargeting ad creatives shown to each user up to that day.
- Day of week effects: In order to control for another obvious source of activity level cyclicality, we also include day of week fixed effects where the coefficient for the first day of week is fixed at 0 for identification.

3.5.2 Data Construction and Estimation

The data window for the visits dataset (May 1st to July 31st) encompasses that for the ad impressions dataset (May 21st to July 13th). Since our model requires information from both datasets, we limit our window for analyses to the latter. Furthermore, we would like to partially account for the "cold start" problem (Padilla and Ascarza 2021) where all users are assumed to

start with zero ad stocks due to the left-truncation of data. As such, we first construct the variables in the dataset (e.g., previous ad exposures of each creative for each user, or the lags of online activity levels) as if we would be using the full dataset, and then truncate the first 7 days of the resulting dataset to arrive at our final dataset for analysis.

Due to the sheer magnitude of the dataset, it is simply impractical to include all users in analysis. As such, we limit our analysis to 12,000 users sampled from more than 800,000 users with both ad impressions and visits data. We first exclude users who are mostly not representative of actual users: users with 25 or more impressions (24 is the 99% quantile) or 100 or more visits to the target firm's website (there were only 156 such users). We surmise these may be bots that are crawling the internet for various purposes such as informing the search engines, which are unlikely to be of managerial interests of the target firm.

Due to the extreme sparsity of ad impressions, as can be encapsulated by the fact that 61.8% being exposed to a single ad throughout the entirety of the data window, we use a sampling mechanism that favors users with more frequent ad impressions. By including more of such users while not completely discarding users with very low impression counts at the same time, we expect to be able to more efficiently estimate the model parameters without unnecessarily exploding the size of our dataset. More specifically, we used the sampling weights $w_i \propto (n_i + 0.25)^2$ where n_i denotes the number of unique days on which the user had at least one ad impression. It is important to note that we are *not* adjusting the sampling weights based on the outcome variable, or the number of visits to the target website, which might bias the parameter estimates due to selection issues. Rather, we are sampling based on a function of one of the predictor variables, which steer clear of such concerns.

For estimation, we use the Stan programming language (Carpenter et al. 2017) that allows one to readily make posterior draws with low autocorrelations using a No U-Turn Sampling (NUTS) mechanism, without the need to explicitly derive the posterior distributions, as is typical in Metropolis-Hastings algorithms under non-conjugate setups. Due to the vastness of the dataset and the heavy computation that goes into calculating the ad stock variables for individual users on each day, we limit ourselves to 800 posterior draws from each of the 6 chains, where the first 400 would be discarded as they arise from a warm-up phase in which Hamiltonian Monte Carlo parameters such as the step size and parameter covariance matrix undergo calibration.

3.5.3 Results

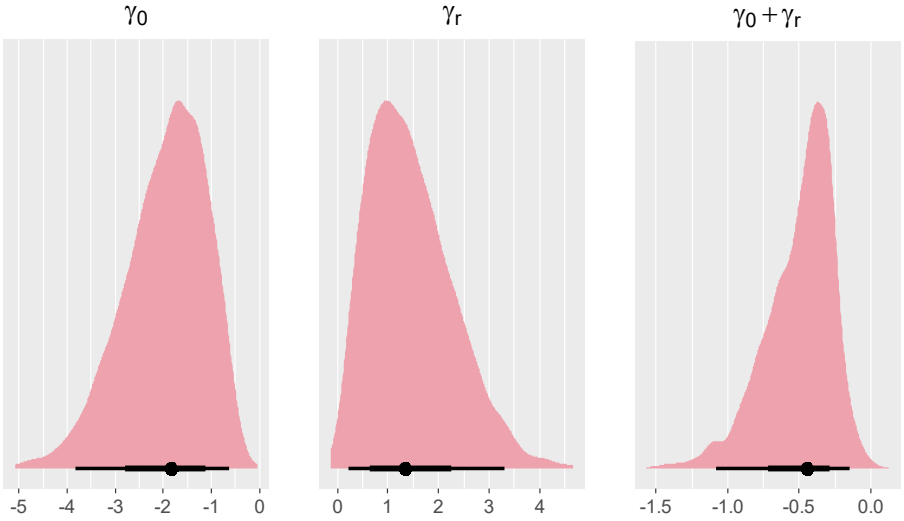
The time (in seconds) it took for each of the 6 MCMC chains to complete sampling is in Table B1 in the Appendix. Based on the trace plots of the key parameters of the model presented in the figure below, it seems all 6 chains have converged to the stationary posterior distributions, and hence all will be used to make inferences about the parameters.

As a measure of convergence and sampling quality, we look into the trace plots alongside the effective sample size (ESS) and Rhat for some of the key parameters of our model. Effective sample size is a measure of the quality of the posterior draws as defined by $\text{ESS} = \frac{n}{1+2\sum_{k=1}^{\infty}\rho(k)}$ where n is the number of posterior samples, and $\rho(k)$ is the autocorrelation of the samples at lag k. The lower the autocorrelation between the posterior draws, the higher the effective sample size, or higher number of independent sample equivalents. Rhat is a measure of the convergence of the chains. It compares the average within-chain variance of samples to the variance of pooled samples. Rhat values of less than 1.1 are often used as criterion for convergence. As our table of Rhats and the trace plots for key parameters of the model in Table B2 in the Appendix indicate,

we see no strong indication that any of the chains have failed to converge, and hence will be using posterior samples from all 6 chains for analyses.

We first report the posterior distributions of the decay-related parameters in Figure 3.6. For each parameter, we will be presenting the empirical distribution of the pooled posterior samples across all 6 chains, along with a bar at the bottom whose thinner and thicker lines represent the range of 2.5% and 97.5% empirical quantiles and 33% and 66% empirical quantiles, respectively. A dot in the middle of the bar represents the sample median. First, the parameter γ_0 dictates the baseline decay rate for the acquisition creatives. As expected, they are estimated to be negative, with the empirical 95% credible interval excluding 0, implying that the contribution of the acquisition ads towards users' goodwill decreases as the time since ad exposure increases. More specifically, the posterior median is around -1.84, implying that not accounting for the time-varying decay rates, acquisition ads lose approximately 84.2% ($\exp(-1.84) - 1 \approx -0.842$) of its impact every day, suggesting a very minimal impact after a few days.

Figure 3.6 Posterior Distributions of Decay Rate Parameters



Next, we examine the posterior distribution of the parameter γ_r, which captures the divergence in decay rate between the retargeting and acquisition creatives. As the empirical 95% posterior quantile for γ_r that excludes 0 in the second panel of the plot below suggests, there does seem to be statistically significant heterogeneity in decay rates according to the content of the ads. Specifically, a positive estimate for it suggests that the retargeting ads retain their impact on users' latent goodwill for a longer period of time compared to their acquisition counterparts. This result does seem to coincide with an *a priori* expectation that suggests these ads are likely to be more memorable and hence "stick" with users more. While not maximally granular in that the analogous parameter is not separately estimated for each creative, that this parameter is statistically different from 0 also implies that the commonly used assumption of a common decay rate parameter across all ads fails to uncover an important pattern in terms of how ad exposures of various creatives differentially affect users' propensity to engage in certain behaviors in the future. As we further detail later while discussing the resulting decay curves as a function of the model parameters, these differences, along with the other parameters might have important practical implications regarding how sustained any impact of ad exposures are.

As we discussed in the model specification, the retargeting ad creatives have a "baseline" decay rate of $\gamma_0 + \gamma_r$, when not accounting for γ_ds or the deviations from a geometric decay assumption. These are again estimated to be negative as one would expect, with 0 lying to the right of the right tail of the empirical 95% posterior credible interval. The posterior median is approximately -0.45, implying a 36.1% ($\exp(-0.45) - 1 \approx -0.361$) daily loss of retargeting ad's impact. While this may appear to be a substantial loss, it is still drastically smaller – in fact less than half – than that of the acquisition ads, again highlighting the importance of accounting for the heterogeneity in decay rates across different types of ads, which may have important

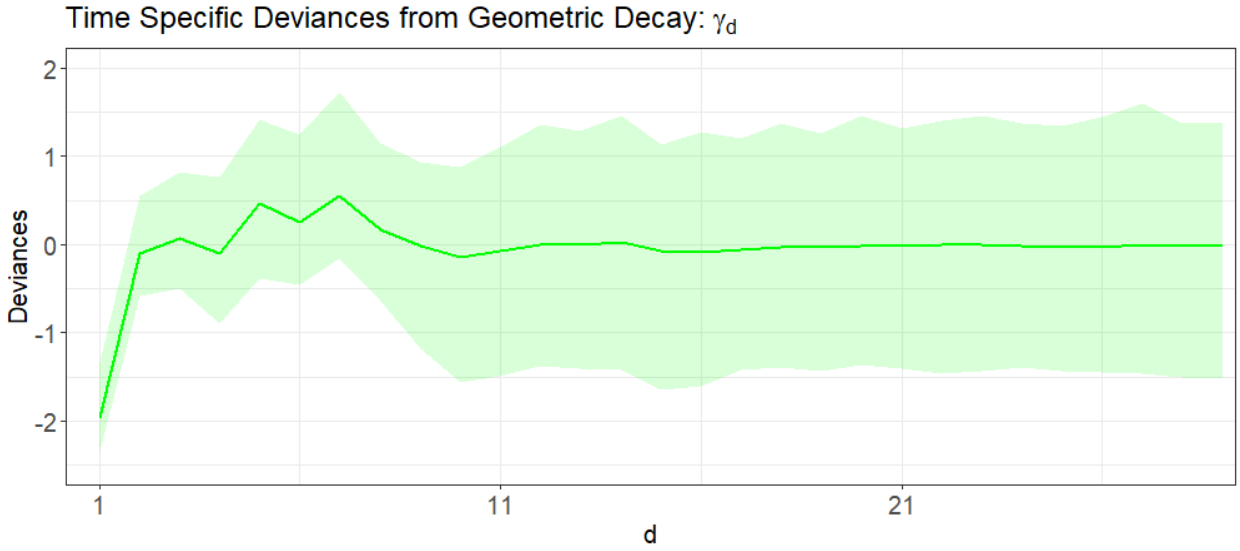
practical implications for managers regarding the intervals at which different types of ads are placed, for example.

We now turn to estimates of another set of variables, $\{\gamma_d\}$, that reveal the degree to which the pattern of decay deviates from the traditional geometric decay. As mentioned during our model specification, $\gamma_d \equiv 0$ would indicate that the traditional geometric decay assumption is indeed what empirical data support. Any deviations from it would manifest as one or more parameters being significantly different from 0. Figure 3.7 depicts the median values of the empirical posterior distribution for γ_{d} at each d, along with a band that represents the empirical 95% credible intervals. We see strong evidence for $\gamma_{d=1} < 0$, with the posterior median around -1.96 and 0 located well outside the credible interval. This implies that after the first day of ad exposure, its impact on users dissipates much faster than what the typical geometric decay would imply, revealing another way in which empirical evidence suggests a violation of the traditional geometric decay assumption with constant decay rates across types of ad contents and over time. Practically, such dramatic decay on the first day of ad exposure may be reflecting the unique nature of online banner ads: unlike ads on traditional media such as the TV, radio, or newspapers, where there tends to be a significant delay in users' behavioral responses from the time of ad exposures, online banner ads offer users to act immediately by clicking on them. As such, online advertisers may witness a much stronger impact of ads immediately after exposure, especially when the outcome behavior being measured is visits to the website.

Nevertheless, the model does not find significant evidence of deviation from geometric decay after the first day of exposure. That is, for $d \geq 2$, the credible intervals for all γ_{d} include 0, failing to reject $\gamma_{\mathrm{d}} = 0$. However, it is important to note that these findings do *not* confirm that the geometric decay assumption is correct for these time periods. Rather, we conjecture that it is

simply the lack of information in the data to make definitive inferences regarding these parameters. Specifically, due to the extremely small magnitude of the resulting weights w_{ct} on ad stock for ad impressions that occurred more than a few days ago, the values of γ_d have minimal impact on the values of w_{ct} and the resulting likelihood of the data, which may be one reason why for $d > 10$ or so, the posterior distribution is almost identical to the prior distribution.

Figure 3.7 Posterior Median and 95% Credible Intervals of γ_d



Other parameters of interest include c_c, or the initial impact parameters for different ad creatives that allow one to accommodate for heterogeneity in their effectiveness right after exposure. For illustrative purposes, we display the posterior distributions of c_c for acquisition and retargeting ads on a separate scale in Figure 3.8. Notice that the retargeting ads are overall estimated to be much more effective than the acquisition ads, with one exception being Creative 10, which has a similar initial impact to some of the acquisition creatives. These results, along with the much slower decay rates for the retargeting ads, seem to imply that retargeting ads are indeed more effective than the acquisition ads in bringing users to the target firm's website.

Figure 3.8 Posterior Distributions of Initial Impacts of Ad Creatives c_c

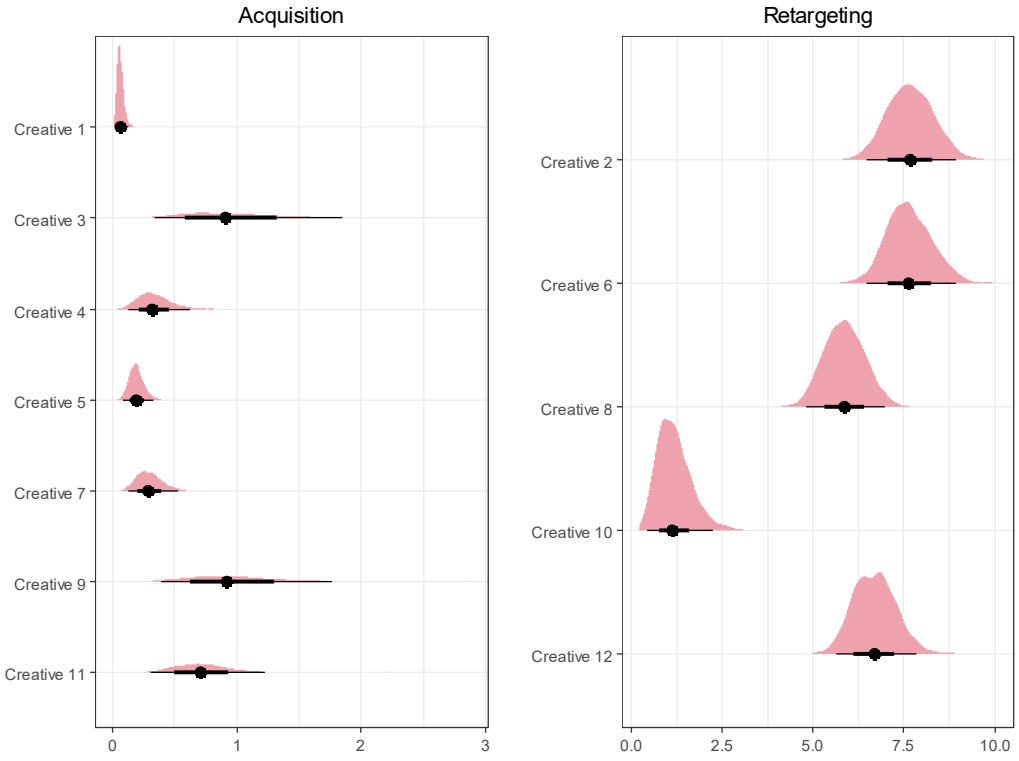
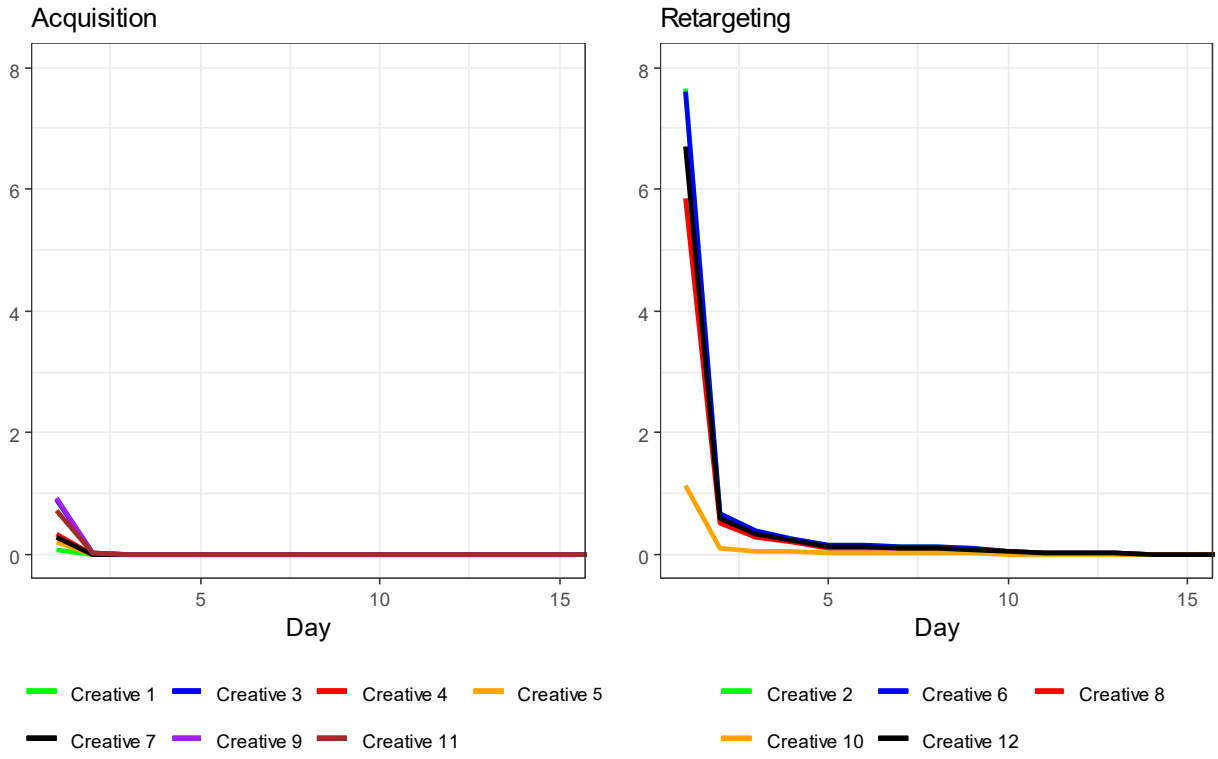


Figure 3.9 depicts the evolution of w_{ct} over time, or the decay curves for various creatives. While we do separate them into two panels — the left for acquisition creatives and the right for retargeting ones — for illustrative purposes, the scale of the y-axis is common to both so that they are easily comparable across the two panels. While we did estimate these curves for 30 days after ad impressions, all of them were basically indistinguishable from 0 after day 15, which corroborates our assumption of 0 impact of ads 30 or more days after exposure, so we truncated the latter halves of these curves in our plot. We can immediately see important patterns that are direct implications of the individual parameters that inform these curves, which we have discussed in detail so far. Specifically, the retargeting ads contribute a noticeably greater amount of goodwill to the ad stock on the day of ad exposure compared to the acquisition ads; the least effective retargeting ad creative seems to be on par with some of the most effective acquisition ones. Another important distinction to be made is that while acquisition ads are estimated to be

mostly unable to retain any meaningful impact after a day of exposure, most retargeting creatives do seem to make material changes to users' propensities to visit the target firm's website. Finally, common across the two panels are curves demonstrating a very dramatic drop after the first day and a relative flattening afterwards. Again, this is likely due to the significant negative estimate for γ_1 and others that are statistically indistinguishable from 0.

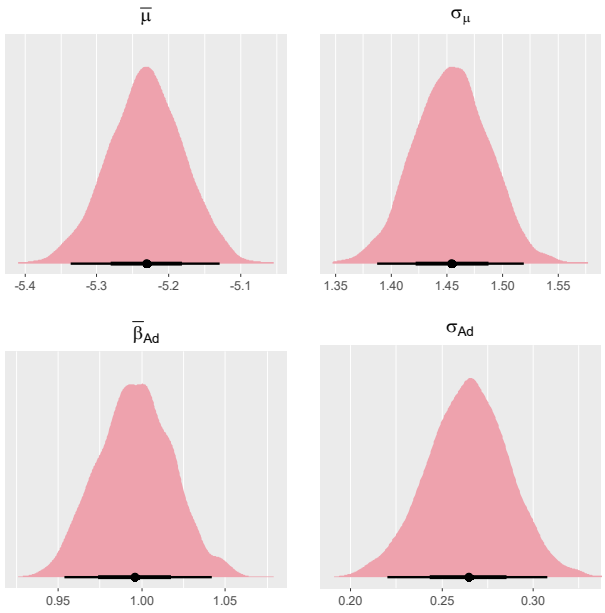
Figure 3.9 Resulting Ad Decay Functions for Acquisition and Retargeting Ad Creatives



While not the central focus our empirical analyses, it is worthy to note the degree of heterogeneity across the users in terms of their baseline propensity to visit the target firm's website and their responsiveness to the ad stock built through ad exposures. Figure 3.10 shows the empirical posterior distributions for each of the relevant parameters using posterior samples from all 6 MCMC chains. We see that the posterior median for $\bar{\mu}$ is well into the negative, meaning that without the goodwill from ad exposures, the overall population's propensity to visit the target firm's website is quite low. But at the same time, we do notice substantial variability in

users' baseline propensities, with the standard deviation of the Gaussian distribution being around 27.8% of the magnitude of the overall mean. A similar pattern arises for users' responsiveness to their respective ad stocks. The overall mean of the heterogeneity distribution is estimated to be positive as expected, again with a substantial degree of heterogeneity, but where it is unlikely for many users to exhibit a negative response to previous ad exposures.

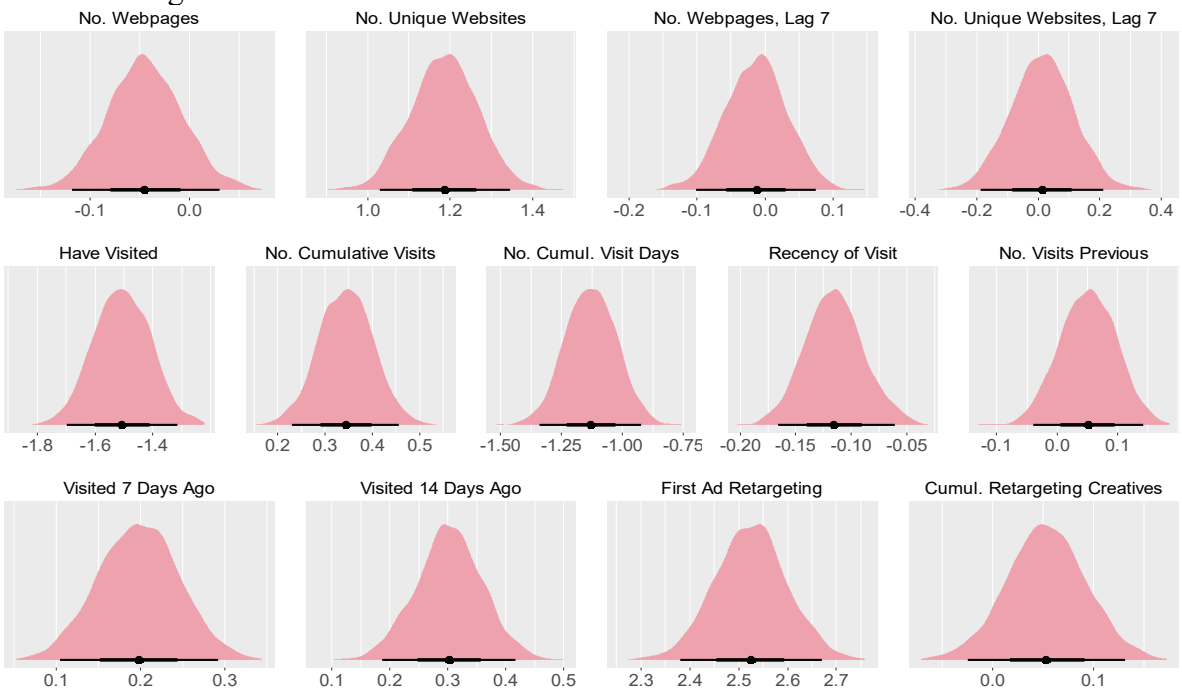
Figure 3.10 Posterior Distributions for User Heterogeneity-Related Parameters



We next briefly touch upon the coefficient estimates for control variables that might capture interesting patterns in our data, shown in Figure 3.11. First, the first two variables are corrections for the so-called "activity bias." While the simple number of webpages visits on a given day does not positively relate to the propensity of users to visit the target firm's website, the unique number of websites visited on that day does, implying that in this particular case, it's the variety of users' activities, not the amount, that needs to be accounted for. On the other hand, the 7-day lag variables for each have close to 0 impact on users' propensity to visit the website. The next sets of variables also show an interesting pattern. Users who have visited the target firm's website before are less likely to visit it again, especially if they have done so on multiple

days a long time ago. However, the user's number of previous visits to it does have a positive relationship with whether they are likely to do so in the future. These seem to suggest that users who have visited the target firm's website multiple times on the same day are those that may have a relatively higher propensity of returning. This makes intuitive sense in that these users may be those that have interacted with the target firm in an intense manner, suggesting an elevated level of interest in the target firm's product offerings. The 7-day and 14-day lag of visit indicators do seem to suggest that there may be a substantial cyclicality in users' browsing activities. Finally, as to be expected, there seems to be a notable difference in users' visit propensities whose first ad exposure in our dataset is that of a retargeting ad, who are most likely those that have visited the target firm's website before the start of the data window. Notice that this variable has to do with the time-invariant baseline propensities for each user, whereas the "Have Visited" variable is a time-varying variable that accounts for feedback of activities within each user but over time.

Figure 3.11 Posterior Distributions for Control Variable Coefficients



3.6 Simulation Studies with Gaussian Process Priors

3.6.1 Simulation Data Generation

In this section, we demonstrate that, using a simulation study, Gaussian Process (GP) priors can be effectively used to recover the latent decay functions, or the function relating the time elapsed since ad exposure and its current impact on the latent goodwill, with functional forms that diverge substantially from the frequently assumed parametric form of geometric decay. For simulation purposes, we mostly focus on the earlier part of the customer journey: we model the relationship between historic ad exposures and visits to the target firm's website of each user. Specifically, we simulate a dataset for 2,000 hypothetical users who are assumed to be targets of an online advertising campaign over a 75-day period, with two main components: 1) the number of ad exposures for each of the various ad creatives on each day, and 2) the number of visits to the target firm's website each day. The natural basis of our data generating process would be a log-linear Poisson count model where the latter has a stochastic multiplicative relationship with the former.

We aim to generate a dataset that closely resembles the real dataset that we plan to analyze with a similar framework in many key aspects, while maintaining a data generating process that is generalizable at the same time. First and foremost, as one of our ultimate goals of data analyses is to uncover how the impact of ads using multiple creatives (henceforth, c) *differentially* decay over time, we allow each user to be exposed to ads of multiple creatives multiple times, even within the same day. Then, the latent ad stock variable ($AdStock_{it}$) for each user (i) at each time period (t) would be calculated as the sum across previous time periods (days) and across creatives, of the number of ad exposures of each creative on each time period ($AdExposure_{ict}$) multiplied by the corresponding values of the decay function ($f_c(t)$) according

to the creative and the time passed since the ad exposure. In essence, the formulation of the ad stock variable is similar to the typical one in that it sums up the time-decaying impact of all previous ad exposures across all types of ads. Yet, we allow the decay functions to be more flexible than they usually are in two meaningful ways: 1) they don't have to be of geometric parametric form, and 2) the rate of decay, or the shape of decay functions are allowed to be different across creatives.

We also include a few randomly generated covariates ($\boldsymbol{x}_{it}$) that would serve as proxies for control variables used in our empirical analyses. For simplicity, we will be making i.i.d. random draws for each of the individual components from a standard normal distribution without imposing any correlations between them. Furthermore, in order to incorporate plausible heterogeneity across users in terms of their baseline propensity to visit the website, we will also include a time-invariant intercept term for each user (μ_i). For the heterogeneity mixing distribution, we will be using a Gaussian distribution, pursuant to the norm.

Finally, we believe the preponderance of zero outcomes, or a user never visiting the website on a given day, is a notable and substantially important pattern found in our empirical data that should be directly accounted for. Naturally, we alter the log-linear Poisson model into a zero-inflated one, in which the probability of a zero visit for each user each day is systematically inflated over what a typical Poisson model, along with the latent ad stock variable and control variables, would predict. Specifically, the final outcome would be exactly zero with a probability of π_0, whereas it will be generated from a Poisson distribution that accounts for the ad stock and the remaining covariates with the remaining probability $1 - \pi_0$.

To recap, the full model used to generate the data for our simulation studies can be specified formally as follows for $i = 1, \cdots, 2000$ and $t = 1, \cdots, 75$:

$$AdStock_{it} = \sum_{t'=0}^{t} \sum_{c=1}^{C} f_c(t - t') \times AdExposure_{ict} \tag{8}$$

$$\mu_i \sim N(\mu, \sigma^2),\ x_{itp} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1^2)\ \text{for}\ p = 1, \cdots, P \tag{9}$$

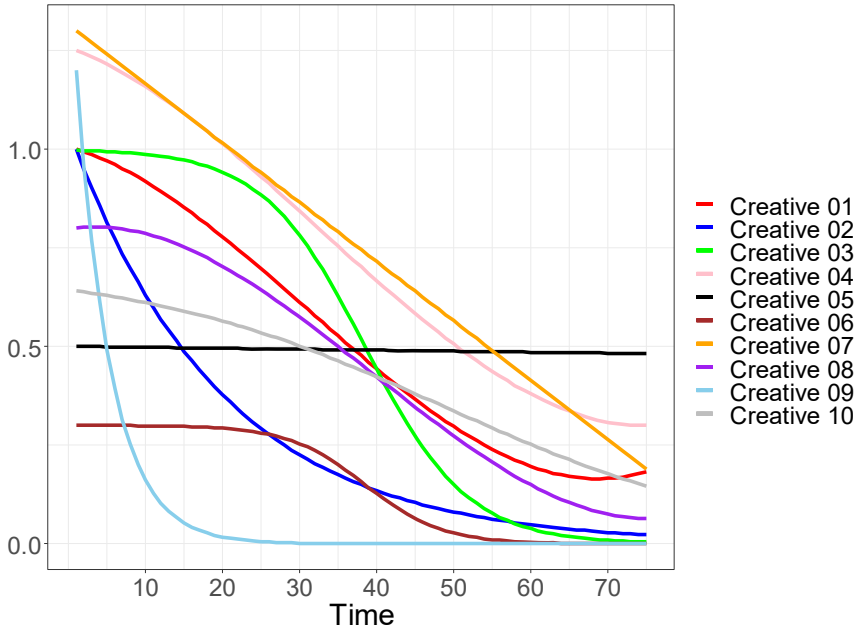
$$\lambda_{it} = \exp\left(\mu_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \beta_{Ad} \times AdStock_{it}\right) \tag{10}$$

$$NumVisit_{it} = \begin{cases} Poisson(\lambda_{it}), & with\ probability\ 1 - \pi_0 \\ 0\ , & with\ probability\ \pi_0 \end{cases} \tag{11}$$

We simulate the ad exposure history of each user by first determining the number of ad exposure they get over the 75-day period, by making a uniform draw from $\{3, 4, \cdots, 10\}$. Then, we stochastically determine the timing of each ad exposure and the ad creative used by making uniform draws from $\{1, 2, \cdots, P\}$ and $\{1, 2, \cdots, C\}$, respectively. While this might not be perfectly representative of the true data generating process in that users' activity levels and hence their ad exposures might be more concentrated in a narrower time window than a uniform draw might imply, we do not think it will materially influence the substantive results and validity of our simulation study.

As for the various ad creatives whose decay functions are to be recovered using GP priors, we introduce $\mathrm{C} = 10$ groups of ad creatives, which is close to the number of creative groups in our empirical data. The decay functions associated with these ten creatives are listed in Table B3 in the Appendix and their respective evolutions over the 75-day time window are depicted in Figure 3.12. We intentionally choose functions with a wide range of parametric characteristics: some are geometric, as the traditional geometric decay assumption would imply, while others are either polynomial or logistics function based. It is of interest whether there are substantial differences in the recoverability of latent decay functions belonging to various parametric families. To better align with our expectation of diminishing impact of ad exposures over time, the coefficients or parameters in each are carefully calibrated so that they are mostly decreasing functions over the relevant time windows.

Figure 3.12 10 Decay Functions Used in the Simulation Study



Other specifications for the simulation studies include $\mu = -9$ and $\sigma = 1$ for the Gaussian user intercept heterogeneity distribution, $P = 3$ and $\boldsymbol{\beta} = (-2, 1, 0.1)^T$ for the linear predictors, and $\beta_{Ad} = 1.5$ for the ad stock sensitivity. Note that the decay curves and the coefficient β_{Ad} are not separately identifiable without imposing specific identifying restrictions (beyond the regularization imposed by the prior distributions), so we will only be attempting to recover the product of these two quantities. Finally, the zero-inflation parameter is set to $\pi_0 = 0.3$ so as to reflect the relatively high proportion of zero outcomes in our empirical data.

3.6.2 Gaussian Process Priors

For decay function estimation, we need to specify the two main components of the Gaussian Process priors: 1) the mean function $m(t)$ that reflect the *a priori* expectations regarding the functional form and values, and 2) the covariance function $k(t, t')$ that determine the correlation between function values evaluated at different points, often called the kernel.

We will set a zero mean function to be $m_c(t) = E[f(t)] \equiv 0$ which assumes no *a priori* knowledge about these decay functions. While it is possible to inform the model with certain

expectations about the overall form of the function to be estimated (e.g., monotonic decrease of function over time) using parametric assumptions for m, the mean function is known to be less consequential in determining the function values compared to the kernels are (Dew and Ansari 2018), and we do not see a particularly strong need to impose any parametric restrictions on it.

While there are many plausible choices for the more estimation-relevant kernel, we will be using a squared exponential kernel here, where the strength of correlation between function values at two time periods would be negatively associated with the distance between those time points. This seems to be a reasonable assumption in that we expect the decay functions to be somewhat "smooth": function values for them evaluated at a pair of nearby points should be *a priori* more related to each other than those evaluated at another pair of values that are further apart. Specifically, $k_c^{SE}(t,t';\eta,\rho) = Cov(f_c(t), f_c(t')) = \eta \times \exp\left(-\frac{(t-t\prime)^2}{\rho}\right)$ for $\eta > 0,\ \rho > 0$. Here, the hyperparameter η, or the amplitude, dictates the overall variance of the function values away from the mean function, which in our case is 0. As η grows larger, the decay functions will tend to have values larger in magnitude, whereas the opposite would be true for small values of η. The other hyperparameter ρ, often referred to as the length-scale, controls the smoothness of the function. Larger values of ρ will result in smoother functions with fewer abrupt kinks whereas smaller values of ρ will imply a more jagged function. So as not to impose specific degrees of variance and jaggedness for our decay function, we impose priors for these hyperparameters as well, as opposed to picking specific values: $\eta \sim \text{Half} - \text{Cauchy}(0,5)$ and $\rho \sim \text{Half} - \text{Cauchy}(150,75)$.

Note that in the current simulation study, we make an implicit assumption of common hyperparameters (η, ρ) for all 10 decay curves. There is a subtle tradeoff being made here. For one, making this assumption dramatically reduces the computation time necessary for estimation,

as we only need to calculate one common covariance matrix for all curves. If we were to have separate sets of hyperparameters uniquely applicable to each curve, we would need computing power and time to achieve convergence in the Markov chains. Another advantage of this assumption is that there is a lot more data that informs the estimation of these hyperparameters. As we see later, the effective sample sizes for these parameters imply some difficulty in estimating them, and having 10 sets of them will most likely require a significantly longer chain for all of them to be robustly estimated. On the other hand, one major drawback for making a common hyperparameter assumption is the imposing of similar amplitude and smoothness across all curves. This might be a substantial restriction when the curves demonstrate substantial heterogeneity in these characteristics. This might also be one of the factors contributing to the minor divergences between the true and estimated decay curves as can be seen in our estimation results.

While GP priors in theory lends itself to a multivariate Gaussian distribution for function values at any number of points, we will only be estimating the function values of each of the 10 decay functions at fixed set of discrete points of time: $\boldsymbol{t} = (1, 2, \cdots, 75)$, in which case our prior specification of $f_c(t) \overset{\text{i.i.d.}}{\sim} \mathcal{GP}(m(t), k_{SE}(t, t'; \eta, \rho))$ essentially reduces into assuming a multivariate Gaussian prior distribution on $f_c(\boldsymbol{t}) \overset{\text{i.i.d.}}{\sim} N(m(\boldsymbol{t}),\ \boldsymbol{K}(\boldsymbol{t}, \boldsymbol{t}))$ for all practical purposes. Also note that in practice, for numerical stability, rather than making posterior draws for $f(\boldsymbol{t})$ directly from a multivariate Gaussian with the corresponding mean ($\mathbf{0}$) and covariance matrix $\boldsymbol{K}(\boldsymbol{t}, \boldsymbol{t}; \eta, \rho)$ at the hyperparameter values, we would be taking a random draw from the 75-dimensional standard normal distribution and then pre-multiplying the lower Cholesky-decomposition of the covariance matrix. Finally, we set priors for other parameter values in the model: $\beta \sim \mathrm{N}(0, 1^2),\ \pi_0 \sim Beta(2, 2),\ \mu \sim N(0, 5^2)$, and $\sigma \sim Half - Cauchy(0, 5)$.

3.6.3 Simulation Study Results

For estimation, we again use the Stan programming language (Carpenter et al. 2017) with 1,500 posterior draws made from each of the 6 chains, where the first 750 from the warmup are discarded. Estimation times for each of the 6 chains in seconds are listed in Table B4 in the Appendix.

We again look into the trace plots as well as the effective sample size (ESS) and Rhat for some of the key parameters of our model to gauge sampling quality and chain convergence, listed in Table B5 in the Appendix. While the effective sample sizes for the GP hyperparameters are somewhat modest, our table of Rhats and the trace plots indicate no overt signs of failed convergence, so we will present the simulation study results using posterior samples from all 6 chains.

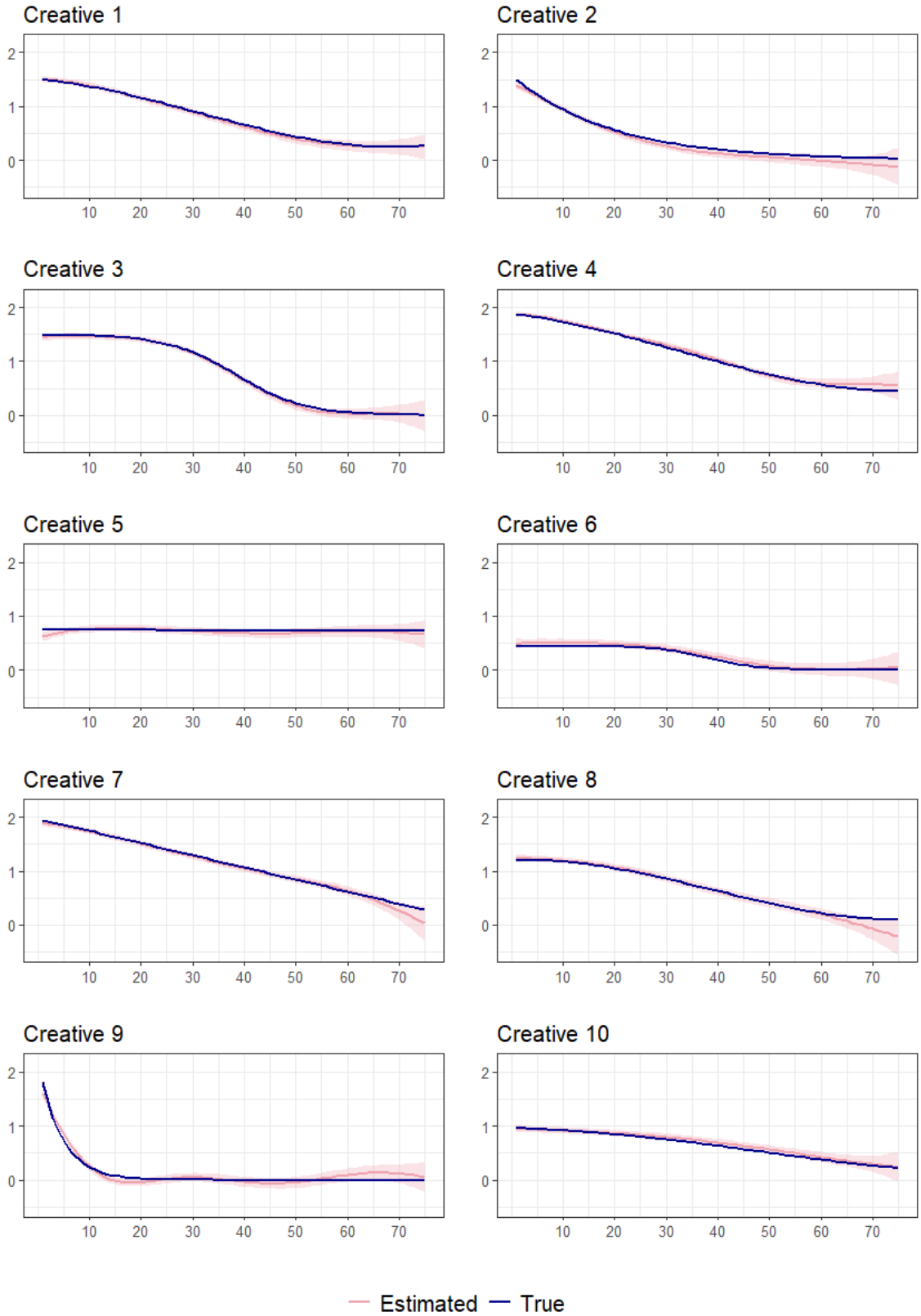
We present in Figure 3.13 the estimation results of 10 decay curves, our focal quantities of interest, recovered using the GP priors by connecting the median function values of the pooled posterior draws across all 6 chains at each point of $t = 1, 2, \cdots, 75$, as well as their respective 95% empirical credible intervals, and overlay them with the respective true decay curves used in generating the dataset used for the simulation study. We again note that the quantities recovered below are the individual decay curves multiplied by $\beta_{Ad} = 1.5$ as they are not empirically separately identifiable.

We can easily see that the recovery of these decay curves were mostly successful; the overall shape and magnitude of the estimated curves by and large overlap with the true functions, and even when minor deviations occur, the true curves mostly lie within the 95% credible intervals, which manifest tight upper and lower bounds for the decay functions. A more careful observation of the credible intervals reveal that they tend to be wider near the right end, and that

this is the region where the estimated curves diverge the most from their respective true ones. This can mostly likely be attributed to the fact that there are relatively fewer ad impressions in the simulated dataset that can inform the estimation of this part of the curves. Whereas the earlier part of the curves that represents the impacts of a recent ad exposure may be informed by ad exposures at nearly all points of time within the data window, the later part that represents the impact of an ad exposure that occurred a long time ago can only be inferred from the behaviors of users with ad exposures near the very beginning of the data window. Nevertheless, we still believe that the estimation of function values near the end of the curve may have been made more robust due to the use of GP priors. Compared to a classical nonparametric approach of function estimation that would almost solely rely on the very sparse signals that are pertinent to the nearby points for which the function values are being estimated, potentially leading to very unstable estimation, GP priors enable a more systematic regularization owing to the covariance structure that is also informed by the earlier parts of the curve where the information is more plentiful.

Another feature of the results we would like to highlight is that GP priors seem to be able to accommodate a wide range of functions that might have different parametric forms. We had made deliberate choices to include decay functions that range from simple polynomials to exponentials and logistic-shaped ones in order to check whether GPs lend themselves more readily to certain families of functions compared to others. While the nonparametric nature of the estimation may suggest that the underlying mathematical properties of the decay functions would not be of grave concerns, we find it assuring that the results do confirm that our approach need not be restricted to instances where the decay functions take a specific functional form.

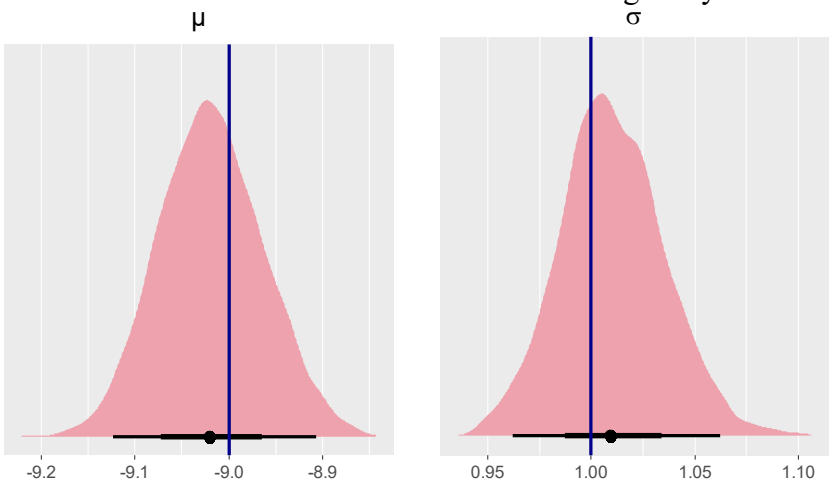
Figure 3.13 True Decay Functions (Blue) and Posterior Median and 95% Credible Bands (Pink) Estimated for 10 Ad Creatives



We next turn to estimation results for the other key parameters in our model in Figure 3.14. As in the previous section, the empirical posterior distributions are presented with a bar at the bottom whose thinner line represent the range of 2.5% and 97.5% empirical quantiles and the thicker line the range of 33% and 66% empirical quantiles, and a dot that represents the sample median. These will also be accompanied by a vertical line within each plot representing the respective true value used while generating the simulated dataset.

First, we look into the estimation of the parameters for the Gaussian heterogeneity mixing distribution. We can clearly see that both $\mu = -9$ and $\sigma = 1$ are successfully recovered with remarkable precision. Both posterior distributions seem to be well-behaved in that there don't seem to be any clusters of samples implying a poor convergence of any chains, and the true values lie near the middle of the range of the respective empirical posterior distributions.

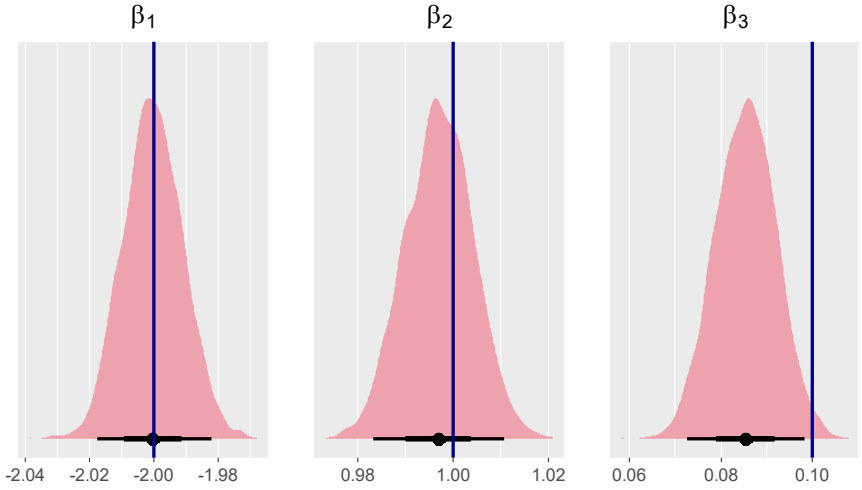
Figure 3.14 Posterior Distributions and True Values of Heterogeneity-Related Parameters



Next, we check the recovery of the three coefficients for the control variables, $\boldsymbol{\beta} = (-2, 1, 0.1)^T$. Figure 3.15 demonstrates that all three posterior distributions do not exhibit troubling signs regarding convergence. The empirical posterior distributions for the first two coefficients include their respective true values near the middle of their range, again confirming successful estimation. The only exception is that for β_3, where the true value of 0.1 does lie

within the range of the empirical posterior distribution, but just barely outside the 95% sample quantile range. However, we regard this slight miss more as a confluence of random generation of the simulated dataset and the stochastic nature of the HMC estimation.

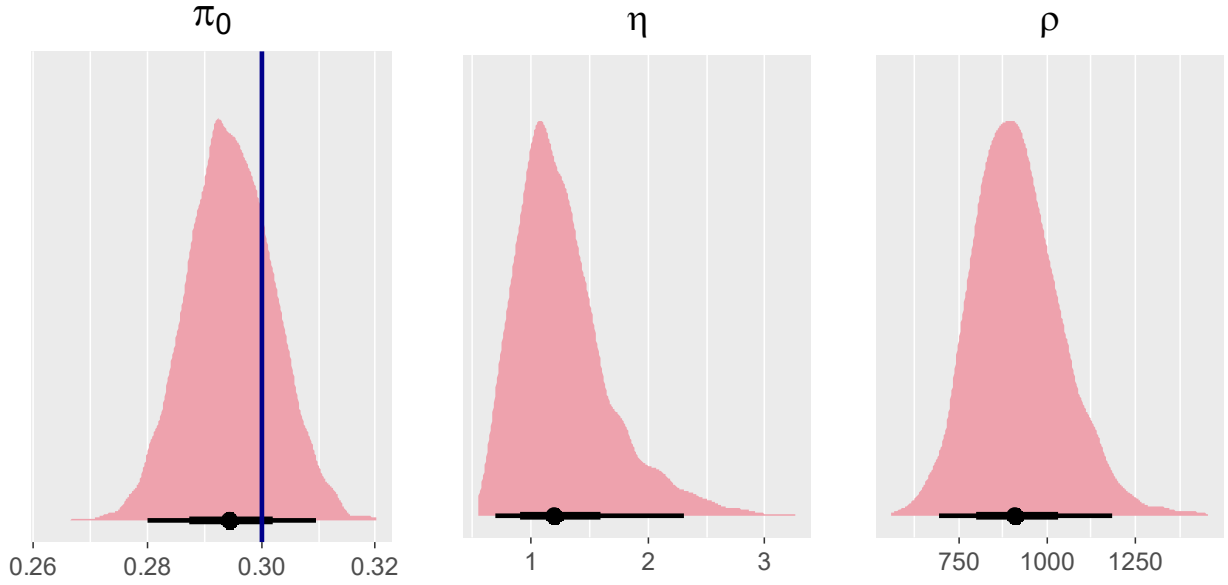
Figure 3.15 Posterior Distributions and True Values of Control Variable Coefficients



We see in Figure 3.16 that the estimation for the zero-inflation parameter for the Poisson distribution, π_0, is also quite accurate in that the true value of 0.3 is near the middle of the range of a well-behaved empirical posterior distribution. Finally, we look into the posterior distributions for two hyperparameters for the GP priors: η, that dictates the amplitude, or variance, of the function values for all 10 decay functions away from our mean function of 0, and ρ that determines the overall smoothness of these functions. While there are no "true values" to which we would compare the posterior distributions, these values contain important information about the overall shape of our decay curves. The empirical posterior median for η is around 1.20. Given that $k_c^{SE}(t,t;\eta,\rho) = Var\big(f_c(t)\big) = \eta$, this parameter could alternatively be interpreted as the prior variance of the function value at any given point. Observing that our decay functions mostly lie between 0 and 2, we would argue that this estimate "makes sense." As for the length-scale parameter ρ, the empirical posterior median is around 906.8. While there is no such direct

intuitive interpretation for this parameter, it seems worthwhile to mention that the prior distribution for it was Half $-$ Cauchy$(150,75)$, the center of mass of which is very far away from the posterior. It seems reasonable to assume that had we used a half-normal distribution for the prior vis-à-vis half-Cauchy that has a much fatter tail and accommodates values distant from the center comparatively freely, we would have ended up with a much smaller value for ρ, which may have resulted in more jagged decay functions than what we currently have.

Figure 3.16 Posterior Distribution and True Value of Zero-Inflation Parameter π_0, and Posterior Distribution for GP Hyperparameters η and ρ



3.7 Conclusion and Future Directions

In this essay, we propose more flexible approaches of capturing the evolution of the impact ad exposures have on consumers over time. Specifically, we try to relax the restrictive assumption extensively used in the marketing literature: a common decay rate across all types of ads that stays constant over time. Through an empirical analysis using a fully parametric Bayesian approach and individual-level data recording users' ad exposures and their web

browsing data via internet cookies, we uncover material evidence that the prevalent common rate geometric decay assumption is violated in multiple ways. For one, we find that retargeting ads seem to be retaining their impact on consumers' propensities to visit the target firm's website over a considerably longer time when compared to their acquisition counterparts, implying a need to account for heterogeneity of decay rates across ad creatives containing different type of information. Our analysis further reveals that, likely owing to the nature of online banner ads, we see a more noticeable drop in the ads' contributions to the latent ad stock after the day of exposure than what the geometric decay specification would allow for, implying that such assumption would most likely underestimate the ads' efficacy on the day of exposure and overestimate it over the following days.

The results of our analyses included in this essay have important implications for marketing researchers and practitioners alike. As detailed in the literature review, the geometric decay assumption has been used indiscriminately when trying to relate past marketing efforts to current or future behavioral reactions from consumers with very few exceptions. However, our results strongly suggest that such assumption may be too restrictive and that a relaxation of these assumptions may be required to more accurately evaluate the efficacy of online banner advertisements that are intended to not only garner immediate reactions in users' clicks but also leave long-lasting goodwill toward the firm or the brand. While our empirical analysis pertains to capturing the dynamic effects of online advertising on internet users, the basic idea and framework need not be limited to such contexts; any marketing efforts that are intended to make long-lasting impact that endures over time would be suitable.

Practitioners also would benefit greatly from this flexible approach of measuring the latent ad stock as it is a pertinent question of which the answer would vastly affect their

decisions on ad placements over time. Specifically, a tradeoff may be relevant when some ad content is more capable of generating a more immediate response from the consumers whereas others, while not incendiary right after exposure, may "stick" in their memory for a longer time and foster persistent goodwill. If we assume a common decay rate for ads of all types, marketers would be advised to almost exclusively go with the former over the latter.

While our analyses illuminate valuable findings that offer substantive contributions to the marketing literature, we believe there are worthwhile set of analyses that would further complement the current ones presented in this essay. One natural direction for future research would be to apply the Gaussian Process priors, of which we demonstrated the effectiveness and reliability via simulation study, to the empirical data that we analyzed in order to uncover even less restricted decay curves that can more readily accommodate creative heterogeneity and dynamic patterns over time. While not directly modeled in the current simulation study, we hope to be able to introduce critical dimensions of user-level heterogeneity as well: that in terms of their baseline propensity to engage in the desired behaviors, reactivity to the marketing activities of the target firm, and potentially, their various decay rates of ads' impacts on their goodwill. One potential challenge that might arise relate to the fact that our empirical individual-level dataset is considerably sparser than the generated dataset used for the simulation study; the number of ad exposures per user may not be sufficient to identify precisely the decay functions using GPs. Whether we can overcome such data limitation is an empirical question and it may become necessary to adjust the degree of discretization of the temporal dimension. We are also currently working on applying a similar nonparametric approach to an aggregate ad spending and sales data from a global cell phone manufacturer through various channels. While we would lose the granularity of our current individual-level data, we are hoping that such aggregate-level

dataset would contain a stronger signal compared to the individual-level data that we have used in the current project.

Another future goal of this project is to develop a more holistic model that relates the ad impressions to site visit and to conversions in the end while introducing a more rigorous correction for the potential endogeneity created by the selection of users into the retargeting campaigns (Li and Ansari, 2014). In our current empirical analysis, we have not yet built a model that simultaneously connects ad exposures to visits to the target firm's website and consequently those visits to conversions, the ultimate goal of the target firm's marketing efforts. We believe doing so will allow us to build a more complete picture of how firms' advertising efforts affect users through the various stages in their customer journey. Furthermore, while we do include many features in the model that would address the potential endogeneity issues – in terms of which users are being retargeted by the advertiser – to a great extent, there may still avenues for further improvements in this regard, such as estimating a joint process model with the ad retargeting mechanism and how the ad exposures affect users' future behavior. We believe such improvements would also allow us to capture a fuller picture of the data generating process of our data.

Once we have finalized the model and its estimation, our goal will be to devise an ad placing mechanism across users and over time while considering the heterogeneous initial impact and decay rates of ads across creatives to maximize the overall visits to the website and ultimately conversions under a budget constraint or a limit on the total number of ad exposures. This task is far from trivial as the space of feasible solutions is astronomically vast and the computation needed for calculating the objective function is also quite costly. Whereas this problem can be viewed as a traditional integer programming under linear constraints (de Reyck

and Degraeve 2003), given the potentially long computation time to evaluate the objective function, adopting the general principles of a Bayesian Optimization approach (Frazier 2018) could be especially well-suited for this application as it is able to reduce the number of evaluations needed to locate the globally optimal solution by imposing a Gaussian Process prior on the objective function and evaluating it on points where there is a maximal value of learning based on previous evaluations. Given that the dimension of solution space is much larger than what is typically recommended for Bayesian Optimization (less than 20), applying it as is might not be feasible. However, it could be useful in coming up with a useful starting point that would dramatically reduce the number of iterations needed to identify the general area of the solution as compared to starting cold. Given that the optimal placement of various types of ads across a large number of users and over time is one of the most pressing issues that face managers, developing a computationally light algorithm that helps solve this high dimensional problem seems to be a promising avenue for future research.

CHAPTER 4

To Whom, When, and What to Ask?: Assisting Smoking Cessation and Detecting Relapse with Real Time Interventions

4.1 Abstract

Marketers frequently face a dilemma between communicating with their customers as frequently as possible to convey product or promotional information or request for their feedback, and maintaining their attention without getting them fatigued. Whereas reaching customers has become increasingly costless due to the prevalence of digital means such as emails and push notifications through mobile apps, consumers often express displeasure due to the incessant efforts to grab their attention. In this work, we examine a similar tradeoff and offer a framework in which marketers and researchers are able to balance the two at the individual level by adjusting the frequency and content of future ones according to the outcomes of previous ones. We do so in the context of mobile health, specifically Ecological Momentary Assessments (EMAs) for a smoking cessation program, which consist of a set of questions regarding the participants' current emotional state and intent to quit smoking. We utilize a Bayesian dynamic factor model that can readily accommodate the inattentiveness or nonresponses of each participant and pool information within each participant over time and also across participants to inform which items within the EMA are the most relevant in predicting future lapsing and how likely each of them is to drop out of the program.

4.2 Introduction and Literature Review

Ecological Momentary Assessments (EMA) are a useful tool in obtaining repeated measures of behavioral or psychological data real-time in a relatively unobtrusive manner (Shiffman 2009). They have been used extensively in a wide variety of psychological studies going back decades (e.g., Weiss et al. 1999, Larson et al. 2001, Dunton et al. 2015), but especially so in the contexts of substance, e.g., studies of alcohol and tobacco use (e.g., Todd et al. 2005, Cooney et al. 2007). These assessments can take various forms, ranging from a simple indication whether a specific event has happened over the course of the day to an extensive set of questions involving measurements of mood and behaviors collected by smart devices. While the information value obtained from frequent EMAs are widely acknowledged (Delfino et al. 2001), concerns for response overload that may increase response fatigue, lead to dropouts, or compromise data integrity (Larson 1989, Litt et al. 1998) spurred many studies to limit their EMA prompting to around 4-5 times day (e.g., Shiffman and Waters 2004, Beckham et al. 2008).

While EMAs can be used as less intrusive forms of measurements of the current psychological or behavioral conditions of participants, they can also be used as intermittent reminders for them to stay their course away from undesirable behaviors (Rowan et al. 2007), acting as a type of intervention themselves. There has been active research regarding adaptive interventions (AI), where the type or degree of interventions adapt according to how the participants react to the previous interventions, as opposed to letting a predesignated intervention scheme carry on regardless of the observed proximal outcomes during the experiment (e.g., Klasnja et al. 2015). This approach may be helpful in improving the distal outcome by assigning treatments that suit each participant's needs at that precise moment, with the proximal outcomes informing the assignment. One framework that allows for randomized testing of effectiveness of

AIs is SMART (Sequential Multiple Assignment Randomized Trial), which uses random assignments at each decision point, while adjusting the set of assignable treatments based on the outcome for each participant, resulting in multiple AIs being embedded into one clinical trial. It has been widely used in a variety of contexts such as substance use (e.g., McKay et al. 2015) and cancer treatment (e.g., Kidwell 2014).

Given the potential dual nature of role of EMAs – assessment and intervention – in this project we aim to improve their deployment in two major ways. First, we want to optimize *when* these EMAs are activated across each participant. Currently, EMAs are often designed to be activated at fixed time intervals (e.g., Larson and Kasimatis 1991, Beckham et al. 2008), or randomly within predetermined time windows for all participants (e.g., Rowan et al. 2007, Dunton et al. 2015). While this approach may be beneficial due to the ease of implementation and consistency across users, it does not fully incorporate the information obtained from previous responses from the individual. For example, if some participants are deemed to be of very low risk of lapsing and are expected to experience minimal changes in their conditions in the near future, the need to activate an EMA at that time may be considerably lower for these participants compared to those who had a high estimated probability of lapsing soon. Given the potential response fatigue that might worsen if pinged too frequently (Beal and Weiss 2003), it may make sense to limit the frequency of EMAs for some participants as long as doing so does not negatively affect the trajectory of their behaviors.

Secondly, we would also like to customize *which* questions should be included in the EMA each time it is activated for each participant. While the frequency of EMA would be an important factor in response fatigue, the length of the EMA itself might be another key contributor to it (Lukasiewicz et al. 2007, Davidson et al. 2017). While repeated measures of a

single construct may enhance the measurement accuracy, such benefits should be carefully measured against the potential for lowering response compliance (Li et al. 2022). There may also be a few questions the responses to which are very highly correlated with those to the others, in which case we may be able to make reasonably accurate guesses without directly asking them (Adigüzel and Wedel 2008). It is also important to note that even for two questions with similar predictive powers for the proximal outcome, responses to one may fluctuate much more over time compared to those to the other, and the magnitude of such variability might also be heterogeneous across participants. It might be possible to customize the content of the EMA for each participant by limiting the repeated inclusion of the latter type of questions without undermining the informational value of the overall EMA. Similarly, some participants might be much more patient or reliable in answering the questions than the others, evidenced by lower nonresponse rate or infrequent straight-line responses, in which case it might make sense to adopt a relatively longer EMA that will help boost the information obtained by it, compared to those used for participants who are more prone to experience response fatigue while filling out the EMAs that can lead to future noncompliance.

However, there may be some practical concerns in implementing such customization of EMA activation and content across users over time. The first stems from a "cold start" problem (Padilla and Ascarza 2021), both in the earlier stages of the experiment where all participants are starting fresh, and when a new participant joins the program, where there is a lack of data and responses that inform the study designers about their individual characteristics. As the individual EMA customization would depend heavily on the patterns found in previous responses of the individual users, such as their past responses to specific questions, their overall attentiveness while responding to the EMAs, when they receive EMAs and which questions are included in

each may undergo initial periods of high volatility in response to the early responses before enough data accumulate for each participant and the EMAs stabilize as a result. In a similar vein, there might be instances where the earlier inputs of a participant have such a disproportionately strong impact on the design and implementation of subsequent EMAs that questionnaires included may keep asking highly irrelevant questions that aren't very informative for the current diagnosis of their current status and future design of EMAs (Toubia et al. 2007).

Another related potential concern has to do with the *dynamic* nature of the quantity EMAs are often trying to measure. The very reason why EMAs are activated multiple times throughout the day or the week is to make sure they reasonably capture the fluid changes participants experience in terms of their moods, urges, and behaviors that vary substantially over time. Researchers would need to consider the possibility that the relationship between the questions within the questionnaire and how predictive each is of future behavior, not just the responses to the questions themselves, may also fluctuate over the course of the study (Cunha et al. 2010). For example, a participant's self-assessed prospects of successfully abstaining from smoking may be heavily correlated with their responses to their current anxiety levels in the beginning of a smoking cessation program but the two may become less related the longer they maintain abstinence and their confidence grows. At the same time, while stress may be an extremely strong predictor of smoking relapse in the near future for many individuals in the initial stages of a smoking cessation program, it may no longer be as some participants may successfully break the immediate reactionary link between stress and smoking. Failing to account for such evolution over time of the factor structure of the questions and how each affect the observable behaviors may lead to poor decision in terms of when each participant should be reassessed using which question, as well as how likely they are to relapse in the future.

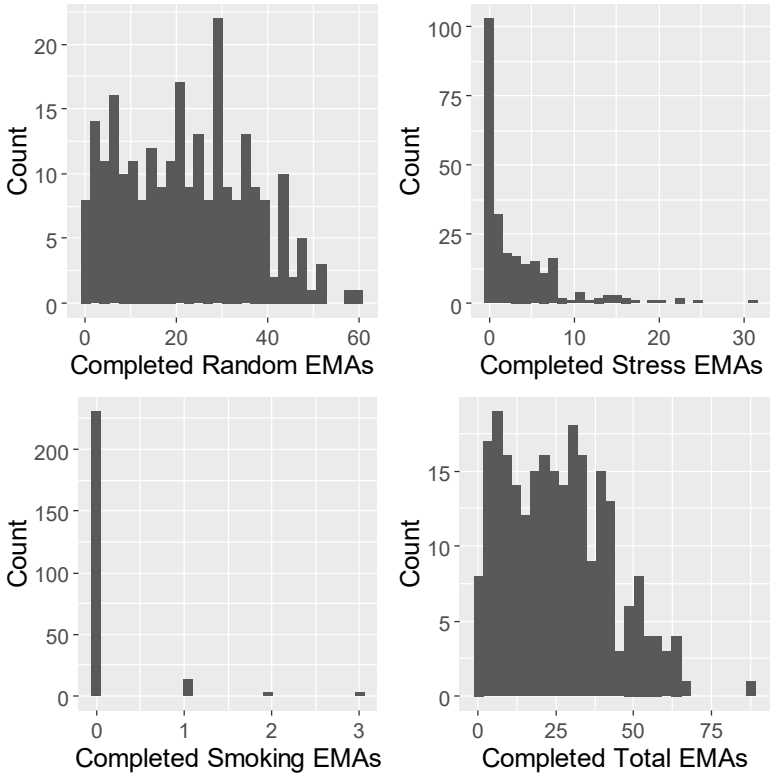
4.3 Data

The dataset we plan to use for this project comes from a smoking cessation program where the goal of the program was to deter smoking behaviors, collected from June 5th, 2017 to December 6th, 2019. It started with 303 participants who were compensated for completing the EMA and wearing sensors to detect their stress levels and smoking. 30 of them dropped out or withdrew in the process, leaving 273 in the final pool, 270 of which were included in our dataset. Among them, only 251 had at least one complete EMA. There were three main types of EMAs utilized during the study: 1) random EMA, 2) stress EMA and, 3) smoking EMA. EMA of the first type was activated 4 times a day, randomly within 4 evenly divided time intervals within each day while the participants were awake, whereas the second type of EMA was activated when the current stress level of the participant was deemed high by the sensor participants were carrying. The final type of EMA was activated when the sensors detected signs of smoking. Two versions of questionnaires were used, with the shorter version consisting of 53 questions and the longer one 62, where questions assess the current general psychological state, the urge to smoke, and how many cigarettes the participant has smoked, if any, since the most recent EMA.

Our dataset has records of 10,010 EMAs prompted, of which 6,417 (64.1%) were completed, 3,233 (32.3%) were missed, and 360 (3.6%) were abandoned for multiple reasons such as timeout. The average number of all (completed) EMAs across participants was around 37.1 (25.6), with a standard deviation of 18.8 (17.1), the median being 39 (23). As can be easily seen from Figure 1, the most common type of EMA is the random EMA (mean: 22.2, SD: 13.7), with smoking EMA (mean: 0.12, SD: 0.34) being the least frequently completed and stress EMA (mean: 3.3, SD: 4.9) in between. We hypothesize that due to the relatively thick file of records for each user, we may be able to uncover not only heterogeneity across users in terms of the

predictive nature of each question to their future smoking behavior, consistency of their responses to each question, and their response fatigue tolerance, but the evolution of each over time.

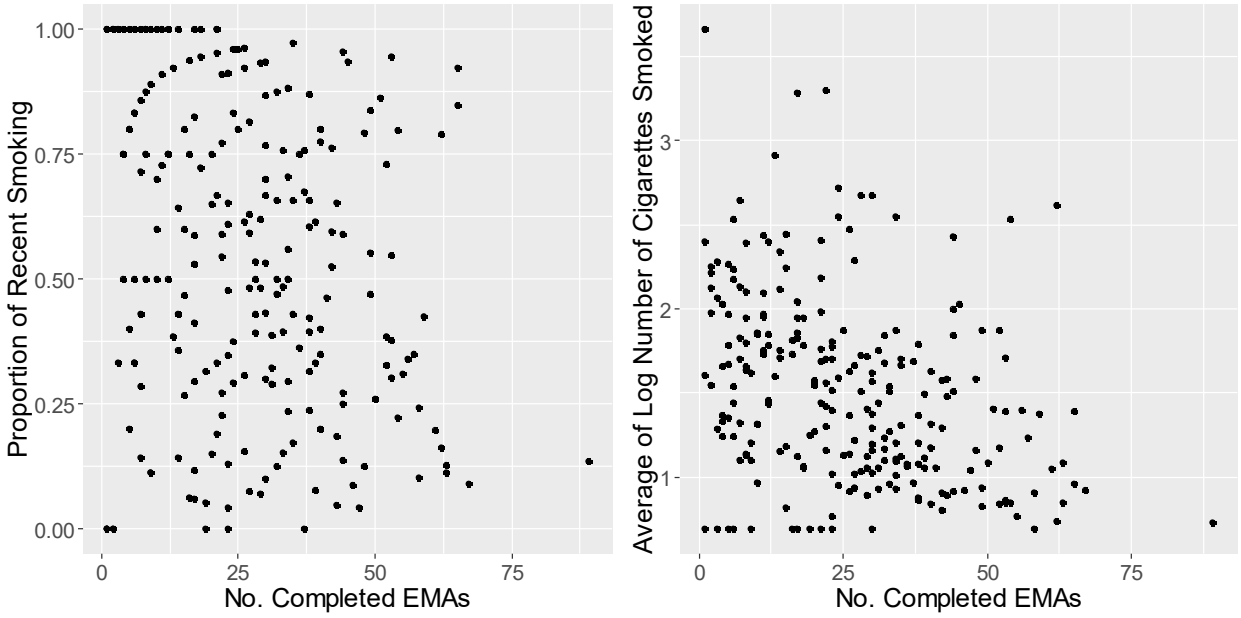
Figure 4.1 Histogram of Number of Completed EMAs Across Participants by Type



In order to estimate a model that can accurately predict whether the participants will relapse based on their previous responses to EMAs, we need reasonable variations in the proportion of EMAs in which users respond positively to the question "Since the last assessment, have you smoked any cigarettes?," which serves as the outcome variables, both within and across users. The left panel of Figure 3, where each dot represents a participant, clearly shows that our dataset does have such qualities: while we have many users who seem to be struggling from relapse (almost 100% of EMAs reporting having smoked after the previous assessment) and those who are reportedly very successful (with close to 0 such EMAs), we also have a lot of respondents in between, implying both within and across-user variation in their cessation relapse.

The right panel of Figure 2 depicts the distribution of the average logged number of cigarettes smoked against the number of completed EMAs, which also demonstrates that there are reasonable variations across participants in terms of the number of cigarettes smoked during each instance of relapse.

Figure 4.2 Scatterplot of Proportions of EMAs with Positive Reports of Recent Smoking Against the Number of Completed EMAs (Left) and Scatterplot of Average of Log Number of Cigarettes Smoked Against the Number of Completed EMAs Across Participants (Right)



The correlation between the number of completed EMAs and the proportion of EMAs reporting recent smoking has a statistically significant negative correlation of -0.29 (with the correlation being -0.26 if we log-transform the number of EMAs), potentially implying that users who actively complete the EMAs tend to be those who are more successful in maintaining abstinence from smoking. Similarly, the correlation between the number of completed EMAs and the average log number of cigarettes smoked is also negative and statistically significant at -0.34. These may be valuable pieces of information in that the continued participation of EMAs itself can be informative in predicting future relapse and the number of cigarettes smoked during

the relapses, which also highlights the importance of dealing with potential selectivity issues while modeling the relationship between participants' responses and smoking relapse.

4.4 Exploratory Analyses on Response Quality

4.4.1 Correlations between Items

Given that users lapsing and their inattention while responding to questionnaires are frequent concerns for these types of studies, we first look at whether there are any worrisome patterns in the participants' responses to individual questions. More specifically, we look into whether the participants' responses to questions that we would expect *a priori* to be strongly correlated actually are, in the direction we would expect them to be. For example, we would expect users who respond affirmatively to one item related to a positive affect (e.g., "Right now, I feel happy.") would tend to do so to another closely related one (e.g., "Right now, I feel joyful."). Similarly, we would also *a priori* anticipate a negative correlation between their responses to two items that concerns a positive emotional state ("Right now, I feel happy.") and a negative one (e.g., "Right now, I feel sad."). Considering that there are many more complex dimensions of emotions beyond the mere positive vs. negative one (Scherer 2005), the data's failure to pass these simple sanity checks, while justifying the need for improving participant attention and engagement throughout similar studies even more, would cause some concerns for using this particular dataset as a source of information that guides such efforts.

Below is a table that shows another item in the questionnaire with the strongest positive and negative correlations for a few items in the EMA (an analogous result along with the list of full set of 37 items can be seen in Table C1 of the Appendix) that are obtained by first calculating the *within*-participant correlation and then averaging across participants. We can

easily see that the results make intuitive sense; for example, happy and joyful share positive valence, whereas irritable inherently possess a negative one. With items particular to participants' attitude towards smoking, the item tied to the pleasure of smoking is positively correlated to the urge to smoke and negatively correlated to the motivation for abstinence. While the direction of the correlations between items mostly makes sense, the magnitude of the correlations – especially the negative ones – are not as substantial as one might expect, which may be hinting that there may be some lack of attentiveness among participants while responding to these questions.

Table 4.1 Sample EMA Items with the Strongest Positive and Negative Correlations

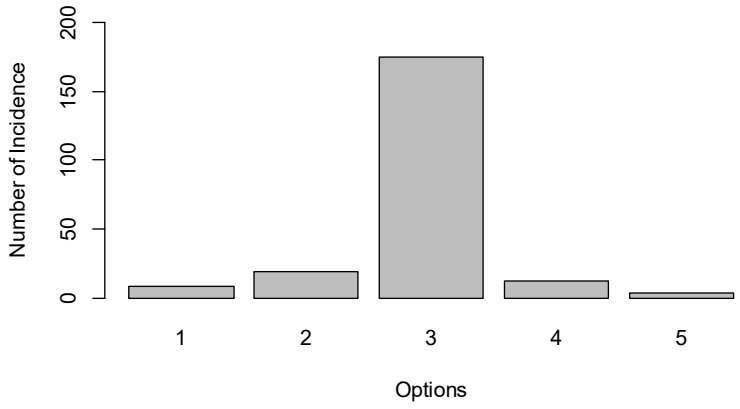
Item	Item with Strongest Positive Correlation	Correlation	Item with Strongest Negative Correlation	Correlation
Right now, I feel happy.	Right now, I feel joyful.	0.460	Right now, I feel irritable.	-0.328
Right now, I feel tired.	Right now, I feel restless.	0.370	Right now, I feel active.	-0.269
Right now, Smoking would improve my mood, be pleasurable, or help me cope with this situation.	Right now, I have an urge to smoke.	0.414	Right now, I am very motivated to be abstinent from cigarettes.	-0.196

4.4.2 Straight-Line Responses

While the above analyses do seem to reassure us that the participants overall seem to have paid attention doesn't mean that every individual in the sample did. For one, it would make sense to see if there were any significant patterns of straight-line responses (i.e., choosing the same option for every item), as it is one of the most straightforward red flags in response quality to detect. Considering that the participants were compensated for completing each EMA, there may have been certain individuals that may have resorted to such behavior in order to maximize their compensation with minimum effort. To identify these patterns, I calculated the "standard deviations" of the responses to 37 questions within each EMA by converting them into a

numerical scale. A complete straight-line responding would be captured by standard deviations of exactly 0, whereas a very small standard deviation would also signal that 1) either the respondent was a bit more sophisticated and tried to avoid complete straight-lining while not actually paying attention, or 2) the users tend to not vary their responses across the items, which might warrant participant-specific re-rendering or re-scaling of their responses if we were to compare their responses to those of other participants (Ying et al. 2006).

Figure 4.3 Distribution of Responses of a User Engaging in an Almost-Complete Straight-Line Responses



In total, we find 26 EMAs that manifest complete straight-line responding (i.e., no variation in their responses for any of the 31 items), distributed across 13 users, an extreme case was being one user who have chosen the middle option for all items on these 8 EMAs. There were also 38 users in total who showed complete or almost-complete straight-line responses at least once, as defined by the bottom 3% in the standard deviations across items within a single EMA. Figure 4.3 shows the distribution of responses to all items across 7 EMAs of a particular user. We can easily see that whereas there is minimal variation in their response, they did overwhelmingly choose the middle option, which might necessitate a special handling of their response in order to prevent any potential data contamination when estimating any model using

their responses. These are also potential causes for attenuated correlations between items as discussed above.

4.5 Bayesian Dynamic Factor Model

4.5.1 Model Specification

With a promising dataset in our hands, we propose modeling the observable behavioral outcome variables (if each participant smokes and if so how many cigarettes), whether or not each participant completes the EMA and their responses to the questions in each EMA jointly. We could consider participants as transitioning in terms of their latent states each time period, where the responses to the EMAs are noisy measurements of their latent states using dynamic factor models (Cunha et al. 2010, Agostinelli and Wiswall 2016, McAlinn et al. 2018). The observable behavioral outcomes would then be modeled as functions of these latent factors, potentially using Tobit-type models (Heckman 1974).

For example, we may model the joint data generating process as follows (adapted from McAlinn et al. 2018), where y_{it} is the number of cigarettes smoked between time $t-1$ and t by participant i at time t, y_{it}^{*} is the latent urge of participant i to smoke during the time window between t and $t-1$, δ_{it} is the willingness to respond to the EMA to, $\boldsymbol{\omega}_{it} \in \mathbb{R}^{P}$ is the latent state (factor) that affect the participants' urge to smoke and their willingness to complete EMA, $\boldsymbol{Y}_{it} \in \mathbb{R}^{\mathrm{K}}$ is the vector of questionnaire response of participant i at time t. $\boldsymbol{\gamma}_t$ is the time-varying coefficient relating the participants' urge to smoke and their latent state in the previous time period. $\boldsymbol{B}_t$ and $\boldsymbol{\Sigma}_t$ represent the time-varying relationship between the latent state of each individual and their questionnaire responses, and the correlation between the responses, respectively.

$$y_{it} = \begin{cases} Poisson(y_{it}^*), & y_{it}^* \geq 0 \; and \; \delta_{it} \geq 0 \\ 0, & y_{it}^* < 0 \; and \; \delta_{it} \geq 0 \\ Missing, & \delta_{it} < 0 \end{cases} \quad (1)$$

$$y_{it}^* = \boldsymbol{\gamma_t}\boldsymbol{\omega}_{i,t-1} + \xi_{it}, \; \xi_{it} \sim N(0, \sigma_\xi^2) \quad (2)$$

$$\boldsymbol{Y}_{it} = \begin{cases} \boldsymbol{B_t}\boldsymbol{\omega}_{it} + \boldsymbol{\eta}_{it}, \; \boldsymbol{\eta}_{it} \sim N(0, \boldsymbol{\Sigma}_t), \; \delta_{it} \geq 0 \\ Missing, \; \delta_{it} < 0 \end{cases} \quad (3)$$

$$\delta_{it} = \boldsymbol{\lambda}\boldsymbol{\omega}_{it} + \zeta_{it}, \; \zeta_{it} \sim N(0, \sigma_\zeta^2) \quad (4)$$

$$\boldsymbol{\omega}_{it} = \boldsymbol{\Phi}\boldsymbol{\omega}_{i,t-1} + \boldsymbol{\epsilon}_{it}, \; \boldsymbol{\epsilon}_{it} \sim N(0, \sigma_\epsilon^2 I) \quad (5)$$

We believe Bayesian estimation methods would be particularly useful in estimating the model. Rather than directly calculating the joint likelihood of all the observed data, which might require complex integrations over the unobserved quantities and computationally expensive approximations, a Bayesian approach allows us to draw posterior samples for these latent quantities conditional on the other parameters, which is much more feasible. Using a Bayesian approach would also help us overcome some of the issues in EMA implementations such as the "cold start" problem, as it can naturally pool information from the other participants and regularize the heterogeneous individual-level parameters so that EMA activations for each participant become less volatile during the initial stages where the model has had less opportunity to learn about them. As more data are collected for each user, EMA customizations can rely more on the observed outcomes and EMA responses of that particular user than on that of the overall population.

Nevertheless, we expect to face some challenges while modeling the data as above. For one, inferring the latent states of participants who do not respond to EMAs would be the biggest practical issue. Whether or not a participant is willing to complete the EMA (δ_{it}) would most likely depend on their unobserved latent state at that time ($\boldsymbol{\omega}_{it}$), which in turn would have to be

inferred based on their previous EMAs (and potentially future ones using the "transition" matrix $\boldsymbol{B_t}$) as we do not have responses to the current EMA ($\boldsymbol{Y}_{it}$). While this is theoretically possible through probabilistic imputations for these values, many quantities may be poorly identified if users' response rates are generally low or when a particular user goes through a long period of hiatus. This may be exacerbated by the fact that we are assuming many key quantities in the model such as $\boldsymbol{B_t}$ and $\boldsymbol{\Sigma}_t$ to be time-varying. These time-varying quantities would most likely be heavily informed by the latent states $\boldsymbol{\omega}_{it}$s, which also have uncertainties around them. While a sparse structure may be induced on them by assigning prior distributions such as the spike-and-slab priors (e.g., Ishwaran and Rao 2005, Rockova and McAlinn 2021), whether our dataset contains enough information to be able to achieve reasonable precision in estimation is an empirical question.

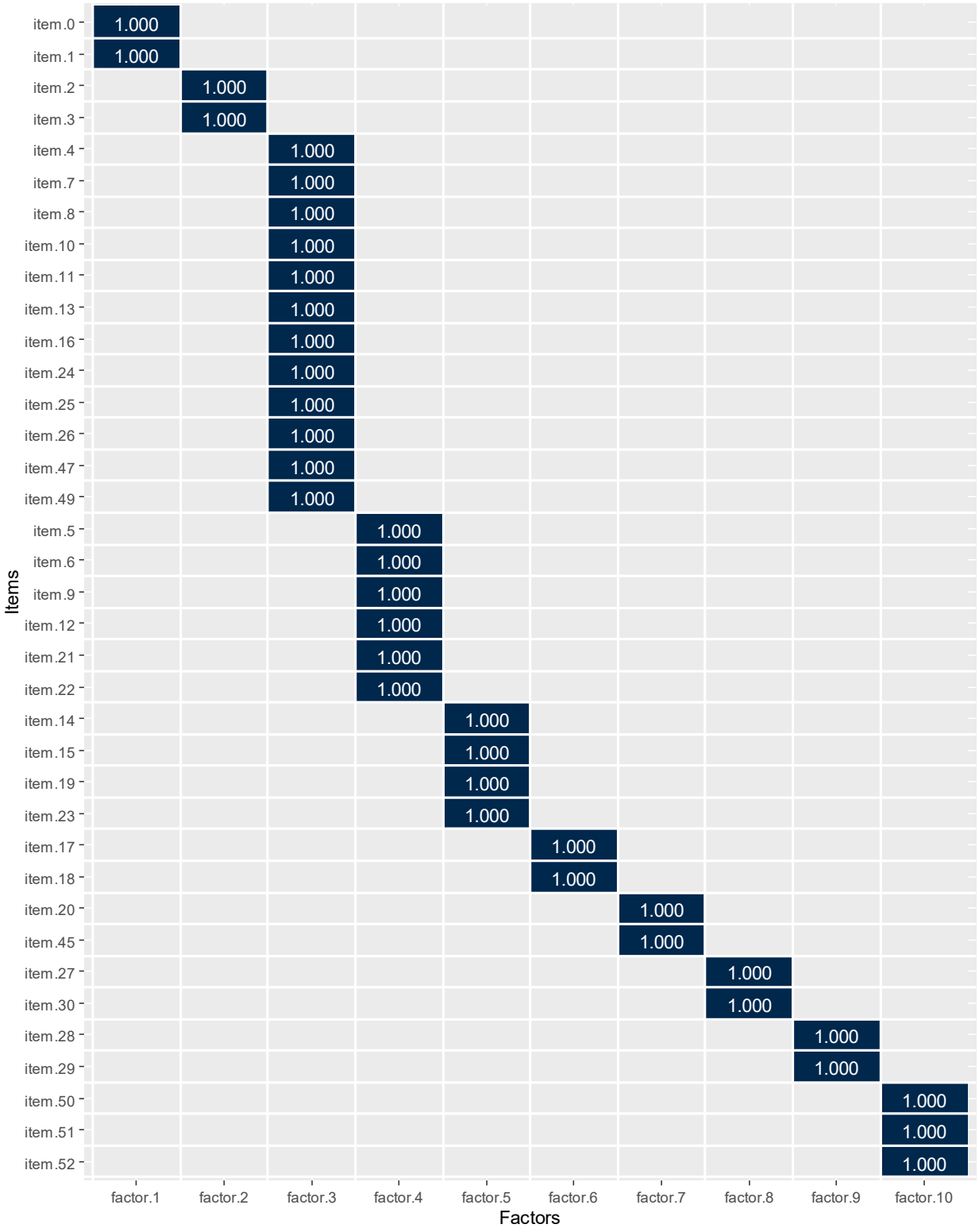
4.5.2 Preliminary Empirical Analyses

Before we estimate the model specified above, we first perform a Bayesian Exploratory Factor Analysis (Conti et al. 2014) that has been developed to reduce the dimensions of a dataset to produce readily interpretable factors by allowing each manifest variable (i.e., item) is allowed to load only on a single latent factor. The practical application of their algorithm on our dataset has been implemented via the `R` function `befa` of the `BayesFM` package (v0.1.5). We have mandated that *at least* 2 items load on each of the latent factors and have tried allowing varying numbers of maximum factors to balance the ease of interpretability and a more accurate representation of the factor structure. While we have also performed analogous analyses while limiting the number of maximum factors to 5 for better interpretability (found in Figure C1), here we present the results where we allow for at most 10 for an appropriate balance between being able to adequately capture the latent structure behind the items and reasonable interpretability.

The results shown below are post-processed using the **`post.column.switch`** and **`post.sign.switch`** functions within the same package so as to address the column and sign switching issues prevalent with similar models using MCMC procedures (Papastamoulis and Ntzoufras 2022). 20,000 posterior samples after a burn-in period of 20,000 iterations indicate that a 10-factor structure best explains the panel dataset at hand. Whereas multiple runs of the MCMC did converge to slightly different solutions in terms of the factor loadings of a few items, the overall structure remains consistent. Figure 4.4 depicts the posterior probabilities of each item belonging to each of the 10 latent factors based on the posterior samples from a representative run.

A more in-depth look at the 10 latent factors reveals that the structure of the factors seems to reflect well the preliminary correlations found in Table C1. For example, items 0 ("When the phone beeped, I was doing things automatically, without being aware of what I was doing.") and 1 ("When the phone beeped, I found myself doing things without paying attention."), as well as items 2 ("When the phone beeped, I was telling myself that I shouldn't be thinking the way I'm thinking.") and 3 ("When the phone beeped, I was thinking that some of my emotions are bad or inappropriate and shouldn't feel them.") were pairs with the highest positive correlations, which are estimated by the model to form their own respective latent factors. Items in the third latent factor mostly relate to positive affects (e.g., calm; item 7, grateful; item 11, happy; item 13, proud; item 16; relaxed; item 26), whereas the those in the fourth (e.g., angry; item 5, ashamed; item 6, and disgusted; item 9) and fifth (e.g., irritable; item 14, lonely; item 15, bored; item 23) factors represent negative ones. Similarly, the eighth and ninth latent factors directly capture desire for smoking abstinence (items 27 and 30) and the urge

Figure 4.4 Posterior Probabilities of Each Item Belonging to Each Latent Factor in a 10-Factor Structure



to smoke (items 28 and 29), respectively. Finally, the tenth latent factor (items 50 to 52) closely relates to spiritual awareness. The fact that these related items are loaded onto their respective latent factors *without* directly utilizing the any semantic meanings of each item through natural language processing or similar technique is both reassuring and promising.

Whereas the use of a pre-existing package makes the analyses much less time-consuming and foolproof, their framework does have some noteworthy restrictive assumptions that we would like to overcome in the future. Most prominently, the proposed model currently used is a "dedicated" factor model in which each item only loads on a single latent factor. However, it seems plausible that an item may be related to multiple constructs. For example, the item "When the phone beeped, I found myself doing things without paying attention." could be highly relevant to both restlessness and the urge to smoke, which might be captured by other latent factors. Another critical missing piece of the current model is the lack of a dynamic evolution of the factor structure; the current model assumes that the relationship between items stay constant over time, whereas it might not be the case in the current setting as participants becomes closer to successfully quitting smoking. Relatedly, the current preliminary model implicitly assumes that each EMA is exchangeable in the sense that it is agnostic about *which* participant it is from and *when*, whereas one of the key features that would allow us to identify any temporal changes in the factor structure is the panel nature of our dataset, with multiple EMA responses from each participant over time. Finally, the model does not account for the potential lack of inattentiveness of respondents – as evidenced by straight-line responding – and the heterogeneous use of scales across participants, which are a critical practical issue as discussed in the previous section. These issues would be addressed by our proposed model above, the estimation of which we leave for future work.

4.6 Future Directions and Discussion

Once we have the model estimation results, we would then proceed to the optimization implementation phase, in which we try optimizing when each participant receives EMAs and which questions are included in each. There is an important consideration to be made when determining the optimization function to base the optimization on. As alluded to in the introduction, EMAs not only serve as a means of collecting valuable information regarding individual participants' behavioral patterns as well as the predictive value of each information that that can be used for future optimization, but also as behavioral interventions that maintain participants' engagement with the smoking cessation program and ultimately assist with reducing cigarette consumption. Therefore, the use of an objective function that pertains to any one of these goals, or even a weighted average of each may be easily justified. Furthermore, it is also plausible that the weight of each objective might shift over time: during the early phases of the trial, learning about the participants might be the main goal, but over time, the expected retention of each participant, or reduced smoking frequency/amount may take precedence. It might be worth investigating whether the use of different objective functions leads to markedly divergent suggestions with respect to the frequency and content of EMAs for each individual.

There are two potential ways by which we can gauge the performance of our optimization. The first is to come up with model-based counterfactuals where we iteratively simulate users' evolution in terms of their latent states and observed behaviors over time. Then, we may compare the proximal and distal outcomes from those of the actual data and our simulated counterfactuals. The other is to do a field experiment, conditional on IRB approval, where we randomly choose some participants to be subject to the traditional EMA activation schedule and the others to be subject to the customized one. Then, we can compare the two

groups in terms of their signs of response fatigue such as dropout rates, the predictive performance of information obtained by the EMAs and the progress they make in terms of smoking cessation. One model validity check we may be able to perform at this stage is to see whether tailoring our intervention based on the suggested solution under one objective function (e.g., longer participant retention) does lead to a improvement in that dimension as opposed to others.

We believe overcoming potential challenges and successfully customizing EMA would bring great value to future researchers trying to implement them in various contexts, not just limited to smoking cessation programs on which we currently plan to work. Specifically, we believe our framework would contribute substantially to addressing the pressing issue of maintaining a sustained level of EMA compliance, which itself may be positively correlated with desired health outcomes as suggested in our data description. For one, customizing the EMA to gauge the current states of participants through shorter questionnaires without substantial loss of information, which may allow researchers to ping them more frequently without dramatically exacerbating their response fatigue and potentially compromising data quality.

The empirical context of EMAs and the main objective of this study are directly comparable to those of managing the accuracy of consumer panels, which have long been a topic of interest for marketing researchers (e.g., Sudman 1964). Whereas the main source of individual consumer behavior data have gradually shifted away from similar manual records due to the wider availability of scanner panel data (e.g., Guadagni and Little 1983) and the advent of digital marketing and resultant databases of consumer behavior (e.g., Ansari and Mela 2003), they still serve as a valuable means of gaining insights on individual-level consumption or detecting shifts in attitude towards various brands, especially in contexts where they may not be readily

accessible otherwise (e.g., Yang et al. 2010, Hansen et al. 2018, Joo et al. 2021). However, it has been well documented that attrition in such panels is not only quite prevalent (Golany et al. 1991) and occasionally can be severe (e.g., Farley et al. 1976), but can also bias model estimates (Winer 1983). Therefore, it is in the best interests of marketing researchers to prevent attrition among panels to the extent possible, and we believe our approach for mitigating response fatigue by adapting questionnaire frequency and content could be directly applicable to addressing such problems, particularly when long surveys are frequently distributed.

However, the basic principles obtained from the current project could also be extended to other marketing practices more broadly as well. There are many instances where firms constantly ask their customers for their opinions regarding their satisfaction with their products or services and brand perception (Bolton and Lemon 1999), which in turn allows firms to engage in proactive interventions targeted at each customer (de Matos et al. 2018). However, there have been reports that indiscriminate preemptive interventions may backfire in that they actually increase customer churn (e.g., Ascarza et al. 2016), indicating that too frequent interventions by marketers are not always optimal (Chae et al. 2019). Wohllebe et al. (2021) specifically demonstrate, based on a field experiment, that users who were given more frequent non-personalized push notifications from a mobile app exhibited on average a lower propensity to react to those notifications and a higher app uninstallation rate. While interactions with the customers can not only reveal what their opinions are towards the brand but also serve as reminders of their brand proposition, we believe that marketers need to practice discipline when engaging in such behaviors, reaching out to them only when doing so is warranted and in a highly customized manner. Our framework proposes a way that can facilitate such measured approaches, one in which we hope marketing practitioners would find value.

CHAPTER 5

Summary and Outlook

This dissertation consists of three essays that collectively contribute to the relatively limited marketing literature on estimating the temporal implications of online marketing interventions and optimizing them with non-immediate outcomes as the objective function. In the first essay, we investigate the capabilities of a novel consumption-decelerating promotion that we call "Wait For Free" applied on a serialized digital media platform. We then look for the optimal promotion set that is expected to maximize the platform-wide revenue using a combinatorial genetic algorithm. The second essay proposes a novel framework in which the non-immediate impact of marketing activities such as those of online banner advertisements can be accounted for in a more flexible manner compared to that of the traditional Nerlove-Arrow framework for ad stock. We apply a semiparametric specification of the ad stock on an empirical setting of online display ads and perform a simulation study in which we show that an even more flexible nonparametric formulation using Gaussian Process priors are able to recover the latent decay functions accurately. The third essay explores the possibility of optimizing the frequency and content of Ecological Momentary Assessments in order to minimize response fatigue and increase retention in mobile health contexts. We propose a Bayesian dynamic factor model that would allow us to reduce the number of items included within each prompt and the frequency of EMAs by putting previous responses of each participant into consideration.

While we believe those we have already addressed in this work are important questions the answers to which will not only benefit marketing practitioners but also academics, we do acknowledge there remain further questions that are managerially and theoretically relevant and warrant future investigation. For example, with regards to the first essay where we have mostly assumed to be exogenous some critical configurations of the Wait For Free promotion such as the length of the waiting period to unlock the next episode and the number of free episodes offered in the beginning, a more thorough investigation of how to better operationalize these for maximal user engagement is an interesting avenue for future work. We believe lab or field experiments where we directly offer an incentive-compatible tradeoff between waiting for free consumption and paying for immediate consumption would be extremely useful in resolving such issues. A more extensive inquiry into the optimal schedule of new episode release especially in the serialized digital media context, not just the optimal monetization of already released episodes, would also be highly relevant for marketing managers. We also postulate that an extension of our framework in the third essay into a more general marketing context could also reveal new insights into how consumers and firms can maintain a long-term relationship. Whereas most participants in a smoking cessation program are aware that the main beneficiary of remaining in such programs are primarily themselves, many consumers do not possess an analogous view regarding their interactions with companies. As such, whereas the fundamental insights from our mobile health context might offer a general guiding principle for marketers who would like to foster a communications channel in which customer feedback occurs regularly as in customer satisfaction surveys, some of the findings might not directly apply. To what extent these two settings diverge would be of interest to customer relationship managers and salespeople alike.

APPENDICES

APPENDIX A

Appendix to Chapter 2

Table A1

WFF Adoption Dates and Number of Comics for Each Wave

Wave Number	WFF Adoption Date	Number of Comics
1	November 6th, 2018	12
2	November 14th, 2018	2
3	November 21st, 2018	2
4	November 28th, 2018	2
5	December 5th, 2018	2
6	December 12th, 2018	2
7	December 19th, 2018	8
8	January 30th, 2019	5
9	March 6th, 2019	5
10	March 27th, 2019	7
11	April 3rd, 2019	5
12	May 1st, 2019	7

Table A2

Variable Operationalization for Episode and Comic Characteristics

Variable	Operationalization
Episode Position	Ratio of the episode number for the focal episode divided by the maximum episode number for the comic series.
Consumption Clumpiness	Proportion of the number of episodes the average user of the series views on a single day to the total number of episodes in the series.
No. Episodes	Log of length of the comic series as defined by the number of episodes within the series.
Weeks since Completion	Log of the number of weeks since the last episode of the comic has been uploaded to the platform. 0 for incomplete comics.
Conversion Rate	Proportion of users who have paid for at least one episode out of all users who have viewed at least one of the comic's episode over the 19-week period before the first adoption of WFF on the platform.
Price	Log of the per-episode price as in coins.
Popularity	Log of the median weekly total (free + paid) views for each comic over the 19-week period before the first adoption of WFF on the platform.

Table A3

Logistic Regression Results for PSM and IPW

	Estimate	Std. Error	Z	P
Intercept	-12.944	2.159	-6.00	0.0000
Consumption Clumpiness	0.218	0.454	0.48	0.6312
No. Episodes	0.754	0.441	1.71	0.0874
Weeks since Completion	0.499	0.141	3.53	0.0004
Conversion Rate	0.301	0.198	1.52	0.1277
Price	1.346	0.839	1.61	0.1084
Popularity	0.584	0.131	4.47	0.0000

Table A4

Comparison of Comic-Level Promotional Effect Estimates using Random Effects vs. Fixed Effects for α_c and α_t

		No Correction	PSM	CEM	IPW
Random Effects	$\hat{\beta}_F$	3.303 (0.071)	3.353 (0.089)	3.347 (0.097)	3.544 (0.030)
	$\hat{\beta}_P$	0.797 (0.043)	0.807 (0.044)	0.746 (0.042)	0.836 (0.016)
Fixed Effects	$\hat{\beta}_F$	3.291 (0.071)	3.446 (0.092)	3.450 (0.101)	3.550 (0.030)
	$\hat{\beta}_P$	0.787 (0.043)	0.840 (0.045)	0.781 (0.043)	0.839 (0.016)

Table A5

Promotional Effect Estimates using 10 Random Seeds for Match Selection during CEM

	1	2	3	4	5	6	7	8	9	10
$\hat{\beta}_P$	3.347	3.427	3.411	3.218	3.395	3.411	3.332	3.502	3.261	3.291
	(0.097)	(0.106)	(0.103)	(0.100)	(0.103)	(0.101)	(0.102)	(0.106)	(0.100)	(0.106)
$\hat{\beta}_F$	0.746	0.745	0.815	0.779	0.858	0.861	0.737	0.786	0.730	0.756
	(0.042)	(0.046)	(0.047)	(0.049)	(0.046)	(0.047)	(0.048)	(0.046)	(0.043)	(0.046)

Table A6

Summary Statistics of Comic Attributes for WFF and Non-WFF Comics, Before and After CEM

	Before CEM		After CEM	
	WFF	Non-WFF	WFF	Non-WFF
Consumption Clumpiness	-2.392 (0.462)	-2.351 (0.784)	-2.425 (0.444)	-2.610 (0.737)
No. Episodes	4.148 (0.378)	3.768 (0.765)	4.162 (0.384)	4.281 (0.606)
Weeks since Completion	3.436 (1.385)	2.542 (1.909)	3.443 (1.414)	3.243 (1.707)
Conversion Rate	-1.314 (0.916)	-1.369 (1.012)	-1.415 (0.837)	-1.483 (1.167)
Price	1.378 (0.049)	1.380 (0.233)	1.378 (0.050)	1.378 (0.050)
Popularity	8.226 (1.288)	7.071 (1.677)	8.143 (1.278)	8.110 (1.292)

Table A7

Estimates of Promotional Effect for Paying vs. Non-Paying Users Using Various Endogeneity Correction Methods

	Free		Paid	
	Paying	Non-Paying	Paying	Non-Paying
No Correction	3.581 (0.070)	3.201 (0.075)	0.731 (0.044)	2.260 (0.068)
PSM	3.647 (0.092)	3.345 (0.097)	0.789 (0.046)	1.419 (0.082)
CEM	3.707 (0.099)	3.275 (0.105)	0.744 (0.044)	1.358 (0.084)
C&S (No Covariates) *	3.965 (0.248)	3.649 (0.280)	0.817 (0.103)	1.917 (0.232)
C&S (6 Covariates)	4.060 (0.229)	3.771 (0.277)	0.821 (0.088)	1.477 (0.220)
IPW	3.846 (0.030)	3.410 (0.032)	0.793 (0.016)	1.916 (0.026)

* Callaway and Sant'Anna (2021)

Table A8

Estimates of Promotional Effect for Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods

	Free		Paid	
	Frequent	Infrequent	Frequent	Infrequent
No Correction	3.661 (0.073)	2.286 (0.064)	0.808 (0.044)	0.353 (0.072)
PSM	3.740 (0.095)	2.484 (0.081)	0.837 (0.047)	0.432 (0.068)
CEM	3.782 (0.103)	2.431 (0.086)	0.807 (0.045)	0.368 (0.068)
C&S (No Covariates) *	4.080 (0.253)	2.713 (0.236)	0.893 (0.078)	0.695 (0.195)
C&S (6 Covariates)	4.167 (0.241)	2.854 (0.209)	0.887 (0.072)	0.745 (0.195)
IPW	3.917 (0.031)	2.580 (0.026)	0.861 (0.016)	0.466 (0.025)

* Callaway and Sant'Anna (2021)

Table A9

Estimates of Promotional Effect for Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods, Using 5 Visit Days as Cutoff for Frequent User

While we chose 10 distinct visits days during the year preceding platform's adoption of WFF as the cutoff for defining "Frequent" visitors in our main analyses, we here provide a robustness check where we use 5 as the cutoff, which leads to 1,246,603 (42.8%) and 1,664,440 (57.2%) users being classified into Frequent and Infrequent ones, respectively. It is apparent that the overall pattern of the results remains the same.

	Free		Paid	
	Frequent	Infrequent	Frequent	Infrequent
No Correction	3.568 (0.072)	2.167 (0.063)	0.780 (0.044)	0.465 (0.075)
PSM	3.656 (0.094)	2.349 (0.080)	0.812 (0.046)	0.627 (0.075)
CEM	3.689 (0.101)	2.294 (0.085)	0.783 (0.044)	0.616 (0.079)
C&S (No Covariates) *	3.985 (0.268)	2.534 (0.240)	0.874 (0.086)	0.624 (0.224)
C&S (6 Covariates)	4.080 (0.244)	2.679 (0.214)	0.871 (0.083)	0.654 (0.215)
IPW	3.828 (0.031)	2.458 (0.026)	0.833 (0.016)	0.648 (0.028)

* Callaway and Sant'Anna (2021)

Table A10

Estimates of Promotional Effect for Frequent (Cont.) vs. Infrequent (Cont.) Users Using Various Endogeneity Correction Methods

Given that our analyses regarding the Frequent vs. Infrequent visitors could be affected by the fact that many users, particularly those in the latter group, never return to the platform post-WFF, we present results limiting the scope to users who view at least one episode post-WFF, resulting in 604,392 Frequent (67.7%) and 288,896 Infrequent (32.3%) users, respectively.

	Free		Paid	
	Frequent (Cont.)	Infrequent (Cont.)	Frequent (Cont.)	Infrequent (Cont.)
No Correction	3.684 (0.073)	2.484 (0.062)	0.832 (0.044)	0.566 (0.073)
PSM	3.729 (0.094)	2.704 (0.074)	0.845 (0.047)	0.721 (0.081)
CEM	3.778 (0.102)	2.651 (0.082)	0.808 (0.045)	0.593 (0.087)
C&S (No Covariates) *	4.086 (0.257)	2.797 (0.239)	0.896 (0.080)	0.890 (0.219)
C&S (6 Covariates)	4.172 (0.248)	2.915 (0.224)	0.888 (0.070)	0.916 (0.229)
IPW	3.927 (0.031)	2.715 (0.026)	0.871 (0.016)	0.628 (0.027)

* Callaway and Sant'Anna (2021)

Table A11

Estimates of Promotional Effect for Paying vs. Non-Paying and Frequent vs. Infrequent Users Using Various Endogeneity Correction Methods

The method we use for this analysis is analogous to what we use for a unidimensional analysis except that we are constructing four distinct panel structure dataset for each group of users. The breakdown of users into each segment is as follows: 515,017 Paying Frequent users (17.69%), 307,051 Paying Infrequent users (10.55%), 399,018 Non-Paying Frequent users (13.71%), and 1,689,957 Non-Paying Infrequent users (58.05%).

Tables below show the estimated promotional effect of WFF of different user groups on free and paid viewership, respectively. Overall patterns we observe are mostly consistent with the previous analyses: WFF is more effective in boosting free viewership for Paying vs. Non-Paying users but the opposite is true for paid viewership. Frequent users in general react more strongly than Infrequent users to WFF both in terms of free and paid views, although the difference is statistically insignificant within the Non-Paying population with regards to paid viewership.

	Free			
	Paying Frequent	Paying Infrequent	Non-Paying Frequent	Non-Paying Infrequent
No Correction	3.707 (0.072)	2.209 (0.062)	3.641 (0.081)	2.375 (0.066)
PSM	3.763 (0.094)	2.359 (0.078)	3.757 (0.104)	2.577 (0.084)
CEM	3.835 (0.101)	2.360 (0.084)	3.715 (0.113)	2.498 (0.091)
C&S (No Covariates) *	4.097 (0.241)	2.651 (0.221)	4.244 (0.301)	2.781 (0.248)
C&S (6 Covariates)	4.179 (0.255)	2.813 (0.198)	4.342 (0.295)	2.908 (0.241)
IPW	3.965 (0.031)	2.514 (0.025)	3.819 (0.034)	2.634 (0.027)

* Callaway and Sant'Anna (2021)

	Paid			
	Paying Frequent	Paying Infrequent	Non-Paying Frequent	Non-Paying Infrequent
No Correction	0.799 (0.044)	0.288 (0.073)	1.808 (0.054)	2.089 (0.066)
PSM	0.837 (0.047)	0.470 (0.078)	1.181 (0.077)	1.289 (0.084)
CEM	0.801 (0.045)	0.426 (0.081)	1.491 (0.076)	1.207 (0.088)
C&S (No Covariates) *	0.884 (0.083)	0.630 (0.186)	1.616 (0.211)	1.789 (0.234)
C&S (6 Covariates)	0.883 (0.080)	0.711 (0.200)	1.334 (0.217)	1.346 (0.211)
IPW	0.852 (0.016)	0.376 (0.027)	1.541 (0.021)	1.741 (0.026)

* Callaway and Sant'Anna (2021)

Table A12

Lags and Correlated Random Effects in Episode-Level Analyses

Other controls for the episode-level model include comic-specific random intercepts $\alpha_{cf}^{intercept}$, to account for remaining unobserved characteristics of each comic; week-lags of the number of free views of focal episode, $\log(Y_{cet-1,f}+1)$, to capture any consumption inertia engendered by previous consumption and correct for any episode-level idiosyncrasies uncorrected for by the comic intercepts; week-specific random intercepts α_{tf}, to flexibly account for overall time trends; and, lastly, other comic-level random effects $\alpha_{cf}^{new_episode}$, to correct for the lack of week-lags for newly published episodes, and $\alpha_{cf}^{first_freed}$, to account for existing paid-only episodes becoming available for free to all users. All four random effects are regularized using Gaussian kernels; critically, the comic-level and week-level random effects are each correlated between the free and paid models (e.g., α_c^F with α_c^P and α_t^F with α_t^P), to allow for underlying comic characteristics or the overall time trend to potentially affect the free and paid viewership in a correlated manner. Corrections for Underpredictions and Overpredictions only occur within the free model, not necessitating any correlations.

Random Effects Estimates			
	σ Free	σ Paid	Corr. (ρ)
Comic Level			
Intercept	0.129	0.362	-0.115
Episode Number	0.145	0.249	-0.580
Episode First Free	1.927	1.192	0.037
New Episode	1.511	1.121	0.855
Week Level			
Intercept	0.054	0.088	0.403
Outliers			
Underpredict	1.017	---	---
Overpredict	0.829		
Residual	0.496		

Table A13

WFF Effect Time Trends Parameter Estimates and Model Fit Measures using Various Model Specifications for Episode-level Analyses

	Comic-Level Time Trend Only		Comic-Level + Episode-Level Time Trends (Eqn. 2)		Both Time Trends + Spillovers & Heterogeneity (Eqn. 3)	
Comic-Level Trend						
	Free	Paid	Free	Paid	Free	Paid
Week 1	4.162 (0.009)	1.070 (0.009)	6.674 (0.013)	0.987 (0.013)	6.194 (0.041)	1.100 (0.040)
Week 2	1.403 (0.008)	0.113 (0.007)	4.721 (0.015)	0.023 (0.017)	4.221 (0.042)	0.046 (0.043)
Week 3	0.939 (0.006)	0.131 (0.006)	4.326 (0.015)	-0.047 (0.017)	3.797 (0.042)	-0.062 (0.043)
Week 4	0.726 (0.005)	0.122 (0.005)	4.239 (0.015)	-0.032 (0.017)	3.693 (0.042)	-0.071 (0.043)
Week 5	0.594 (0.006)	0.114 (0.005)	4.185 (0.015)	-0.034 (0.017)	3.645 (0.043)	-0.077 (0.043)
Week 6	0.536 (0.005)	0.255 (0.005)	4.158 (0.015)	0.118 (0.016)	3.609 (0.043)	0.056 (0.043)
Week 7	0.504 (0.005)	0.079 (0.005)	4.140 (0.015)	-0.037 (0.016)	3.588 (0.043)	-0.115 (0.043)
Week 8	0.456 (0.005)	0.157 (0.005)	4.145 (0.015)	0.045 (0.016)	3.585 (0.043)	-0.052 (0.044)
Week 9	0.453 (0.005)	0.109 (0.005)	4.152 (0.015)	0.004 (0.016)	3.584 (0.044)	-0.112 (0.044)
Week 10	0.431 (0.005)	0.137 (0.005)	4.161 (0.015)	0.020 (0.017)	3.587 (0.044)	-0.118 (0.045)
Week 11+	0.399 (0.003)	0.098 (0.003)	4.200 (0.015)	0.025 (0.016)	3.614 (0.045)	-0.150 (0.045)
Episode-level Trend						
	Free	Paid	Free	Paid	Free	Paid
Week 1			0 (fixed)	0 (fixed)	0 (fixed)	0 (fixed)
Week 2			-0.935 (0.017)	-0.266 (0.021)	-0.902 (0.017)	-0.216 (0.021)
Week 3			-3.730 (0.015)	0.203 (0.016)	-3.710 (0.015)	0.312 (0.017)
Week 4			-3.867 (0.015)	0.149 (0.016)	-3.833 (0.015)	0.279 (0.017)
Week 5			-3.866 (0.015)	0.145 (0.016)	-3.824 (0.015)	0.291 (0.017)
Week 6			-3.879 (0.015)	0.131 (0.016)	-3.828 (0.015)	0.297 (0.017)
Week 7			-3.872 (0.015)	0.090 (0.016)	-3.814 (0.015)	0.276 (0.017)
Week 8			-3.874 (0.015)	0.097 (0.017)	-3.808 (0.015)	0.302 (0.017)
Week 9			-3.877 (0.015)	0.091 (0.017)	-3.805 (0.016)	0.316 (0.017)
Week 10			-3.902 (0.015)	0.128 (0.017)	-3.821 (0.016)	0.379 (0.018)
Week 11+			-3.937 (0.015)	0.037 (0.017)	-3.868 (0.016)	0.308 (0.018)
Model Fit Measures						
Log-likelihood	-2,841,452		-2,768,650		**-2,762,740**	
AIC	5,683,080		5,537,516		**5,525,759**	
BIC	5,684,216		5,538,910		**5,527,566**	
Conditional R^2	0.9181		0.9228		**0.9229**	

Table A14

Observed Heterogeneity in WFF Effects across Genres

All coefficients below represent the genre-specific observed heterogeneity in the promotional lift from WFF, relative to the "Others" genre, which represents all genres not represented in the table. For effects estimation, a comic belonging to multiple genres was taken to lie equally across them, e.g., a comic labeled "[adult, fantasy]" half belongs to "Adult" and half to "Fantasy."

	Free	Paid
Adult	0.020 (0.040)	-0.048 (0.040)
Romance	0.004 (0.034)	-0.145 (0.034)
BL	-0.074 (0.036)	-0.097 (0.036)
Drama	-0.222 (0.036)	-0.081 (0.036)
School	0.006 (0.039)	0.260 (0.039)
Day	-0.150 (0.034)	-0.035 (0.034)
Fantasy	-0.116 (0.047)	-0.174 (0.047)
Mystery	-0.003 (0.041)	0.121 (0.041)

Table A15

Set of Rules Applied to WFF Structure of Each Comic for Optimization

Number of initial free episodes	We retain the number of free episodes for each comic on the week prior to the platform-wide adoption of the promotion. If there were no such episodes, this is set at 1 to maintain the core structure of the promotion.
Number of pay-to-access-only episodes at the end	We round 10% of the number of remaining episodes to the nearest integer.
Time to unlock the subsequent episode	We fix it at 1 day for all comics; this appears reasonable, considering that over 60% of WFF comics adopt this timing specification and there is little empirical basis for optimizing this quantity.

Table A16

Average and Median Characteristics of Optimal Promotion Set
vs. Platform-chosen Promotion Set

	Optimization Solution Promotion Set	**Platform-Chosen Promotion Set**
Total Views Average	307,229	89,004
Total Views Median	267,899	58,878
Total Views SD	207,333	105,318
Pop Rank Average	14	112
Pop Rank Median	7	55
Pop Rank Min	1	4
Pop Rank Max	67	343
ADULT	29%	25%
BL	21%	21%
DAY	0%	0%
DRAMA	0%	13%
FANTASY	0%	0%
MYSTERY	4%	4%
ROMANCE	21%	13%
SCHOOL	8%	8%
OTHERS	17%	17%
Consumption Clumpiness	-2.547 (0.557)	-2.249 (0.540)
No. Episodes	4.467 (0.392)	4.000 (0.487)
Weeks since Completion	2.736 (1.792)	3.558 (1.374)
Conversion Rate	-0.671 (1.028)	-1.153 (1.017)
Price	1.386 (0.000)	1.362 (0.083)
Popularity	9.925 (0.814)	8.381 (1.594)

Table A17

Optimized Comic Characteristics, Revenue vs. No. Paid Views

The Tables below compare the two sets of comics identified as optimal under the Revenues (left) and Number of Paid Views (right) as objective functions for the 16-week time horizon. Results suggest the optimized solution set is robust to the choice of the objective function, which is not surprising given the minimal variation in prices across comics.

Revenue		
Comic	Genres	Price
A	[bl, school]	3
B	[bl, romance]	3
C	[adult, bl]	3
D	[adult, romance]	3
E	[adult, bl]	3
F	[romance, other]	3
G	[mystery, other]	3
H	[school, other]	3
I	[adult]	3
J	[adult, other]	3
K	[adult, bl]	3
L	[romance]	3

No. Paid Views		
Comic	Genres	Price
A	[bl, school]	3
B	[bl, romance]	3
C	[adult, bl]	3
D	[adult, romance]	3
E	[adult, bl]	3
F	[romance, other]	3
G	[mystery, other]	3
H	[school, other]	3
I	[adult]	3
J	[adult, other]	3
K	[adult, bl]	3
L	[romance]	3

Table A18

Cox Proportional Hazards Negative Sampling Simulation Study

We conducted a simulation study in order to verify whether our approach of negative sampling and reweighting yielded mostly similar results to when using the full dataset. Specifically, we performed negative sampling where we kept roughly 10% of the zero outcomes and assigned them a weight of 1/0.1 = 10, and estimated a model with one covariate (Promoted) and also one random effect (comic intercepts). We can easily see that our approach seems to be effective; key parameter estimates and the log-likelihood are all reasonably similar when using the two approaches, with the negative sampling approach taking substantially less time for estimation.

Data	# Nonzero Outcomes	# Zero Outcomes	Estimation Time	Coefficient Estimate	Coefficient Std. Error	Random Effect SD	Log-likelihood
Full	79,455	47,160,951	27.8 mins	2.294	0.016	2.178	-1,286,677
Sampled	79,455	4,715,563	3.3 mins	2.313	0.016	2.175	-1,286,649

Figure A1

New Users Per Week Over Data Window

Figure AF01 depicts the number of new active users (users who were first observed to view any comics) to the platform each week, with a clear downtrend seemingly unaffected by the adoption of WFF (red dashed lines); we do not observe strong evidence that the promotion has a sizable role in bringing new users to the platform. More formally, a Regression Discontinuity analysis using the **`rdrobust`** package (v1.0.8) in **`R`** is ns non-significance (ps > 0.3) for both the conventional and robust detection of a time trend change at the time of WFF adoption to the platform. Given this lack of observable variation, we find no strong reasoning for modeling the impact of the promotion on the size of the user base, although we stop short of claiming that this effect is uniformly zero.

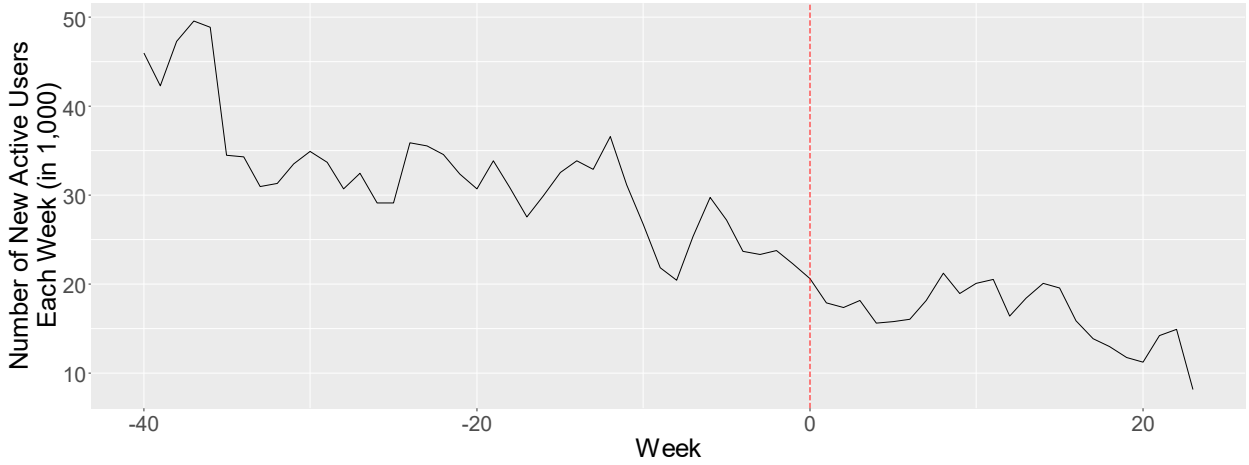


Figure A2

Time Trend of Promotional Lift of WFF on Free and Paid Viewership
(All Endogeneity Correction Methods)

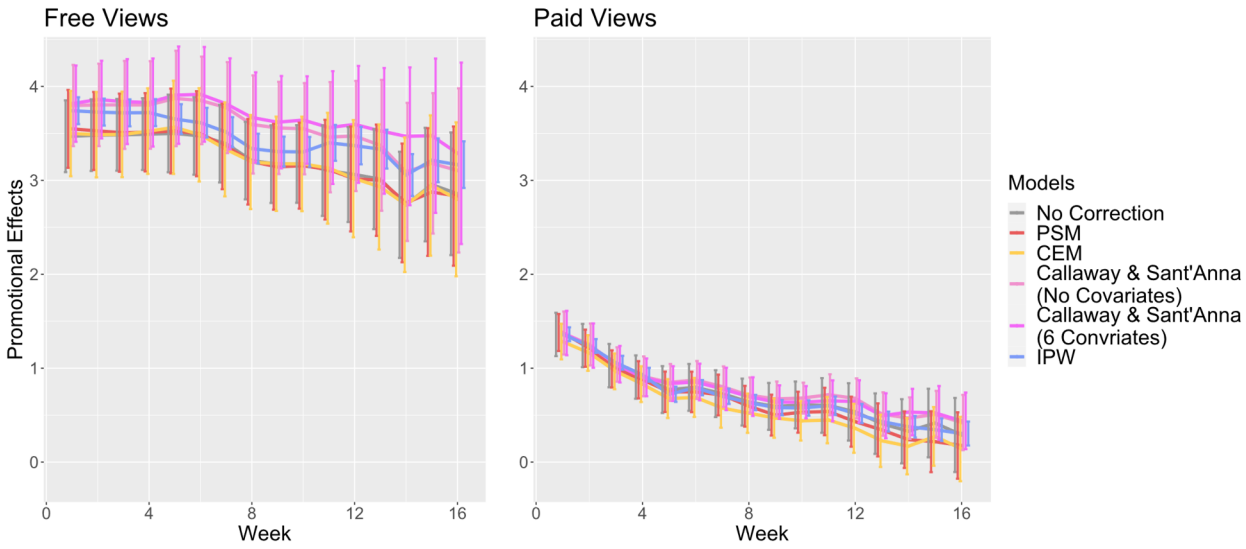


Figure A3

Distribution of Propensity Scores for WFF and Non-WFF Comics, Before and After PSM

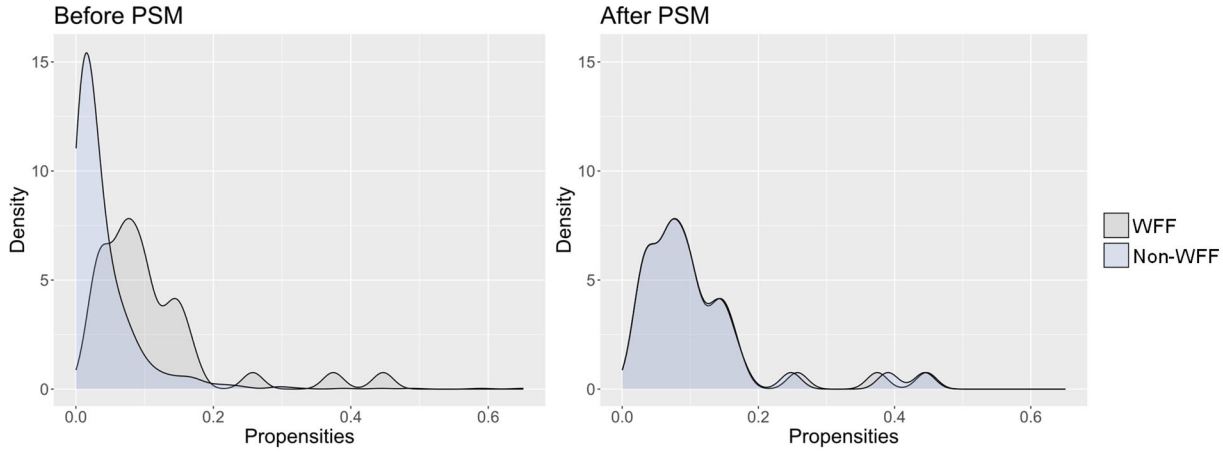


Figure A4

Robustness Check: Promotional Effects, IPW-Weighted (solid) vs. No Correction (dashed)

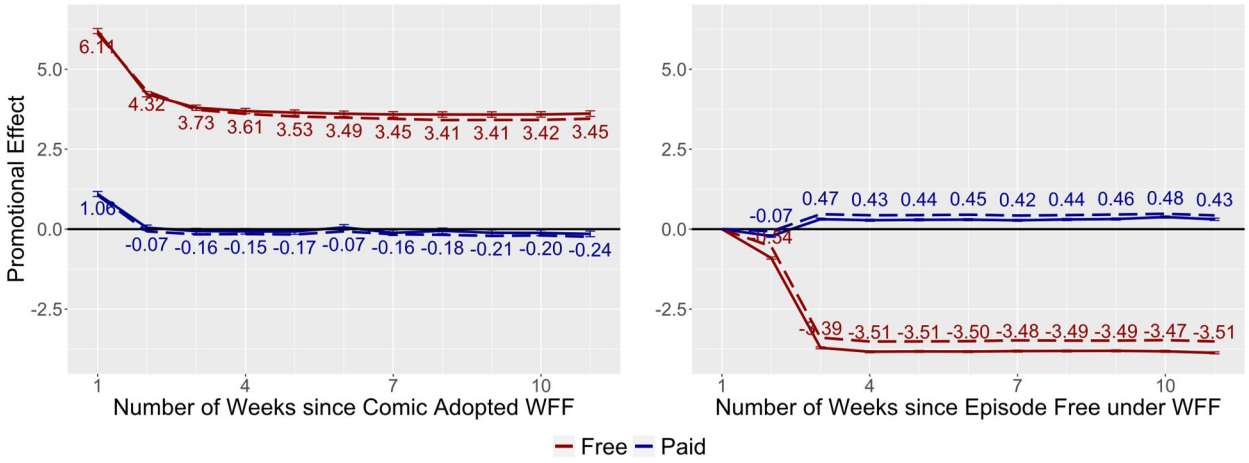


Figure A5

Robustness Check: Promotional Effects, Free and Paid Models

Estimated Jointly (solid) vs. Separately (dashed)

We also estimated the free and paid models independently, maintaining the correlation of the random effects within each model, but dropping those between models (i.e., $\rho_c \equiv 0$, $\rho_t \equiv 0$). We observe that the coefficient estimates are basically identical when the free and paid models are estimated separately vs. jointly.

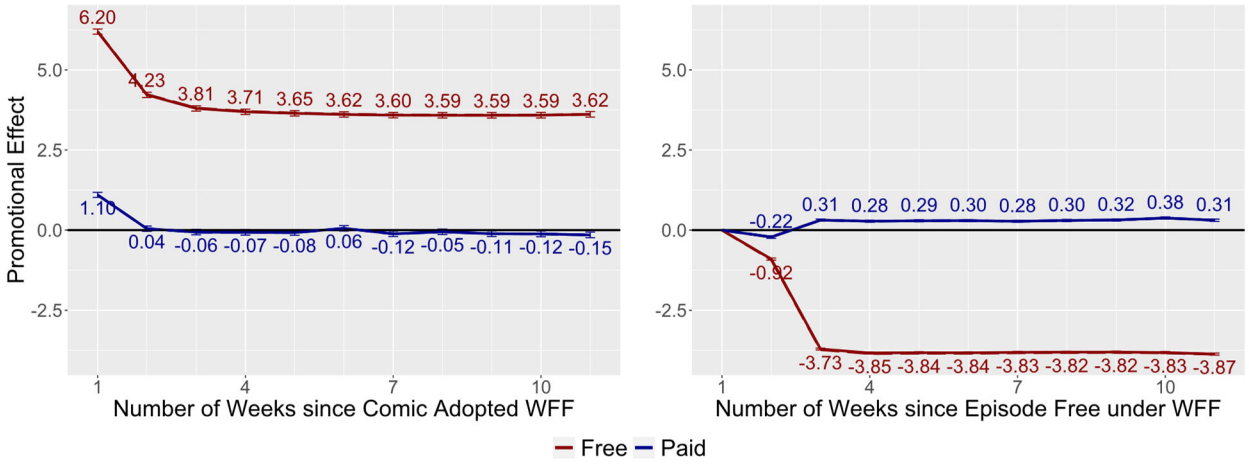


Figure A6

Robustness Check: Promotional Effects, None (solid) vs. Error Correlation (dashed)

The model implicitly assumes that the Free and Paid errors (ϵ_{cetF} and ϵ_{cetP}) for a given episode on a given week are not necessarily correlated. Searching the model space using a model allowing for such a correlation is prohibitively computationally costly. In our final model (M), we observe a substantively minor residual correlation, $\hat{\rho} \approx 0.07$, suggesting that the Free and Paid accounts operate largely independently after accounting for covariates. Nevertheless, as a robustness check, we include the cross-residuals as covariates to assess how model implications change. This figure depicts the strong parallelism in promotional effects with vs. without error correlation.



Figure A7

Robustness Check: Promotional Effects, 800 (solid) vs. 602 (dashed) Comics Included

We also checked robustness to the analyst-based decision to include the top 800 comics (including all WFF comics) in our analysis, accounting for roughly 93% of platform-wide revenue for the year-long period before WFF was introduced. Specifically, we repeated our analysis using the top 600 comics along with the 2 WFF comics that were not among them; these 602 comics account for roughly 90% of revenues. It is apparent that the results are hardly affected by this change.

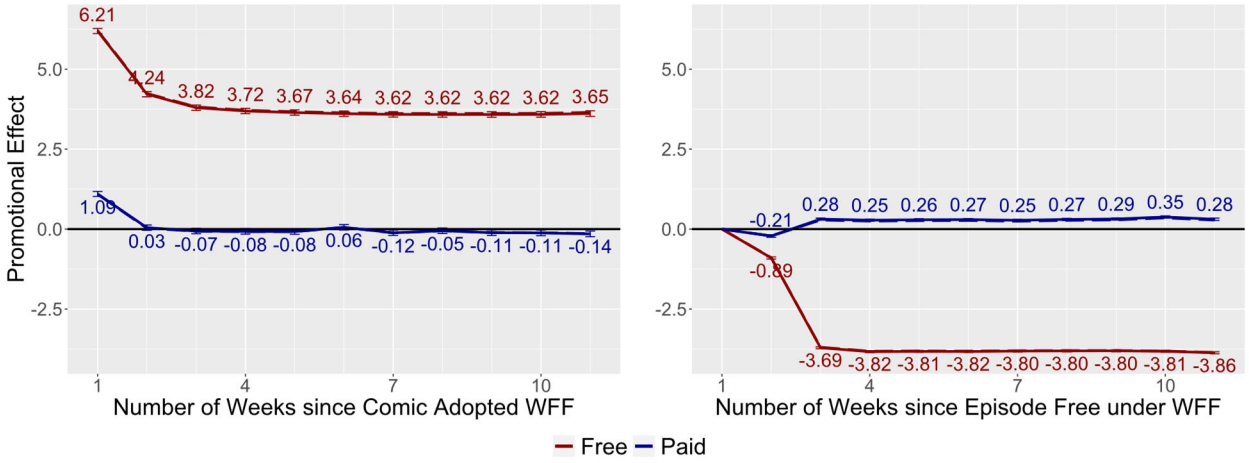


Figure A8

Robustness Check: Promotional Effects, With (solid) vs. Without (dashed) Outlier Correction

Despite the models' incorporation of all available data on promotions and other episode-specific information, a small number of outliers in the free model seemed to manifest inexplicable spikes or dips in viewership in certain weeks. Specifically, there were 1,912 under-predictions (0.13%) and 2,274 over-predictions (0.15%) out of 1,489,609 entries in the "free" model that were accounted for using comic-specific random effects. Instead of performing ad hoc removals of these observations, we flagged them by first estimating the model as is, noting whether the residuals were excessively positive or negative, and then including separate Gaussian-regularized (i.e., so-called "random effects") comic-specific correction terms for under- and over-predictions. A comparison of estimated promotional effects demonstrates that this procedure leads to only small differences for the free and virtually none for paid views model.

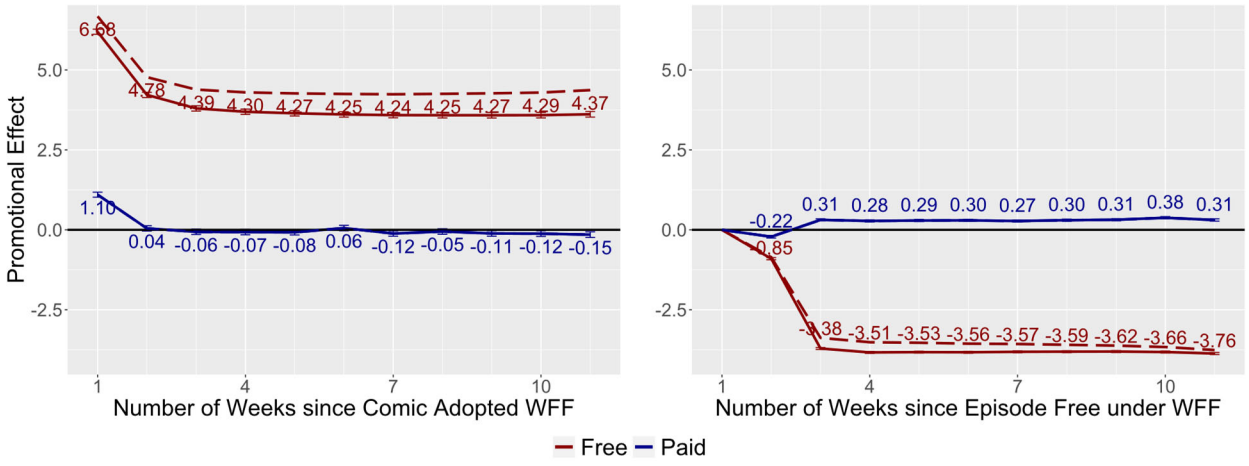
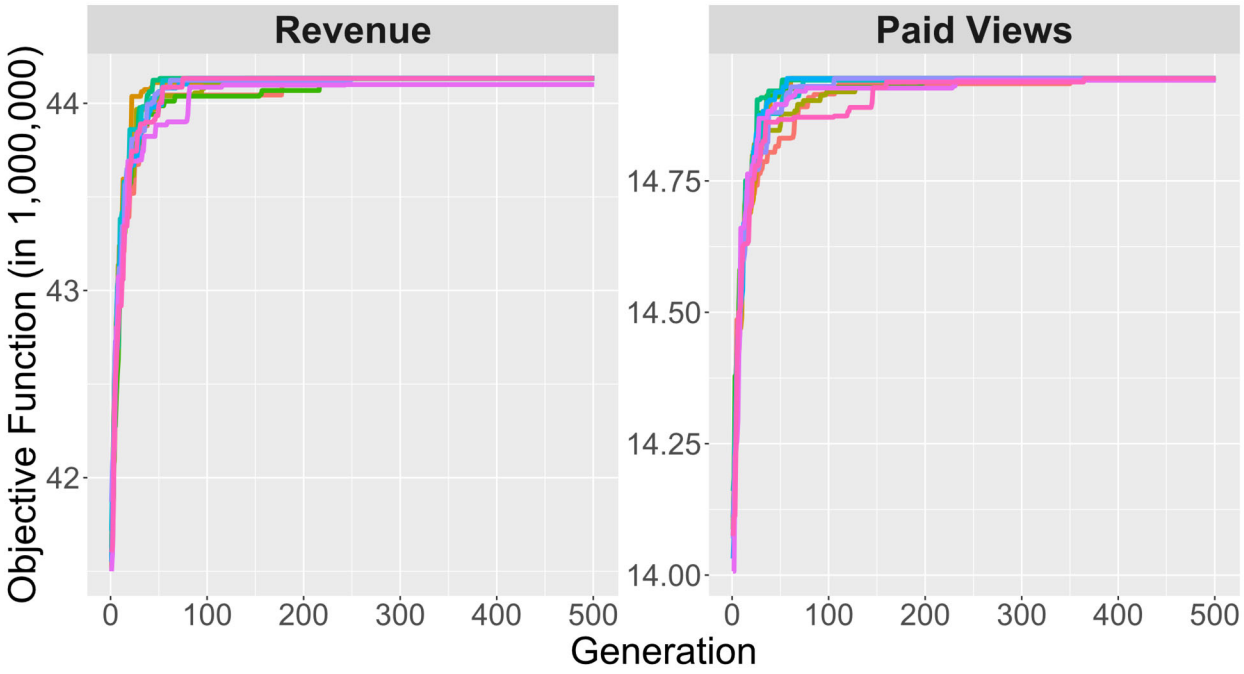


Figure A9

Revenue and Paid Views Convergence Plots for Genetic Algorithm Optimization



APPENDIX B

Appendix to Chapter 3

Table B1

Estimation Time for 6 MCMC Chains in the Empirical Analysis

	Warmup (400)	Sampling (400)	Total (800)
Chain 1	49,936.8	41,392.1	91,328.9
Chain 2	52,603.1	40,414.8	93,017.9
Chain 3	52,081.5	40,484.6	92,566.1
Chain 4	43,707.0	22,957.7	66,664.7
Chain 5	49,763.7	41,342.2	91,105.9
Chain 6	47,037.2	42,128.9	89,166.1
Average	49,188.2	38,120.1	87,308.3

Table B2

Effective Sample Sizes and Rhats for Key Parameters in the Empirical Analysis

Parameter	Effective Sample Size	Rhat
γ_0	1,484.24	1.00
γ_{r}	1,602.29	1.00
$\bar{\mu}$	949.66	1.01
σ_{μ}	446.23	1.01
$\bar{\beta}_{Ad}$	2,053.63	1.00
σ_{Ad}	380.42	1.01

Table B3

List of Decay Functions Used in the Simulation Study

Creative	Decay Functions
Creative 1	$f_1(t) = \frac{t(t+15)(t-105)}{250000} + 1$
Creative 2	$f_2(t) = 0.95^t$
Creative 3	$f_3(t) = \frac{1}{1 + exp\big(0.15(t-37.5)\big)}$
Creative 4	$f_4(t) = \frac{t(t+20)(t-115)}{300{,}000} + 1$
Creative 5	$f_5(t) = 0.5 \times 0.9995^t$
Creative 6	$f_6(t) = \frac{0.3}{1 + exp\ (0.2(t-37.5))}$
Creative 7	$f_7(t) = 1.3 - 0.015t$
Creative 8	$f_8(t) = \frac{t(t-5)(t-110)}{250{,}000} + 0.8$
Creative 9	$f_9(t) = 1.2 \times 0.8^t$
Creative 10	$f_{10}(t) = \frac{0.7}{1 + exp\ (0.05(t-47.5))}$

Table B4

Estimation Time for 6 MCMC Chains in the Simulation Study

	Warmup (750)	Sampling (750)	Total (1,500)
Chain 1	60,481.2	71,061.2	131,542.4
Chain 2	60,349.8	70,394.8	130,744.6
Chain 3	60,195.4	69,241.8	129,437.2
Chain 4	61,440.3	70,322.5	131,762.8
Chain 5	60,555.0	70,728.6	131,283.6
Chain 6	62,180.8	70,845.0	133,025.8
Average	60,867.1	70,432.3	131,299.4

Table B5

Effective Sample Sizes and Rhats for Key Parameters in the Simulation Study

Parameter	Effective Sample Size	Rhat
μ	422.91	1.00
σ	114.92	1.06
β_1	712.78	1.01
β_2	1,079.93	1.00
β_3	1,180.33	1.00
π_0	1,147.15	1.00
η	97.77	1.07
ρ	136.27	1.05

Figure B1

Trace Plots for Key Parameters in the Empirical Analysis

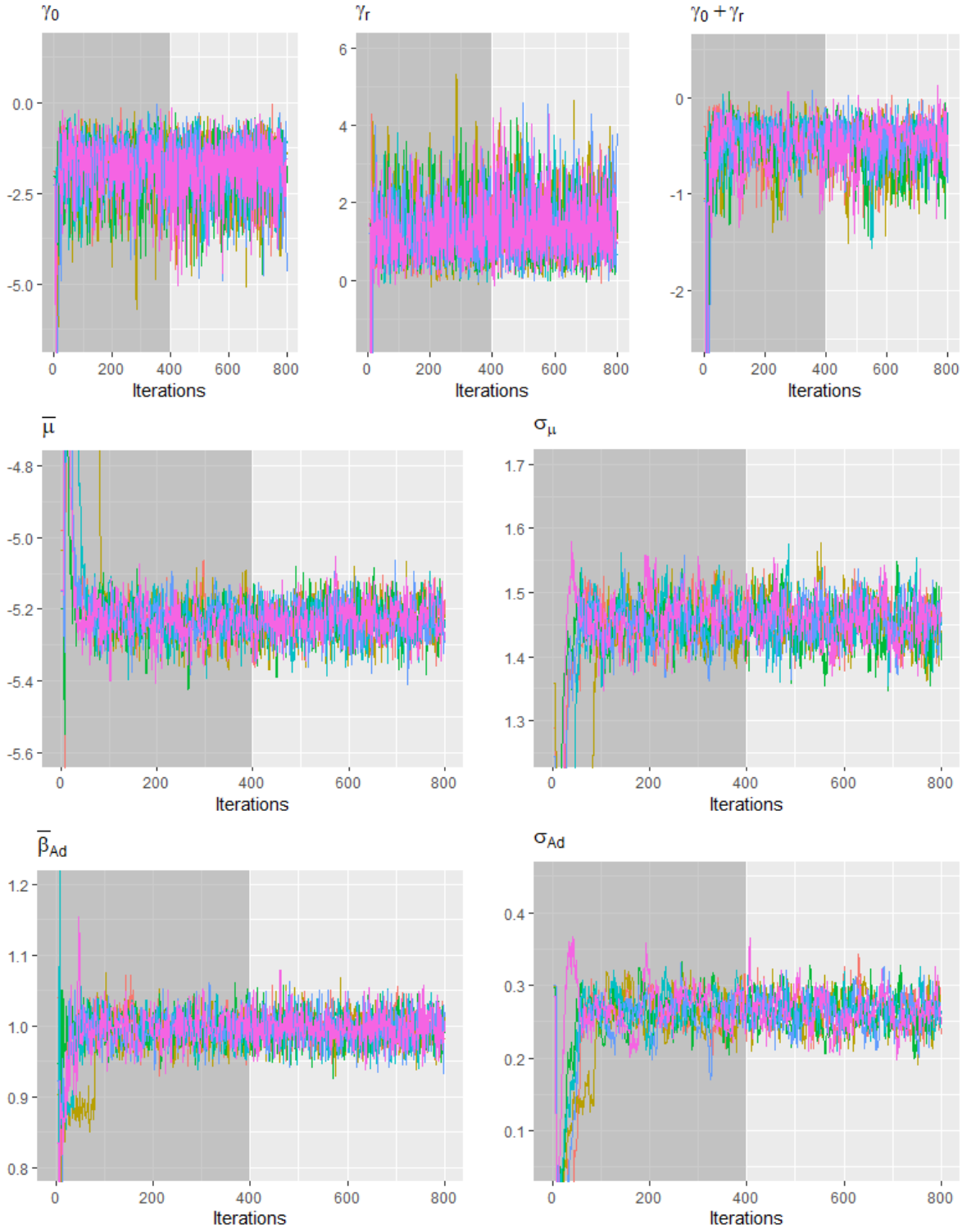
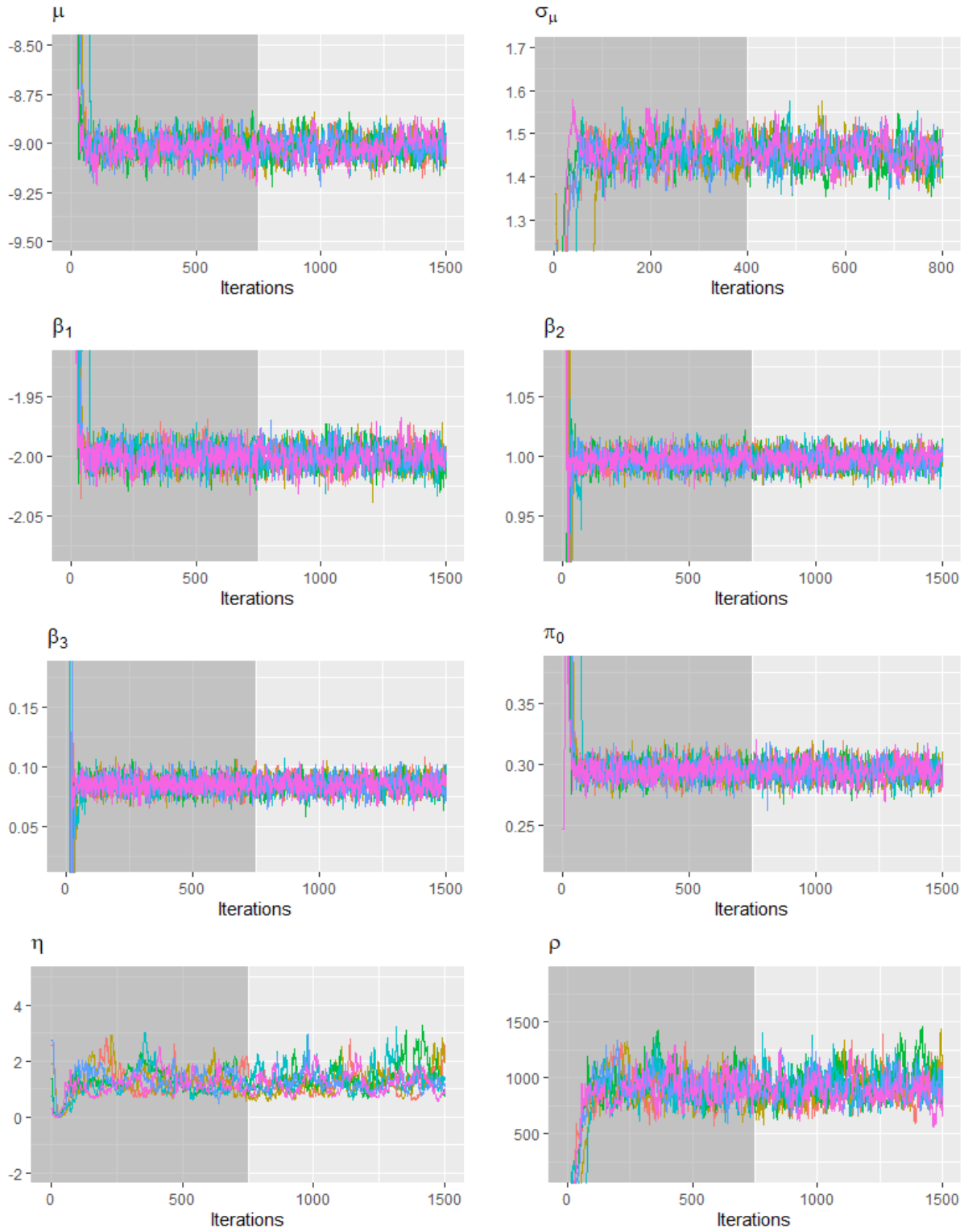


Figure B2

Trace Plots for Key Parameters in the Simulation Study



APPENDIX C

Appendix to Chapter 4

Table C1
EMA Items with the Strongest Positive and Negative Correlations

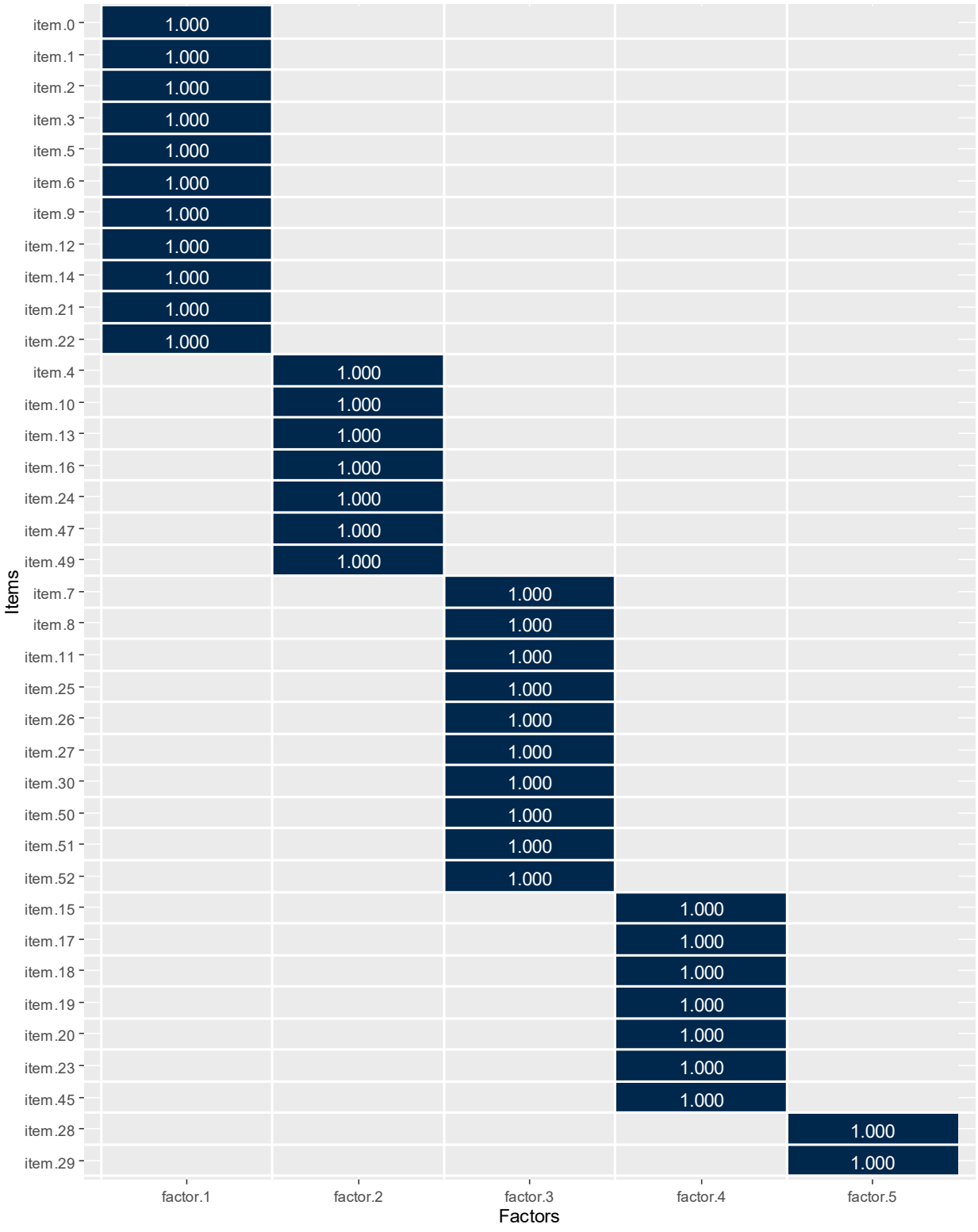
Item Number	Question	Item with Strongest Positive Correlation	Correlation	Item with Strongest Negative Correlation	Correlation
0	When the phone beeped, I was doing things automatically, without being aware of what I was doing.	When the phone beeped, I found myself doing things without paying attention.	0.596	At this moment, I feel that a Higher Power/God/Universal Spirit is with me.	-0.096
1	When the phone beeped, I found myself doing things without paying attention.	When the phone beeped, I was doing things automatically, without being aware of what I was doing.	0.596	Right now, I feel attentive.	-0.106
2	When the phone beeped, I was telling myself that I shouldn't be thinking the way I'm thinking.	When the phone beeped, I was thinking that some of my emotions are bad or inappropriate and shouldn't feel them.	0.507	Right now, I feel grateful.	-0.111
3	When the phone beeped, I was thinking that some of my emotions are bad or inappropriate and shouldn't feel them.	When the phone beeped, I was telling myself that I shouldn't be thinking the way I'm thinking.	0.507	Right now, I feel joyful.	-0.149

4	Right now, I feel active.	Right now, I feel enthusiastic.	0.312	Right now, I feel tired.	-0.269
5	Right now, I feel angry.	Right now, I feel disgusted.	0.466	Right now, I feel calm.	-0.312
6	Right now, I feel ashamed.	Right now, I feel disgusted.	0.432	Right now, I feel happy.	-0.198
7	Right now, I feel calm.	Right now, I feel relaxed.	0.383	Right now, I feel angry.	-0.312
8	Right now, I feel determined.	Right now, I feel enthusiastic.	0.370	Right now, I feel sad.	-0.202
9	Right now, I feel disgusted.	Right now, I feel angry.	0.466	Right now, I feel happy.	-0.307
10	Right now, I feel enthusiastic.	Right now, I feel happy.	0.398	Right now, I feel sad.	-0.236
11	Right now, I feel grateful.	Right now, I feel happy.	0.308	Right now, I feel sad.	-0.233
12	Right now, I feel guilty.	Right now, I feel ashamed.	0.429	Right now, I feel happy.	-0.229
13	Right now, I feel happy.	Right now, I feel joyful.	0.460	Right now, I feel irritable.	-0.328
14	Right now, I feel irritable.	Right now, I feel disgusted.	0.423	Right now, I feel happy.	-0.328
15	Right now, I feel lonely.	Right now, I feel sad.	0.392	Right now, I feel happy.	-0.220
16	Right now, I feel proud.	Right now, I feel happy.	0.397	Right now, I feel sad.	-0.226
17	Right now, I feel nervous.	Right now, I feel sad.	0.438	Right now, I feel relaxed.	-0.234
18	Right now, I feel sad.	Right now, I feel hopeless.	0.476	Right now, I feel happy.	-0.310
19	Right now, I feel restless.	Right now, I feel sad.	0.392	Right now, I feel relaxed.	-0.238
20	Right now, I feel tired.	Right now, I feel restless.	0.370	Right now, I feel active.	-0.269
21	Right now, I feel hopeless.	Right now, I feel scared.	0.520	Right now, I feel happy.	-0.224
22	Right now, I feel scared.	Right now, I feel hopeless.	0.520	Right now, I feel relaxed.	-0.213
23	Right now, I feel bored.	Right now, I feel scared.	0.383	Right now, I feel joyful.	-0.193
24	Right now, I feel joyful.	Right now, I feel happy.	0.460	Right now, I feel irritable.	-0.278
25	Right now, I feel attentive.	Right now, I feel joyful.	0.415	Right now, I feel irritable.	-0.183
26	Right now, I feel relaxed.	Right now, I feel calm.	0.383	Right now, I feel irritable.	-0.308
27	Right now, I am very motivated to be abstinent from cigarettes.	Right now, I am confident in my ability to be abstinent from cigarettes.	0.417	Right now, Smoking would improve my mood, be pleasurable, or help me cope with this situation.	-0.196
28	Right now, I have an urge to smoke.	Right now, Smoking would improve my mood, be pleasurable, or help me cope with this situation.	0.414	Right now, I am very motivated to be abstinent from cigarettes.	-0.148

29	Right now, Smoking would improve my mood, be pleasurable, or help me cope with this situation.	Right now, I have an urge to smoke.	0.414	Right now, I am very motivated to be abstinent from cigarettes.	-0.196
30	Right now, I am confident in my ability to be abstinent from cigarettes.	Right now, I am very motivated to be abstinent from cigarettes.	0.417	Right now, Smoking would improve my mood, be pleasurable, or help me cope with this situation.	-0.165
45	Since the last assessment, I experienced or thought about a stressful issue or problem.	Right now, I feel sad.	0.166	Right now, I feel relaxed.	-0.170
47	Since the last assessment, I experienced or thought about a positive experience.	Right now, I feel joyful.	0.198	Right now, I feel irritable.	-0.128
49	Since the last assessment, I felt little interest or pleasure while doing things.	Right now, I feel happy.	0.237	Right now, I feel disgusted.	-0.190
50	Since the last assessment, someone in my life provided me with help for a worry, problem or difficulty.	At this moment, I feel spiritually connected to other people.	0.181	Right now, I feel lonely.	-0.066
51	At this moment, I feel that a Higher Power/God/Universal Spirit is with me.	At this moment, I feel spiritually connected to other people.	0.415	Right now, I feel scared.	-0.149
52	At this moment, I feel spiritually connected to other people.	At this moment, I feel that a Higher Power/God/Universal Spirit is with me.	0.415	Right now, I feel angry.	-0.163

Figure C1

Posterior Probabilities of Each Item Belonging to Each Latent Factor in a 5-Factor Structure



BIBLIOGRAPHY

Adigüzel, Feray and Michel Wedel (2008), "Split Questionnaire Design for Massive Surveys," *Journal of Marketing Research*, 45 (5), 608–617.

Aggarwal, Praveen, and Rajiv Vaidyanathan (2003), "Use it or lose it: Purchase acceleration effects of time-limited promotions," *Journal of Consumer Behaviour*, 2 (4), 393–403.

Agostinelli, Francesco and Matthew Wiswall (2016), "Identification of Dynamic Latent Factor Models: The Implications of Re-normalization in a Model of Child Development," *NBER Working Paper 22441*.

Ailawadi, Kusum L., Bari A. Harlam, Jacques César, and David Trounce (2006), "Promotion Profitability for a Retailer: The Role of Promotion, Brand, Category, and Store Characteristics," *Journal of Marketing Research*, 43 (4), 518–535.

Ameri, Mina, Elisabeth Honka, and Ying Xie (2019), "Word of Mouth, Observed Adoptions, and Anime-Watching Decisions: The Role of the Personal vs. the Community Network," *Marketing Science*, 38 (4), 567–583.

Ansari, Asim and Carl F. Mela (2003), "E-customization," *Journal of Marketing Research*, 40 (2), 131–145.

Aral, Sinan, and Paramveer S. Dhillon (2021), "Digital Paywall Design: Implications for Content Demand and Subscriptions," *Management Science*, 67 (4), 2381–2402.

Aribarg, Anocha and Eric M. Schwartz (2019), "Native Advertising in Online News: Trade-offs Among Clicks, Brand Recognition, and Website Trustworthiness," *Journal of Marketing Research*, 57 (1), 20–34.

Ascarza, Eva, Oded Netzer, and Bruce G. S. Hardie (2018), "Some Customers Would Rather Leave Without Saying Goodbye," *Marketing Science*, 37 (1), 54–77.

Ascarza, Eva, Raghuram Iyengar, and Martin Schleicher (2016), "The Perils of Proactive Churn Prevention Using Plan Recommendations: Evidence from a Field Experiment*," Journal of Marketing Research*, 53 (1), 46–60.

Atasoy, Ozgun, and Carey K. Morewedge (2018), "Digital Goods Are Valued Less Than Physical Goods," *Journal of Consumer Research*, 44 (6), 1343–1357.

Basuroy, Suman, Kalpesh Kaushik Desai, and Debabrata Talukdar (2006), "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research*, 43 (2), 287–295.

Bawa, Kapil, and Robert Shoemaker (2004), "The Effects of Free Sample Promotions on Incremental Brand Sales," *Marketing Science*, 23 (3), 345–363.

Beal, Daniel J. and Howard M. Weiss (2003), "Methods of Ecological Momentary Assessment in Organizational Research," *Organizational Research Methods*, 6 (4), 440–464.

Beckham, Jean C., Matthew T. Wiley, Susannah C. Miller, Michelle F. Dennis, Sarah M. Wilson, F. Joseph McClernon, and Patrick S. Calhoun (2008), "Ad lib smoking in post-traumatic stress disorder: An electronic diary study," *Nicotine & Tobacco Research*, 10 (7), 1149–1157.

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

Berger, Jonah, Yoon Duk Kim, and Robert Meyer (2021), "What Makes Content Engaging? How Emotional Dynamics Shape Success," *Journal of Consumer Research*, 48 (2), 235–250.

Bolton, Ruth 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.

Bolton, Ruth N. and Katherine N. Lemon (1999), "A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction," *Journal of Marketing Research*, 36 (2), 171–186.

Braun, Michael and Wendy W. Moe (2013), "Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories," *Marketing Science*, 32 (5), 753–767.

Brecko, Kristina (2022), "New Features Free of Charge? Intertemporal Product Versions and Pricing in the Software Market," *Marketing Science*, 42 (1), 61–86.

Brown, Steven P., Pamela M. Homer, and J. Jeffrey Inman (1998), "A Meta-Analysis of Relationships between Ad-Evoked Feelings and Advertising Responses," *Journal of Marketing Research*, 35 (1), 114–126.

Bruce, Norris I., B. P. S. Murthi, and Ram C. Rao (2017), "A Dynamic Model for Digital Advertising: The Effects of Creative Format, Message Content, and Targeting on Engagement," *Journal of Marketing Research*, 54 (2), 202–218.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021), "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76 (1) 1–32.

Chae, Inyoung, Hernán A. Bruno, and Fred M. Feinberg (2019), "Wearout or Weariness? Measuring Potential Negative Consequences of Online Ad Volume and Placement on Website Visits," *Journal of Marketing Research*, 56 (1), 57–75.

Chae, Inyoung, Jihyeon Ha, and David Schweidel (2022), "The Impact of Paywall Suspensions on Digital Subscriptions," *Marketing Science*, forthcoming.

Chatterjee, Patrali, Donna L. Hoffman, and Thomas P. Novak (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science*, 22 (4), 520–41.

Clarke, Darral G. (1976), "Econometric Measurement of the Duration of Advertising Effect on Sales," *Journal of Marketing Research*, 13 (4), 345–357.

comScore (2007), "Cookie-Based Counting Overstates Size of Web Site Audiences," Accessed July 13, 2019.

Conti, Gabriella, Sylvia Frühwirth-Schnatter, James J. Heckman, and Rémi Piatek (2014), "Bayesian exploratory factor analysis," *Journal of Econometrics*, 183, 31–57.

Cooney, Ned L., Mark D. Litt, Judith L. Cooney, David T. Pilkey, Howard R. Steinberg, and Cheryl A. Oncken (2007), "Alcohol and Tobacco Cessation in Alcohol Dependent Smokers: Analysis of Real-Time Reports," *Psychology of Addictive Behaviors*, 21 (3), 277–286.

Cunha, Flavio, James J. Heckman and Susanne M. Schennach (2010), "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78 (3), 883–931.

Danaher, Peter J. and Tracey S. Dagger (2013), "Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign," *Journal of Marketing Research*, 50 (4), 517–534.

Davidson, Collin L., Michael D. Anestis, and Peter M. Gutierrez (2017), "Ecological Momentary Assessment is a Neglected Methodology in Suicidology," *Archives of Suicide Research*, 21 (1), 1–11.

de Reyck, Bert and Zeger Degraeve (2003), "Broadcast Scheduling for Mobile Advertising," *Operations Research*, 51 (4), 509–517.

de Matos, Miguel Godinho, Pedro Ferreira, and Rodrigo Belo (2018), "Target the Ego or Target the Group: Evidence from a Randomized Experiment in Proactive Churn Management," *Marketing Science*, 37 (5), 793–811.

Delfino, Ralph J., Larry D. Jamner, and Carol K. Whalen (2001), "Temporal analysis of the relationship of smoking behavior and urges to mood states in men versus women," *Nicotine & Tobacco Research*, 3 (3), 235–248.

Deng, Yiting, Anja Lambrecht, and Yongdong Liu (2022), "Spillover Effects and Freemium Strategy in the Mobile App Market," *Management Science*, forthcoming.

Dew, Ryan and Asim Ansari (2018), "Bayesian Nonparametric Customer Base Analysis with Model-Based Visualizations," *Marketing Science*, 37 (2), 216–235.

Dew, Ryan, Asim Ansari, and Yang Li (2020), "Modeling Dynamic Heterogeneity Using Gaussian Processes," *Journal of Marketing Research*, 57 (1), 55–77.

Dhar, Ravi, Joel Huber, and Uzma Khan (2007), "The Shopping Momentum Effect," *Journal of Marketing Research*, 44 (3), 370–378.

Dias, M. Benjamin, Dominique Locher, Ming Li, Wael El-Deredy, and Paolo J.G. Lisboa (2008), "The Value of Personalised Recommender Systems to E-Business: A Case Study," *Proceedings of the 2008 ACM conference on Recommender systems*, 291–294.

Dubé, Jean-Pierre, Zheng Fang, Nathan Fong, and Xueming Luo (2017) "Competitive Price Targeting with Smartphone Coupons", *Marketing Science*, 36 (6), 944–975.

Dunton, Genevieve Fridlund, Yue Liao, Stephen Intille, Jimi Huh, and Adam Leventhal (2015), "Momentary Assessment of Contextual Influences on Affective Response During Physical Activity," *Health Psychology*, 34 (12), 1145–1153.

Farley, John U., John A. Howard, and Donald R. Lehmann (1976), "A 'Working' System Model of Car Buyer Behavior," *Management Science*, 23 (3), 235–247.

Fong, Nathan, Yuchi Zhang, Xueming Luo, and Xiaoyi Wang (2019), "Targeted Promotions on an E-Book Platform: Crowding Out, Heterogeneity, and Opportunity Costs," *Journal of Marketing Research*, 56 (2), 310–323.

Fradkin, Andrey, Elena Grewal, and David Holtz (2021), "Reciprocity and Unveiling in Two-Sided Reputation Systems: Evidence from an Experiment on Airbnb," *Marketing Science*, 40 (6), 1013–1029.

Frazier, Peter I. (2018), "Bayesian Optimization," *INFORMS TutORials in Operations Research*, 255–278.

Gartner (2022), "Gartner Survey Reveals Marketing Budgets Have Increased to 9.5% of Overall Company Revenue in 2022", Accessed March 13, 2023.

Golany, Boza, Fred Young Phillips, and John J. Rousseau (1991), "Few-wave vs. continuous consumer panels: Some issues of attrition, bias and variance," *International Journal of Research in Marketing*, 8 (3), 273–280.

Goldfarb, Avi and Catherine Tucker (2011), "Online Display Advertising: Targeting and Obtrusiveness," *Marketing Science*, 30 (3), 389–404.

Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky (2019), "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," *Marketing Science*, 38 (2), 193–225.

Gu, Xian, P. K. Kannan, and Liye Ma (2018), "Selling the Premium in Freemium," *Journal of Marketing*, 82 (6), 10–27.

Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203–238.

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

Halbheer, Daniel, Florian Stahl, Oded Koenigsberg, and Donald R. Lehmann (2014), “Choosing a digital content strategy: How much should be free?,” *International Journal of Research in Marketing*, 31 (2), 192–206.

Hansen, Nele, Ann-Kristin Kupfer, and Thorsten Hennig-Thurau (2018), “Brand crises in the digital age: The short- and long-term effects of social media firestorms on consumers and brands,” *International Journal of Research in Marketing*, 35 (4), 557–574.

He, Sherry, Brett Hollenbeck, and Davide Proserpio (2022), “The Market for Fake Reviews,” *Marketing Science*, 41 (5), 896–921.

Heckman, James (1974), “Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, 42 (4), 679–694.

Hoban, Paul R., and Randolph E. Bucklin (2014), “Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment,” *Journal of Marketing Research*, 52 (3), 375–393.

Hoch, Stephen J., Eric T. Bradlow, Brian Wansink (1999), “The Variety of an Assortment,” *Marketing Science*, 18 (4), 527–546.

Hwang, Minha, Bart J. Bronnenberg, and Raphael Thomadsen (2010), “An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets,” *Marketing Science*, 29 (5), 858–879.

IMARC Group (2023), “Online Entertainment Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2023-2028,” Accessed April 2, 2023.

Interactive Advertising Bureau and PricewaterhouseCoopers (2021), “IAB/PwC Internet Ad Revenue Report, FY 2020,” Accessed March 17, 2022.

Ishwaran, Hemant and J. Sunil Rao (2005), "Spike and slab variable selection: Frequentist and Bayesian strategies," *Annals of Statistics*, 33 (2), 730–773.

Joo, Hailey Hayeon, Minki Kim, Jungmin Lee, and Pradeep Chintagunta (2021), “Drinking Through Good Times and Bad: The Role of Consumer Differences,” *Journal of Marketing Research*, 58 (4), 721–741.

Kahn, Barbara E. and Brian Wansink (2004), “The Influence of Assortment Structure on Perceived Variety and Consumption Quantities,” *Journal of Consumer Research*, 30 (4), 519–533.

Kim, Sungjin, Sachin Gupta, and Clarence Lee (2021), "Managing Members, Donors, and Member-Donors for Effective Nonprofit Fundraising," *Journal of Marketing*, 85 (3), 220–239.

Klasnja, Predrag, Eric B. Hekler, Saul Shiffman, Audrey Boruvka, Daniel Almirall, Ambuj Tewari, and Susan A. Murphy (2015), "Microrandomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions." *Health Psychology*, 34 (Suppl), 1220-1228.

Kumar, Anuj, Amit Mehra, and Subodha Kumar (2019), "Why Do Stores Drive Online Sales? Evidence of Underlying Mechanisms from a Multichannel Retailer," *Information Systems Research*, 30 (1), 319–338.

Kumar, V., JeeWon Brianna Choi, and Mallik Greene (2017), "Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects," *Journal of the Academy of Marketing Science*, 45 (2), 268–288.

Kupor, Daniella and Zakary L. Tormala (2015), "Persuasion, Interrupted: The Effect of Momentary Interruptions on Message Processing and Persuasion," *Journal of Consumer Research*, 42 (2), 300–315.

Laibson, David (1997), "Golden eggs and hyperbolic discounting," *Quarterly Journal of Economics*, 112 (2), 443–477.

Lambrecht, Anja and Catherine Tucker (2013), "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 50 (5), 561–576.

Lambrecht, Anja, and Kanishka Misra (2017), "Fee or Free: When Should Firms Charge for Online Content?," *Management Science*, 63 (4), 1150–1165.

Lambrecht, Anja, Avi Goldfarb, Alessandro Bonatti, Anindya Ghose, Daniel G. Goldstein, Randall Lewis, Anita Rao, Navdeep Sahni, and Song Yao (2014), "How do firms make money selling digital goods online?," *Marketing Letters*, 25 (3), 331–341.

Larson, Randy J. and Margaret Kasimatis (1991), "Day-to-day physical symptoms: Individual differences in the occurrence, duration, and emotional concomitants of minor daily illnesses," *Journal of Personality*, 59 (3), 387–423.

Larson, Reed (1989), "Beeping Children and Adolescents: A Method for Studying Time Use and Daily Experience," *Journal of Youth and Adolescence*, 18 (6), 511–530.

Larson, Reed, Suman Verma, and Jodi Dworkin (2001), "Men's Work and Family Lives in India: The Daily Organization of Time and Emotion," *Journal of Family Psychology*, 15 (2), 206–224.

Lee, Lillian (1999), "Measures of distributional similarity," *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 25–32.

Lewis, Randall A., Justin M. Rao, and David H. Reiley (2011), "Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising," *Proceedings of the 20th ACM International World Wide Web Conference (WWW20)*, 157–166.

Li, Hongshuang (Alice), Sanjay Jain, and P.K. Kannan (2019), "Optimal Design of Free Samples for Digital Products and Services," *Journal of Marketing Research*, 56 (3), 419–438.

Li, Hui (2019), "Intertemporal Price Discrimination with Complementary Products: E-Books and E-Readers," *Management Science*, 65 (6), 2665–2694.

Li, Yang and Asim Ansari (2014), "A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models," *Management Science*, 60 (5), 1161–1179.

Li, Ye, Antonia Krefeld-Schwalb, Daniel G. Wall, Eric J. Johnson, Olivier Toubia, and Daniel M. Bartels (2022), "The More You Ask, the Less You Get: When Additional Questions Hurt External Validity," *Journal of Marketing Research*, 59 (5), 963–982.

Lin, Tung-Ching, Jack Shih-Chieh Hsu, and Hui-Ching Chen (2013), "Customer Willingness to Pay for Online Music: The Role of Free Mentality," *Journal of Electronic Commerce Research*, 14 (4), 315–333.

Litt, Mark D., Ned K. Cooney, and Priscilla Morse (1998), "Ecological Momentary Assessment (EMA) With Treated Alcoholics: Methodological Problems and Potential Solutions," *Health Psychology*, 17 (1), 48–52.

Liu, Charles Zhechao, Yoris A. Au, and Hoon Seok Choi (2014), "Effects of Freemium Strategy in the Mobile App Market: An Empirical Study of Google Play," *Journal of Management Information Systems*, 31 (3), 326–354.

Liu, Xiao, Dokyun Lee, and Kannan Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, 56 (6), 918–943.

Lu, Joy, Eric T. Bradlow, and J. Wesley Hutchinson (2022), "Testing Theories of Goal Progress in Online Learning," *Journal of Marketing Research*, 59 (1), 35–60.

Lukasiewicz, Michael, Marion Fareng, Amine Benyamina, Lisa Blecha, Michel Reynaud, and Bruno Falissard (2007), "Ecological momentary assessment in addiction," *Expert Review of Neurotherapeutics*, 7 (8), 939–950.

Malthouse, Edward C., Bobby J. Calder, and Ajit Tamhane (2007), "The Effects of Media Context Experiences On Advertising Effectiveness," *Journal of Advertising*, 36 (3), 7–18.

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K. Chintagunta (2006), "The Effect of Banner Advertising on Internet Purchasing," *Journal of Marketing Research*, 43 (1), 98–108.

McAlinn, Kenichiro, Veronika Rockova, and Enakshi Saha (2018), "Dynamic Sparse Factor Analysis," *arXiv:1812.04187*.

McAlister, Leigh (1982), "A Dynamic Attribute Satiation Model of Variety-Seeking Behavior," *Journal of Consumer Research*, 9 (2), 141–150.

Moon, Sangkil, Wagner A. Kamakura, and Johannes Ledolter (2007), "Estimating Promotion Response When Competitive Promotions Are Unobservable," *Journal of Marketing Research*, 44 (3), 503–515.

Naik, Prasad A., Murali K. Mantrala, and Alan G. Sawyer (1998), "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17 (3), 214–235.

Nair, Harikesh (2007), "Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games," *Quantitative Marketing and Economics*, 5, 239–292.

Nerlove, Mark and Kenneth J. Arrow (1962), "Optimal advertising policy under dynamic conditions," *Economica*, 29 (114), 129–142.

Neslin, Scott A., Caroline Henderson, and John Quelch (1985), "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4 (2), 147–165.

Nevskaya, Yulia and Paulo Albuquerque (2019), "How Should Firms Manage Excessive Product Use? A Continuous-Time Demand Model to Test Reward Schedules, Notifications, and Time Limits," *Journal of Marketing Research*, 56 (3), 379–400.

Padilla, Nicolas and Eva Ascarza (2021), “Overcoming the Cold Start Problem of Customer Relationship Management Using a Probabilistic Machine Learning Approach,” *Journal of Marketing Research*, 58 (5), 981–1006.

Papastamoulis, Panagiotis and Ioannis Ntzoufras (2022), “On the identifiability of Bayesian factor analytic models,” *Statistics and Computing*, 32, 23.

Park, Sehoon and Minhi Hahn (1991), “Pulsing in a Discrete Model of Advertising Competition,” *Journal of Marketing Research*, 28 (4), 397–405.

Pattabhiramaiah, Adithya, S. Sriram, and Puneet Manchanda (2019), “Paywalls: Monetizing online content,” *Journal of Marketing*, 83 (2), 19–36.

Prasad, Ashutosh, Vijay Mahajan, and Bart Bronnenberg (2003), “Advertising versus pay-per-view in electronic media,” *International Journal of Research in Marketing*, 20 (1), 13–30.

Rafieian, Omid (2022), “Optimizing User Engagement through Adaptive Ad Sequencing,” *Marketing Science*, forthcoming.

Rao, Ram C. (1986), “Estimating Continuous Time Advertising-Sales Models,” *Marketing Science*, 5 (2), 125–142.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman (1999), “Choosing Less-Preferred Experiences for the Sake of Variety,” *Journal of Consumer Research*, 26 (1), 1–15.

Raut, Sumit, Sanjeev Swami, Eunkyu Lee, and Charles B. Weinberg (2008), “How Complex Do Movie Channel Contracts Need to Be?,” *Marketing Science*, 27 (4), 627–641.

Research and Markets (2022), “Webcomics Global Market Report 2022,” Accessed March 21, 2023.

Rockova, Veronika and Kenichiro McAlinn (2021), "Dynamic Variable Selection with Spike-and-Slab Process Priors," Bayesian Analysis, 16 (1), 233–269.

Rolls, Barbara J., E. A. Rowe, E. T. Rolls, Breda Kingston, Angela Megson, and Rachel Gunary (1981), “Variety in a meal enhances food intake in man,” *Physiology & Behavior*, 26 (2), 215–221.

Roos, Jason M. T., Carl F. Mela, and Ron Shachar (2020), “The Effect of Links and Excerpts on Internet News Consumption,” *Journal of Marketing Research*, 57 (3), 395–421.

Rossi, Peter (2014), “Even the Rich can make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications,” *Marketing Science*, 33 (5), 655–672.

Rowan, Paul J., Ludmila Cofta-Woerpel, Carlos A. Mazas, Jennifer Irvin Vidrine, Lorraine R. Reitzel, Paul M. Cinciripini, and David W. Wetter (2007), “Evaluating Reactivity to Ecological Momentary Assessment During Smoking Cessation,” *Experimental and Clinical Psychopharmacology*, 15 (4), 382–389.

Rubin, Donald B. and Richard P. Waterman (2006), “Estimating the causal effects of marketing interventions using propensity score methodology,” *Statistical Science*, 206–222.

Rutz, Oliver J. and George F. Watson IV (2019), “Endogeneity and marketing strategy research: an overview,” *Journal of the Academy of Marketing Science*, 47, 479–498.

Rutz, Oliver J. and Randolph E. Bucklin (2011), “From Generic to Branded: A Model of Spillover Dynamics in Paid Search Advertising,” *Journal of Marketing Research*, 48 (1), 87–102.

Sahni, Navdeep S. (2015), “Effect of temporal spacing between advertising exposures: Evidence from online field experiments,” *Quantitative Marketing and Economics*, 13, 203–247.

Sahni, Navdeep S., Sridhar Narayanan, and Kirthi Kalyanam (2019), “An Experimental Investigation of the Effects of Retargeted Advertising: The Role of Frequency and Timing,” *Journal of Marketing Research*, 56 (3), 401–418.

Scherer, Klaus R. (2005), “What are emotions? And how can they be measured?,” *Social Science Information*, 44 (4), 695–729.

Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2017), “Customer acquisition via display advertising using multi-armed bandit experiments,” *Marketing Science*, 36 (4), 500–522.

Schweidel, David A., and Wendy W. Moe (2016), “Binge Watching and Advertising,” *Journal of Marketing*, 80 (5), 1–19.

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

Sherman, Lee and John Deighton (2001), "Banner Advertising: Measuring Effectiveness and Optimizing Placement," *Journal of Interactive Marketing*, 15 (2), 60–64.

Shiffman, Saul (2009), "Ecological Momentary Assessment (EMA) in Studies of Substance Use," *Psychological Assessment*, 21 (4), 486–497.

Shiffman, Saul and Andrew J. Waters (2004), "Negative Affect and Smoking Lapses: A Prospective Analysis," *Journal of Consulting and Clinical Psychology*, 72 (2), 192–201.

Shively, Thomas, Greg M. Allenby, and Robert Kohn (2000), "A Nonparametric Approach to Identifying Latent Relationships in Hierarchical Models," *Marketing Science*, 19 (2), 149–162.

Shu, Suzanne B., and Ayelet Gneezy (2010), "Procrastination of Enjoyable Experiences," *Journal of Marketing Research*, 47 (5), 933–944.

Statista (2023a), "Average time spent per day with digital media in the United States from 2011 to 2024," Accessed March 12, 2023.

Statista (2023b), "Retail e-commerce sales worldwide from 2014 to 2026," Accessed March 12, 2023.

Su, Xuanming (2007), "Intertemporal Pricing with Strategic Customer Behavior," *Management Science*, 53 (5), 726–741.

Sudman, Seymour (1964), "On the Accuracy of Recording of Consumer Panels: I," *Journal of Marketing Research*, 1 (2), 14–20.

Tellis, Gerard J. (2009), "Generalizations about Advertising Effectiveness in Markets," *Journal of Advertising Research*, 49 (2), 240–245.

Tellis, Gerard J. and Philip Hans Franses (2006), "Optimal Data Interval for Estimating Advertising Response," *Marketing Science*, 25 (3), 217–229.

Todd, Mike, Stephen Armeli, Howard Tennen, Margaret Anne Carney, Samuel A. Ball, and Henry R. Kranzler (2005), "Drinking to cope: A comparison of questionnaire and electronic diary reports," *Journal of Studies on Alcohol*, 66 (1), 121–129.

Toubia, Olivier, John R. Hauser, and Rosanna Garcia (2007), "Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application," *Marketing Science*, 26 (5), 596–610.

van Heerde, Harald J., Peter S. H. Leeflang, and Dick R. Wittink (2004), "Decomposing the Sales Promotion Bump with Store Data," *Marketing Science*, 23 (3), 317–334.

Villas-Boas, Miguel (1993), “Predicting Advertising Pulsing Policies in Oligopoly: A Model and Empirical Test,” *Marketing Science*, 12 (1), 88–102.

Weiss, Howard M., Jeffrey P. Nicholas, and Catherine S. Daus (1999), “An Examination of the Joint Effects of Affective Experiences and Job Beliefs on Job Satisfaction and Variations in Affective Experiences over Time,” *Organizational Behavior and Human Decision Processes*, 78 (1), 1–24.

Winer, Russel S. (1983), “Attrition Bias in Econometric Models Estimated with Panel Data,” *Journal of Marketing Research*, 20 (2), 177–186.

Wohllebe, Atilla, Dirk-Siegfried Hübner, Uwe Radtke, and Szilárd Podruzsik (2021), “Mobile apps in retail: Effect of push notification frequency on app user behavior,” *Innovative Marketing*, 17 (2), 102–111.

Wright, Peter L. (1973), “The Cognitive Processes Mediating Acceptance of Advertising,” *Journal of Marketing Research*, 10 (1), 53–62.

Yan, Shunyao, Klaus M. Miller, and Bernd Skiera (2022), “How Does the Adoption of Ad Blockers Affect News Consumption?,” *Journal of Marketing Research*, 59 (5), 1002–1018.

Yang, Sha, Yi Zhao, and Ravi Dhar (2010), “Modeling the Underreporting Bias in Panel Survey Data,” *Marketing Science*, 29 (3), 525–539.

Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), “Leveraging Missing Ratings to Improve Online Recommendation Systems,” *Journal of Marketing Research*, 43 (3), 355–365.

Zhang, Shuo, Tat Y. Chan, Xueming Luo, and Xiaoyi Wang (2021), “Time-Inconsistent Preferences and Strategic Self-Control in Digital Content Consumption,” *Marketing Science*, 41 (3), 616–636.

Zippia (2023), "How Many People Use The Internet? [2023]: 35 Facts About Internet Usage In America And The World", Accessed March 12, 2023.