Consumer Inattention, Uncertainty, and Marketing Strategy
by
Xinyu Cao
B.S. in Mathematics and Physics, Tsinghua University, 2011
M.S. in Industrial Engineering and Operations Research,
University of California, Berkeley, 2013
S.M. in Management Research,
Massachusetts Institute of Technology, 2016
Submitted to the MIT Sloan School of Management
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN MANAGEMENT
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 2018
© Massachusetts Institute of Technology 2018. All rights reserved.

Signature redacted
Author ....................
MIT Sloan School of Management
May 8, 2018

Signature redacted
Certified by .........
Juanjuan Zhang
Epoch Foundation Professor of International Management
Professor of Marketing
Thesis Supervisor

Signature redacted
Accepted by ..............................................
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program
Consumer Inattention, Uncertainty, and Marketing Strategy

by

Xinyu Cao

Submitted to the MIT Sloan School of Management
on May 8, 2018, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY IN MANAGEMENT

Abstract

This dissertation investigates the implications of consumer inattention and uncer-
tainty for firms’ advertising and pricing decisions.

The first chapter is an overview of the problems addressed in the dissertation and
the main findings.

The second chapter develops a theory-based, cost-effective method to estimate the
demand for new products using choice experiments. The premise is that consumers
are uncertain about their valuation of a new product and need to spend costly effort to
learn their valuation. The effort consumers spend is affected by the probability of their
choice being realized, and as a result will change the manifested demand curve derived
from choice experiments. We run a large-scale choice experiment on a mobile game
platform, where we randomize the price and realization probability of a new product.
Data support our theoretical hypothesis. We then estimate a structural model of
consumer decisions. The structural estimates allow us to accurately infer actual
demand based on choice experiments of small to moderate realization probabilities.

The third chapter examines firms’ advertising strategy on social media under
consumers’ limited attention. Advertising on social media faces a new challenge as
consumers can actively select which advertisers to follow. A Bayesian learning model
suggests that consumers with limited attention may rationally choose to unfollow a
firm. This happens if consumers already know about the firm’s value well and if the
firm advertises too intensely. However, we find that intensive advertising may still
be the optimal strategy for firms. If a firm is perceived as providing low value, it
will want to advertise aggressively to change consumers’ mind; if a firm is perceived
as providing higher value, it will also want to advertise intensively, but in an effort
to crowd-out advertising messages from its competitors. Tracking company accounts
of 49 TV shows on the most popular tweeting website in China provides empirical
evidence that both popular and non-popular firms advertise intensively, although the
number of followers does go down when a firm advertises too intensively.

The fourth chapter investigates channel coordination in search advertising. Given
that consumers have limited attention, there are only a limited number of advertising
slots on search engine platforms that can attract positive number of clicks. A manu-
facturer can sponsor retailers to advertise its products while at the same time compete with them in a position auction with limited number of slots. We prescribe the optimal cooperative search advertising strategies for the manufacturer. We find that it may not be optimal for a manufacturer to cooperate with all of its retailers, even when these retailers are ex ante the same. This finding reflects the manufacturer's tradeoff between higher demand and higher bidding cost caused by more intensified competition. With two asymmetric retailers, the manufacturer should support the retailer with the higher channel profit per click to get a higher position than the other retailer. The manufacturer should take a higher position than a retailer when its profit per click via direct sales exceeds the channel profit per click of the retailer. We also investigate how a manufacturer uses both wholesale and advertising contracts to coordinate channels with endogenous retail prices.

Thesis Supervisor: Juanjuan Zhang
Title: Epoch Foundation Professor of International Management
Professor of Marketing
Acknowledgments

In the first place, I would like to express my gratefulness to my thesis supervisor, Juanjuan Zhang. I deeply feel that having Juanjuan as my advisor is one of the most fortunate things in my life. Before formally working with Juanjuan, I was perplexed about my research direction. It was Juanjuan that helped me find the research topics that I am truly excited about. Every meeting with Juanjuan was inspiring and could spark new ideas. Juanjuan gives me not only invaluable guidance in research, but also incredible support in many other aspects of life. She is always very positive and encouraging. Every time when I feel frustrated due to an obstacle in research or a failed interview, she can always cheer me up and make my heart full of hope. She generously shares all her experiences, and always encourages me to be myself. Since I started to work with Juanjuan, I became more confident and comfortable in various situations. She is definitely my role model in my academic career.

I would also like to thank all other faculty members who have helped me throughout my Ph.D. study. John Hauser encouraged me to explore my research interests early on and provided invaluable support. I also learned many concepts and techniques from his carefully designed Ph.D. seminar, which helped me a lot in my research. Tony Ke and I had been known each other before I came to MIT. He was my mentor when I was a graduate student at UC Berkeley. It is like a miracle that after several years, we met again at MIT. I am incredibly lucky to have Tony as my mentor, coauthor, and friend. Working with him is efficient and enjoyable. He can always quickly get to the key point, pushing me to think deeper and more rigorously. Birger Wernerfelt gave me precious opportunities to be his research assistant and teaching assistant for many times, and I benefited tremendously from these experiences. From his class and our individual meetings, I learned how to formulate and screen research ideas at the early stage, and how to investigate managerial problems at a deep level. From Catherine Tucker, I learned how to dig into data, as well as how to teach in a very effective and inspiring way. I have also got extremely helpful suggestions on my research from Sinan Aral, Sharmila Chatterjee, Dean Eckles, John Little, Drazen Pr-
elec, and Duncan Simester. I deeply admire them as superb researchers and charming professors.

I am very lucky to be surrounded by a brilliant group of peers and friend at MIT. Thanks to the marketing family — Matthew Cashman, Yiqun Cao, James Duan, Madhav Kumar, Song Lin, Nell Putnam-Farr, Xiang Song, Artem Timoshenko, Jeremy Yang, Shuyi Yu, Yunhao Zhang, and Yuting Zhu, my Ph.D. life has been not lonely at all. I am especially thankful to my former officemate and also my marketing “brother” Song Lin for his tremendous help and guidance. I was touched by his diligence and unselfishness. My friends from other departments, Jie Bai, Yu Shi, Fei Song, Shujing Wang, Yufei Wu, Xin Xu, and many others, have made my life at MIT full of joy.

Words cannot describe my gratitude to my family. I have always been feeling that I have the best parents in the world. Though they are not very wealthy, they supported me to get the best possible education in my hometown and in my home country. Though I am their only child, they encouraged me to go abroad to see the world. They are always supportive and cheerful. No matter how many difficulties I met, I never lose hope, because I know my family is there supporting me. Bin, thank you for your unconditional love and support. Life has not been easy for a family with two Ph.D. students in two different cities. The many flights on Friday midnight and Monday morning, as well as the long phone calls everyday have recorded our love and memorable doctoral life. After six years being in long distance, we will finally unite to start our new chapter of life.
Contents

1 Overview

2 Prelaunch Demand Estimation
   2.1 Introduction .................................................. 19
   2.2 Literature Review ............................................. 22
   2.3 Theory Model .................................................. 26
   2.4 Field Experiment .............................................. 30
   2.5 Reduced-Form Evidence ....................................... 32
   2.6 Evaluating the Demand Estimation Method ................. 34
      2.6.1 A Structural Model of Consumer Product Choice..... 35
      2.6.2 Estimation Procedure ................................... 37
      2.6.3 Identification ........................................... 38
      2.6.4 Estimation Results ...................................... 38
      2.6.5 Forecasting Demand in Real Purchase Settings ...... 39
   2.7 Concluding Remarks .......................................... 43

2.A Appendix ......................................................... 44
   2.A.1 Figures ...................................................... 44
   2.A.2 Tables ....................................................... 49
   2.A.3 Proof of Proposition 1 ..................................... 53
   2.A.4 Proof of Proposition 2 ..................................... 54
   2.A.5 Details of Structural Estimation and Extrapolation .... 57
3 Rational Spamming 61
  3.1 Introduction ............................................. 61
  3.2 Background and Motivating Phenomena .................... 66
    3.1 Background and Data .................................... 66
    3.2 Summary Statistics and Motivating Phenomena .......... 68
  3.3 Model .................................................. 74
    3.1 Rational Unfollowing ................................... 77
    3.2 Rational Spamming ...................................... 81
    3.3 Extensions ........................................... 87
  3.4 Concluding Remarks ...................................... 92
  3.A Appendix ............................................... 94
    3.A.1 Figures ............................................. 94
    3.A.2 Tables ............................................. 97
    3.A.3 Proof of Proposition 4 ............................ 101
    3.A.4 Proof of Proposition 5 ............................ 110
    3.A.5 Online Survey .................................... 112

4 Cooperative Search Advertising 117
  4.1 Introduction ............................................. 117
  4.2 Basic Model ............................................ 125
    4.1 Position Auctions .................................... 125
    4.2 Coordination Game .................................... 127
    4.3 Equilibrium Analysis ................................ 129
    4.4 Integrated Channel ................................... 135
  4.3 Manufacturer's Direct Participation in Search Advertising 139
  4.4 Extensions ............................................ 143
    4.1 Identity-Dependent Click-Through Rate ................ 143
    4.2 Endogenous Retail Prices with Price Competition .... 144
    4.3 Endogenous Wholesale Contracts and Retail Prices .... 149
  4.5 Conclusion ............................................ 155
4.A Appendix ............................................ 157
4.A.1 Figures ........................................... 157
4.A.2 Proof of Theorem 2 ............................... 161
4.A.3 Analysis of the Case with \( v_A \) between \( \theta_1 r_1 \) and \( \theta_2 r_2 \) .......... 163
4.A.4 Proof of Proposition 6 ............................ 165
4.A.5 Proof of Theorem 3 ............................... 172
4.A.6 Proof of Theorem 4 ............................... 187
4.A.7 Proof of Theorem 5: ............................. 189
4.A.8 Endogenous Wholesale Contracts and Retail Prices: ............ 194
List of Figures

2-1 Intended Contribution of the Chapter .......................... 44
2-2 Illustration of Theory Prediction ................................. 45
2-3 Screenshot of the Choice Task ................................. 45
2-4 Content of the Player Package ................................. 46
2-5 Aggregate-Level Experiment Result ......................... 47
2-6 Compare Actual and Forecast Demand Curves ............... 48

3-1 Average Posting Intensity vs. Popularity ....................... 94
3-2 The Consumer’s Optimal Strategy of Unfollowing ............... 95
3-3 Comparative Statics .................................................. 96
3-4 The Probability of Being Chosen as a Function of Posting Intensity,
Given All Other Firms Post to the Upper Bound Intensity ........... 96
3-A1 $\Sigma_S(1, 1; n_1, n_2)$ as a function of $n_1$ when $\sigma^2_{01} < \sigma^2_{02}, M > M_0$ ........ 108

4-1 An Example of Google Ads on“Laptop” ............... 157
4-2 Manufacturer’s Optimal Cooperative Search Advertising Strategy, Given
Two Symmetric Retailers. ............................................ 158
4-3 Manufacturer’s Optimal Advertising and Cooperative Search Advert-
tising Strategies, Given Two Symmetric Retailers and Equal Channel
Profit. ................................................................. 159
4-4 Timeline of Events with Endogenous Wholesale Contracts and Retail
Prices. ................................................................. 159
4-5 Manufacturer’s Optimal Wholesaling and Cooperative Advertising Strate-
gies under Linear Wholesale Contracts. ........................ 160
List of Tables

2.1 Number of Users by Realization Probability and by Price 49
2.2 Summary Statistics of User Characteristics 49
2.3 Individual-Level Experiment Result 50
2.4 Price Sensitivity Increases with Realization Probability 50
2.5 Summary Statistics of Effort Measures 50
2.6 Effort Increases with Realization Probability 51
2.7 Estimation Results 51
2.8 Estimated Effort Level and Effort Cost 51
2.9 Performance Comparison across Demand Estimation Methods 52

3.1 Summary Statistics of Accounts 97
3.2 Hourly Number of Posts, Conditional on Posting 98
3.3 Advertising Intensively May Drive Followers Away, Hourly 99
3.4 Advertising Intensively May Drive Followers Away, 4-Hour Period 100
3.A1 When $n_1 + n_2 \geq M, n_2 > 0$, the relationship between $f(\psi) = \Sigma S(1, 1; n_1, n_2)$ and $\psi = \frac{n_1}{n_2}$ 104
3.A2 Summary Statistics of the 5-level Questions and Time of Taking the Survey 115

4.A1 Profits and NE Condition for each Position Configuration – Integrated Channel 170
4.A2 Profits under each Position Configuration for Integrated Channel and Non-integrated Channel 171
4.A3 Optimal Participation Rates and Manufacturer’s Profit 183
Chapter 1

Overview

When faced with a product (or a set of products), consumers are uncertain about their valuation of the products and need to spend the effort to learn about the product's valuation (e.g. Wernerfelt 1994, Villas-Boas 2009, Wathieu and Bertini 2007, Guo and Zhang 2012) before making a purchase decision. Such effort is costly as people have limited attention in receiving and processing information (e.g. Sims 2003, DellaVigna 2009). Thus, consumers can rationally choose the amount of effort they put into learning about a product’s valuation before purchasing, and rationally decide the amount and type of information they would like to obtain about the product, which should have significant implications on firms’ pricing and advertising decisions. This dissertation investigates firms’ pricing and advertising strategies considering consumer inattention.

Chapter 2 develops a demand estimation method using choice experiments considering consumers’ uncertainty in product valuation and their rational choice of effort level when making a choice. Demand estimation is important for new products to succeed, as it is the key to making the pricing and production decisions. Researchers have developed hypothetical and incentive-aligned experiments to elicit consumers’ product choices. These approaches are much less expensive than test marketing, since consumers’ choices will be realized with zero or a small probability in the experiments. However, this study suggests that these approaches can lead to biased estimation of demand because consumers choose a lower level of effort than in the real purchase.
setting when making a choice. A theory model is first proposed, which shows that consumers’ level of effort increases with the probability of their choice being realized in a choice experiment, and as a result, the manifested demand curve derived from a choice experiment becomes steeper as the realization probability increases. We conduct a large-scale choice experiment on a mobile game platform to validate the theory by randomizing the realization probability and price, and the data support the theoretical hypothesis. Then a structural model is built based on the theory mechanism, and estimated using data from the incentive-aligned conditions in the field experiment. We extrapolate the structural estimates to the real purchase setting, and find that the structural extrapolation is an accurate forecast of the real demand curve. In other words, the structural estimation and extrapolation allow for accurate inference of the actual demand based on choice experiments of small to moderate realization probabilities. Hence this method can help companies determine their optimal price and product plan in a cost-effective way.

Chapter 3 examines firms’ advertising strategies on social media under consumers’ limited attention. Advertising on social media faces a new challenge as consumers are exposed to a vast amount of information and they can actively select which advertisers to follow. A Bayesian learning model is built, which suggests that consumers with limited attention may rationally choose to unfollow a firm. This happens if consumers already know about the firm’s value well and if the firm advertises too intensely. However, intensive advertising may still be the optimal strategy for firms. The intuition is that consumers have noisy beliefs about firms’ product quality \textit{ex ante}. If a firm is perceived as providing low value, it will want to advertise aggressively to change consumers’ perception about its product quality; if a firm is perceived as providing higher value, it will also want to advertise intensively, but in an effort to crowd-out advertising messages from its competitors. Tracking company accounts of 49 TV shows on the most popular tweeting website in China provides empirical evidence that both popular and non-popular firms advertise intensively, although the number of followers does go down when a firm advertises too intensively.

Chapter 4 investigate firms’ cooperative search advertising strategies. Search ad-
vertising is an advertising outlet where consumers’ limited attention has been made salient: the market is highly concentrated and there are only a limited number of advertising slots on search engine platforms that can attract positive attention (number of clicks) from consumers. A manufacturer can sponsor retailers to advertise its products while at the same time compete with them in a position auction with a limited number of slots. Due to the high market concentration of search advertising, a manufacturer and its retailers’ ads compete instead of complement with each other. We consider a manufacturer, who can sponsor its retailers by sharing a fixed percentage (called “participation rate”) of each retailer’s advertising cost while at the same time compete with its retailers and outside advertisers in search advertising. We prescribe the optimal cooperative advertising strategy for the manufacturer, in terms of how many and which retailers to sponsor, as well as when the manufacturer should participate in search advertising directly. We find that it can be optimal for a manufacturer to cooperate with only a subset of its retailers even if they are ex ante symmetric. This finding reflects the manufacturer’s tradeoff between higher demand and higher bidding cost caused by more intensified competition. With two asymmetric retailers, the manufacturer should support the retailer with the higher channel profit per click to get a higher position than the other retailer, which indicates the efficiency of participation rate mechanism. The manufacturer should take a higher position than a retailer when its profit per click via direct sales exceeds the channel profit per click of the retailer. We also investigate how a manufacturer uses both wholesale and advertising contracts to coordinate the channel when search advertising is a major source of demand.
Chapter 2

Prelaunch Demand Estimation

2.1 Introduction

Accurate demand estimation is important for new products to succeed, but is challenging in the absence of historical sales data (e.g., Braden and Oren 1994, Urban et al. 1996, Hitsch 2006, Desai et al. 2007, Bonatti 2011). For decades, researchers have spent considerable effort developing market research strategies to estimate product demand before actual launch. Solutions to date can be classified into three categories. 

*Hypothetical approaches* ask participants to either state their product valuation or make hypothetical product choices which are then used to infer their product valuation (e.g., Miller et al. 2011). *Incentive-aligned approaches* further engage respondents by requiring them to actually purchase the product at the price they are willing to pay with a “realization probability” (e.g., Becker et al. 1964, Ding 2007). *Test marketing*, which can be seen as fully incentive-aligned choice experiments, sells the product in trial markets to gather consumer choice data in real purchase environments.

These solutions are imperfect. Hypothetical approaches are known to generate hypothetical biases (e.g., Frykblom 2000, Wertenbroch and Skiera 2002). Incentive alignment can improve the accuracy of demand estimation compared with the hypothetical approach (e.g., Ding 2007, Miller et al. 2011), but may not recover demand in real purchase settings accurately (e.g., Miller et al. 2011, Kaas and Ruprecht 2006). Test marketing achieves the highest external validity among the three methods (Silk
and Urban 1978). However, the gain in external validity comes at a cost. Other things being equal, the higher the realization probability, the more actual products the company must provide for market research. Besides higher operational costs, more products means greater opportunity costs of selling at suboptimal prices – by definition, the company would not know the optimal price before it is able to estimate demand.\(^1\) As a result, existing market research methods often have to trade off external validity against cost control.

In this chapter, we try to resolve this cost-validity conundrum by developing a theory-based, cost-effective method to estimate the demand of new products. This method is low-cost because it relies only on moderate to small realization probabilities. It is effective because it is able to approximate the demand estimation results of test marketing. Figure 2-1 presents the intended contribution of this chapter.

The idea is as follows. We posit that consumers must make a costly effort to learn their true product valuation. For example, consumers may spend time inspecting product features, searching through alternative options, or thinking about possible usage scenarios (e.g., Shugan 1980, Wathieu and Bertini 2007, Guo and Zhang 2012, Guo 2016, Kleinberg et al. 2017). Whether consumers are willing to make this costly effort depends on the realization probability. Intuitively, if a consumer knows that her product choice is unlikely to be realized, she will have little incentive to uncover her true product valuation and will make her choice based on her prior belief. On the contrary, if a consumer knows that her product choice is for real, she will want to think about how much she truly values the product and make her choice prudently. As a result, there exists a structural relationship between realization probability and manifested demand. Our proposed demand estimation method thus proceeds in two steps: first, estimate this structural relationship using product choice data under smaller realization probabilities; second, use the estimation results to forecast product demand in actual purchase settings.

We formalize the above mechanism with a theory model, in which consumers

---

\(^1\)The company we collaborate with confirmed that it had refrained from test marketing for the same reason.
decide whether they are willing to purchase a product at a given price and a given realization probability. The model predicts that manifested price sensitivity increases with realization probability. To understand the intuition, imagine that the company had offered the product for free. Agreeing to buy the product had been a no-brainer. Now, suppose the company raises the price gradually. As the price approaches a consumer’s prior valuation for the product, she will have a greater incentive to zoom in and think carefully about her true need for the product, and the only change this thinking brings to her decision is to not buy the product. A higher realization probability increases the gravity of the purchase decision and amplifies this negative effect of price on demand. Therefore, it will appear as if consumers are more price-sensitive under higher realization probabilities.

To test the theory model and to evaluate the proposed demand estimation method, we run a large-scale field experiment. We choose the field, as opposed to the lab, in order to minimize factors that may affect external validity other than the realization probability (Simester 2017). We collaborate with a mobile soccer game platform. The new product is a new game package that may enhance user performance. We set four realization probabilities: 0, 1/30, 1/2, and 1, where the 0-probability group is designed to capture the effect of hypothetic approaches and the 1-probability group is designed to mirror test marketing. We randomly assign prices and realization probabilities across users exposed to the experiment.

The experiment results support the theory prediction – consumers are more price-sensitive under higher realization probabilities. We rule out a number of competing explanations of this effect using data from a post-choice survey. Moreover, we obtain process measures of consumers’ decision effort. We find that decision effort increases with realization probability, consistent with the behavioral mechanism underlying the theory prediction.

Having validated the theory foundation of the proposed demand estimation method, we develop a structural model of consumer effort choice and purchase decision based

\[2\text{This choice experiment can be seen as a form of incentive-aligned choice-based conjoint analysis with price being the only product attribute. Marketing practitioners call the hypothetical version of this type of experiment a “Monadic pricing survey.”}\]
on the mechanism developed in the theory. This forms the core of our proposed demand estimation method. More specifically, we estimate the structural model using data from the subsample of smaller realization probabilities (1/30 and 1/2 in the field experiment). To assess the external validity of the proposed method, we use the estimation results to forecast product demand in real purchase settings and compare the forecast against the holdout sample where realization probability equals 1. The structural forecast performs remarkably well. For example, the forecast error in price sensitivity is only 4.49% compared against the holdout sample. Simple extrapolation of data from smaller realization probabilities to actual purchase settings, in contrast, yields forecast errors of around 20%. This suggests that the external validity of the proposed method relies on a detailed, structural understanding of the behavioral process.

The rest of the chapter proceeds as follows. We continue in Section 2.2 with a review of the related literatures. In Section 2.3, we develop a theory model to illustrate the mechanism and to formulate testable predictions. We then present the field experiment in Section 2.4 and discuss model-free evidence of the theory in Section 2.5. In Section 2.6, we draw on the theory to develop and evaluate a method to estimate new product demand based on structural use of choice data from smaller realization probabilities. We conclude in Section 2.7 with discussions of future research.

2.2 Literature Review

Researchers have long been exploring ways to estimate product demand, or equivalently, consumers' product valuation. The most reliable way to estimate demand is to use actual sales data or test market data (Silk and Urban 1978). These types of data have high external validity because they are observed in real purchase settings. However, actual sales data is not available for new products prior to launch, whereas test market data is costly to obtain. Even in the 1970s, the cost of test marketing could surpass one million US dollars. Furthermore, test marketing can be risky for
a firm as it allows competitors to obtain the firm’s product information and respond strategically.

As a result, researchers have developed pre-test-market methods, usually called laboratory or simulated test markets, in which recruited consumers are given the opportunity to buy in a simulated retail store (Silk and Urban 1978). Pre-test-market methods also have high external validity, because they provide a realistic purchase environment and consumers' choices are realized for certain (Silk and Urban 1978, Urban and Katz 1983, Urban 1993). However, pre-test-market methods are still costly – the company incurs not only the logistical costs of actual selling, but also the opportunity cost of selling at potentially suboptimal prices. It can even be infeasible as the company may not have enough product samples to sell at the prelaunch stage.

A different approach to estimating product demand is to use hypothetical surveys or hypothetical choice experiments. Marketing researchers have developed hypothetical choice-based conjoint analysis to measure consumers' tradeoffs among multi-attribute products (see Hauser and Rao 2004, Rao 2014 for an overview), and choice-based conjoint analysis can be augmented to estimate product valuation (e.g., Kohli and Mahajan 1991, Jedidi and Zhang 2002). Economists have used "contingent valuation methods" to estimate people's willingness-to-pay for public goods (Mitchell and Carson 1989), where participants are asked to either state their valuation directly (open-ended contingent evaluation) or to choose whether they are willing to purchase a good at a given price (dichotomous choice experiments).

These hypothetic methods ask participants to answer questions or make choices without actual consequences. As a result, these methods are riskless, low-cost, and widely applicable to concept testing. However, researchers have often found hypothetical methods unreliable. Both hypothetical open-ended contingent valuation and hypothetical choice experiments are found to over-estimate product valuation compared to actual purchases (Diamond and Hausman 1994, Cummings et al. 1995, Balistreri et al. 2001, Lusk and Schroeder 2004, Miller et al. 2011). This happens due to participants' lack of incentive to expend cognitive efforts needed to provide an accurate answer, ignorance of their budget constraints, or tendency to give socially
desirable answers in hypothetical settings (Camerer et al. 1999, Ding 2007).

A stream of literature tries to derive more reliable demand estimates using data from hypothetic methods, but the results are mixed. One solution is to use "calibration techniques" but the calibration factors vary significantly and are specific to the product and the context (Blackburn et al. 1994, Fox et al. 1998, List and Shogren 1998, Murphy et al. 2005). Cummings and Taylor (1999) propose a "cheap-talk" design of questionnaire to reduce the hypothetical bias. List (2001) applies this design to a well-functioning marketplace that auctions off sports cards. He finds that the cheap-talk design mitigates the hypothetical bias, but only for inexperienced bidders.

Another stream of research tries to overcome the hypothetic bias by making participants responsible for the consequences of their choices with a probability, called the "realization probability." Becker et al. (1964) design such a mechanism (hereafter BDM), where a participant is obliged to purchase a product if the price drawn from a lottery is less than or equal to her stated product valuation. The BDM mechanism has been widely used to elicit willingness-to-pay in behavioral decision experiments (e.g., Kahneman et al. 1990, Prelec and Simester 2001, Wang et al. 2007). Wertenbroch and Skiera (2002) compare BDM with hypothetical contingent valuation methods, and find that BDM yields lower willingness-to-pay.

Extending the BDM approach, Ding et al. (2005) and Ding (2007) design an incentive-aligned mechanism for conjoint analysis by replacing stated product valuation with inferred product valuation from conjoint responses. Again, participants must adopt the product they chose with a realization probability. The authors show that incentive-aligned choice-based conjoint analysis outperforms its hypothetical counterpart in out-of-sample predictions of actual purchase behavior. Based on this idea, researchers have developed more-advanced incentive-aligned preference measurement methods (e.g., Park et al. 2008, Ding et al. 2009, Dong et al. 2010, Toubia et al. 2012), and confirm that incentive alignment leads to substantial improvement in predictive performance when compared to hypothetical methods.

However, whether incentive alignment can fully recover actual choices remains a question. Notably, Yang et al. (2015a) find that consumers become more price sensi-
tive as realization probability increases in incentive-aligned choice experiments. Our work reinforces this finding and shows that, indeed, incentive-aligned choice experiments may misrepresent demand in actual purchase settings, although they forecast demand more accurately compared with hypothetical approaches. We propose and empirically validate a theory of decision effort that can explain the bias in incentive-aligned choice experiments. Based on the theory, we develop a method to correct the bias in incentive-aligned experiments, which allows us to estimate the real demand curve in a cost-effective way.

Our decision effort mechanism emphasizes the idea that consumers need to incur a cost to learn their product valuation. Consumers are often uncertain about product performance and individual preferences (e.g., Urbany et al. 1989, Kahn and Meyer 1991, Ariely et al. 2003, Ofek et al. 2007, Wang et al. 2007). It is costly to evaluate product features (e.g., Wernerfelt 1994, Villas-Boas 2009, Kuksov and Villas-Boas 2010) or to think through one’s subjective preferences (e.g., Shugan 1980, Wathieu and Bertini 2007, Guo and Zhang 2012, Huang and Bronnenberg 2015, Guo 2016). In a recent paper, Kleinberg et al. (2017) show that the notion of costly valuation learning has important implications for mechanism design; in particular, it renders the popular increasing-price auction ineffective. As a result of costly preference learning, consumers face an effort-accuracy tradeoff when making choices instead of maximizing decision accuracy as conventionally assumed (Hauser et al. 1993, Payne et al. 1993, Yang et al. 2015b). Wilcox (1993) shows that increased incentives raise subjects’ willingness to incur decision effort and hence influence decision outcomes. Smith and Walker (1993) survey 31 experimental studies and find that higher rewards shift the experiment results towards the prediction of rational models. They also explain this result with effort theory – that is, higher rewards induce agents to exert more cognitive effort. Yang et al. (2015a) find that consumers’ amount of attention to the choice problem increases with the scale of incentives as measured by the realization probability. There is also evidence from neuroeconomics that, when choosing consumer goods, brain activation is stronger and more widespread in the real choice condition than in the hypothetical condition (Camerer and Mobbs 2017).
In this chapter, we further investigate the role of costly decision effort on consumer response in choice experiments, where consumers' effort incentive depends on the probability of their decisions being realized. This allows us to portray the structural relationship between realization probability and product demand. In the following session, we develop a theory model to describe this mechanism and to form testable predictions.

2.3 Theory Model

Consider a market with a unit mass of consumers. The true valuation of a new product, \( v \), is heterogeneous across consumers, following a distribution \( f(\cdot) \) unknown to the firm and consumers (otherwise there is no need for demand estimation). Consider a representative consumer \( i \). She does not know her true product valuation \( v_i \) ex ante. The mean of her prior belief about the true valuation is \( \mu_{0i} \), which can be decomposed as \( \mu_{0i} = v_i + e_i \), where the perception error \( e_i \) follows a distribution \( g(\cdot) \). We assume that \( g(\cdot) \) is continuous and symmetric around 0, is the same across consumers, and that consumers know \( g(\cdot) \) ex ante.

The consumer can expend a decision effort to learn about her true valuation of the product. If the consumer devotes effort \( t \), she will know the true value of \( v_i \) with probability \( t \), and her belief about \( v_i \) stays at \( \mu_{0i} \) with probability \( 1 - t \). An example of a choice context this formulation captures is a consumer’s search of whether she already has a product in her possession that is a good substitute for the new product. Alternatively, we can model the decision effort as smoothly reducing a consumer’s uncertainty about her true product valuation, but the qualitative insight of the theory model remains the same. We write the cost of effort as \( \frac{1}{2}ct^2 \) to capture the idea of increasing marginal cost. We assume that consumer \( i \)'s utility from purchase is \( U_i = v_i - p \) and the consumer has a reservation utility of zero. Thus consumer \( i \) will purchase the product priced at \( p \) if and only if \( \mathbb{E}[v_i] \geq p \), where \( \mathbb{E}[v_i] \) denotes the consumer’s expected value of \( v_i \).

The timing of the choice experiment unfolds as follows. In the first stage, the
consumer observes the price \( p \) and the realization probability \( r \). She is told that if she chooses “willing to buy,” she will have to pay \( p \) and receive the product with probability \( r \), and will pay nothing and not receive the product with probability \( 1 - r \).

In the second stage, the consumer chooses the level of her decision effort, \( t \). In the third stage, the consumer decides whether to choose “willing to buy” based on the outcome of her decision effort. If she is willing to buy, a lottery will be drawn and with probability \( r \) she will pay price \( p \) and receive the product as promised in stage one.

We first derive the optimal effort of the representative consumer. The consumer chooses effort \( t \) to maximize her expected net utility:

\[
E[U(t, \mu_0; p, r)] = r \left( tE[(v_i - p)^+] + (1 - t)(\mu_0 - p)^+ \right) - \frac{1}{2}ct^2, \tag{2.1}
\]

where the expectation is taken over consumer \( i \)’s prior perceived distribution of \( v_i \) before she expends any decision effort.

The first-order condition of \( \partial E[U(t, \mu_0; p, r)]/\partial t = 0 \) yields the optimal effort level:

\[
t^*(\mu_0; p, r) = \frac{r}{c} \left( E[(v_i - p)^+] - (\mu_0 - p)^+ \right) \tag{2.2}
\]

The second-order condition is trivially satisfied for this optimization problem.

We obtain the following comparative statics results.

**Proposition 1** Suppose \( p - \mu_0 \) is strictly within the support of \( g(\cdot) \). The consumer’s optimal decision effort increases with realization probability \( r \), and decreases with the distance between price and her prior belief of her valuation \( |p - \mu_0| \). A greater realization probability amplifies the latter effect.

Proof: see the Appendix.

Intuitively, expending effort helps a consumer make a better informed purchase decision based on her true product valuation. The higher the realization probability, the higher the value of this effort. When realization probability equals 1, the consumer makes the same effort as in real purchase decisions. When realization probability
equals 0, choices become purely hypothetical with no impact on consumer utility, and the consumer makes no effort to learn her product valuation. Moreover, when product price is extremely low (or high), the consumer may trivially decide to buy (or not buy) regardless of her true valuation, which makes the decision effort unnecessary. On the other hand, when price is closer to a consumer's prior valuation, making a purchase decision based on the prior belief alone is more likely to lead to a mistake, and the consumer will want to expend more effort to discover her true valuation.

Knowing consumers' optimal effort decisions, we can derive the “manifested demand” for the product, i.e., the expected fraction of consumers who choose “willing to buy” given price $p$ and realization probability $r$:

$$D(p, r) = \int_{v_i} \int_{e_i} [t^*(v_i + e_i; p, r)1(v_i \geq p) + (1 - t^*(v_i + e_i; p, r))1(v_i + e_i \geq p)]g(e_i)f(v_i)de_idv_i.$$  \hfill (2.3)

We emphasize the notion of manifested demand, as opposed to estimated demand, to highlight the theoretical effect of realization probability on consumer choices. In other words, even if consumers are behaving truthfully based on their expected product valuation and even if there is no empirical error, manifested demand may still differ from actual demand because consumer choices are not fully realized.

Now we investigate how realization probability affects the manifested demand curve. Let $D_p(p, r)$ denote the local slope of the demand curve at price $p$, measuring consumers' price-sensitivity at price $p$. To facilitate presentation, we define the slope of the demand curve at the center of the true valuation distribution, $D_p(p, r)|_{p=\mu_v}$, as the “central” slope of the demand curve. The following proposition summarizes our finding.

**Proposition 2** Suppose $g(\cdot)$ is symmetric around 0 and is weakly decreasing on $(0, \infty)$. Suppose $f(\cdot)$ has a unique mode $\mu_v$, and is weakly increasing on $(-\infty, \mu_v)$ and weakly decreasing on $(\mu_v, \infty)$. Then the following results hold.

$D(p, r)$ is weakly decreasing in the realization probability $r$ when $p > \mu_v$, and is weakly increasing in $r$ when $p < \mu_v$. Denote $Z^-(p) = \{z > 0 : f(p+z) - f(p-z) < 0\}$,
\( Z^+(p) = \{ z > 0 : f(p+z) - f(p-z) > 0 \} \), and \( S_g = \{ z > 0 : g(z) > 0 \} \). For \( p > \mu_v \), if the (Lebesgue) measure of \( Z^-(p) \cap S_g \) is greater than 0, then \( D(p, r) \) strictly decreases in \( r \). For \( p < \mu_v \), if the (Lebesgue) measure of \( Z^+(p) \cap S_g \) is greater than 0, then \( D(p, r) \) strictly increases in \( r \).

The central slope of the demand curve, defined as \( \frac{\partial D(p, r)}{\partial p} \bigg|_{p=\mu_v} \), is weakly decreasing in \( r \). If \( Z(\mu_v) = \{ z > 0 : f(\mu_v+z) + f(\mu_v-z) < 2f(\mu_v) \} \) has a non-empty intersection with the set of \( z \) where \( g(z) \) is strictly decreasing, then when \( r \) increases, the central slope of the demand curve strictly decreases, i.e., the demand curve becomes steeper.

Proof: see the Appendix.

It should be noted that many commonly seen distribution functions satisfy the conditions for the results in the above proposition to hold strictly. We illustrate this fact using normal and uniform distributions, respectively. In the first example, the true valuation \( v_i \) follows a normal distribution \( N(\mu_v, \sigma_v^2) \) and the perception error follows a normal distribution \( N(0, \sigma_v^2) \). Hence the perception error distribution \( g(\cdot) \) is always positive, so that \( S_g = (0, \infty) \). When \( p > \mu_v \), the true valuation distribution \( f(\cdot) \) is strictly decreasing, which implies \( Z^-(p) = (0, \infty) \). When \( p < \mu_v \), \( f(\cdot) \) is strictly increasing, so that \( Z^+(p) = (0, \infty) \). Therefore, the manifested demand \( D(p, r) \) strictly decreases with \( r \) for any \( p > \mu_v \) and strictly increases with \( r \) for any \( p < \mu_v \). We also have \( Z(\mu_v) = (0, \infty) \), which is the same as the set of \( z \) where \( g(z) \) is strictly decreasing. Therefore, the central slope of the demand curve \( \frac{\partial D(p, r)}{\partial p} \bigg|_{p=\mu_v} \) strictly decreases with \( r \).

In the second example, the true valuation \( v_i \) is uniformly distributed on \([-\sigma_v, \mu_v + \sigma_v]\) and the perception error is uniformly distributed on \([-\sigma_0, \sigma_0]\). It follows that \( S_g = (0, \sigma_0) \). When \( p > \mu_v \), \( Z^-(p) = (|p - (\mu_v + \sigma_v)|, p - (\mu_v - \sigma_v)) \), so manifested demand \( D(p, r) \) strictly decreases with \( r \) if and only if \( \sigma_0 > |p - (\mu_v + \sigma_v)| \), that is, if and only if \( \mu_v + (\sigma_v - \sigma_0) < p < \mu_v + (\sigma_v + \sigma_0) \). When \( p < \mu_v \), \( Z^+(p) = (|\mu_v - \sigma_v| - p, (\mu_v + \sigma_v) - p) \), manifested demand \( D(p, r) \) strictly decreases with \( r \) if and only if \( \sigma_0 > |(\mu_v - \sigma_v) - p| \), that is, if and only if \( \mu_v - (\sigma_v + \sigma_0) < p < \mu_v - (\sigma_v - \sigma_0) \).

Figure 2-2 presents how the demand curve changes with realization probability.
assuming $f(\cdot)$ and $g(\cdot)$ are both uniform distributions.\(^4\) We can see that the demand curve rotates at $p = \mu_v$ as realization probability changes. Specifically, demand increases with realization probability for any price below $\mu_v$ and decreases with realization probability for any price above $\mu_v$. As realization probability increases, the demand curve becomes steeper, and consumers appear to be more price sensitive.

2.4 Field Experiment

We run a field experiment to validate the prediction and the mechanism of the theory and to evaluate the proposed demand estimation method. We choose the field experiment approach to minimize threats to external validity such as the decision context. This allows us to identify the effect of realization probability on the external validity of demand estimation methods.

We collaborate with a top mobile platform of soccer games in China. Founded in 2013, the platform currently hosts 80,000 daily active users, generating 2 million US dollars in monthly revenue. In the game, each user manages a soccer team and the goal is to win as many times as possible. A team’s likelihood of winning depends on the number of high-quality players it enlists. The new product we sell in the field experiment is a “lucky player package” that consists of six high-quality players. This player package had never been sold on the game platform prior to the experiment.

We want to randomize realization probability and price. We set four different realization probabilities: 0, 1/30, 1/2, and 1. The 0-probability group is designed to replicate hypothetical surveys, and the 1-probability group is meant to mirror the actual purchase setting. We add two interim realization probability groups because the proposed demand estimation method needs at least two realization probability levels for empirical identification and we choose only two for a conservative evaluation of the method’s predictive power. We assign a 1/2-probability group to observe the effect of moderate realization probabilities. In addition, we create a 1/30-probability group because, in many experiments, the rule-of-thumb is to recruit 30 subjects per

\(^4\)The figure is plotted under the condition that $\sigma_0 > \sigma_v$. 

30
condition. For future applications of the proposed demand estimation method using 30 subjects per condition, this realization probability can be more tangibly interpreted as one out of the 30 subjects buying the product for real, which makes the experiment looks more trustworthy than using a smaller realization probability.

We set five price levels, measured as 1600, 2000, 2400, 2800, or 3200 “diamonds,” which is the currency used in the game. Users need to pay real money to obtain diamonds. The exchange rate is about 1 US dollar for 100 diamonds. We discuss with the company to make sure this price range is reasonable and at the same time the gap between prices is large enough to elicit different purchase rates. The five price levels, combined with the four realization probabilities, lead to 20 conditions for the experiment. Once a user enters the experiment, she is randomly assigned to one of the conditions.

Each condition presents the user with a screen of the choice task. (Figure 2-3 presents the screen for the 1/30-probability group.) On this screen, we inform the user that she has a chance to purchase a lucky player package at price $p$ and ask her to choose between “willing to purchase” and “not willing to purchase.” For the 0-probability group, we inform the user that this is a hypothetic survey and no actual transaction will take place. For the 1-probability group, we notify the user that she has the chance to actually purchase the package. For the interim probability groups, we explain that if the user chooses “willing to purchase,” a lottery will be drawn and there is probability $r$ that she will actually receive the player package and will be charged the price $p$ automatically. If the user chooses “not willing to purchase” or does not win the lottery, she will not receive the player package and will not be charged anything. Users can click on the player package and see the set of players contained therein (see Figure 2-4). They can also click on each player and see what skills the player has. After making the purchase decision, the user will be directed to a follow-up survey, which we designed to obtain auxiliary data to test the theory.

The experiment took place from 12AM, December 2, 2016 to 12PM, December 4, 2016. We randomly selected half of the platform’s Android servers, and all users on these servers automatically entered the experiment once they accessed the game.
during the period of the experiment. A total of 5,420 users entered the experiment, 271 in each condition. Among these users, 3,832 (70.7%) completed the choice task. Among those who completed the choice task, 2,984 (77.87%) filled out the survey. Table 2.1 reports the number of users assigned to each probability and price group, and the number that completed the choice task or the survey. We notice higher completion rates in the 0-probability group. However, reassuringly, for all groups with positive realization probabilities, neither completing the choice task nor completing the survey is significantly correlated with the assigned realization probability ($Corr = -0.0163, p = 0.2993$ and $Corr = -0.0195, p = 0.3080$, respectively) or the assigned price level ($Corr = 0.0023, p = 0.8660$ and $Corr = 0.0265, p = 0.1013$, respectively).

For each user who completed the choice task, we collect data on her characteristics at the time of the experiment, including the number of diamonds the user has (Diamond) and the VIP level of the user (VIP). The VIP level is determined by how much money the user has spent in the game. Table 2.2 presents the summary statistics. We can see that Diamond is a highly right-skewed variable, hence we convert it into a new variable $\text{Log-Diamond} = \log(\text{Diamond} + 1)$ and will use this new measure in subsequent analysis.

### 2.5 Reduced-Form Evidence

In this section, we present reduced-form evidence of the theory prediction and of the decision effort mechanism, using data from the field experiment.

We first examine aggregate-level demand. By demand, we mean the proportion of users who chose “willing to purchase” out of those who completed the choice task in each condition. Figure 2-5 shows how aggregate-level demand changes with price under each realization probability. We see a pattern – as the realization probability increases, demand seems to decrease faster with price; in addition, the overall level of demand seems to decrease with realization probability.

To verify these observations, we fit a linear demand curve for each realization probability group by regressing individual-level purchase decisions on price. The de-
pendent variable Purchase equals 1 if the user chooses “willing to purchase” and 0 if the user chooses “not willing to purchase.” For the ease of presentation, we normalize the five price levels to 4, 5, 6, 7, 8 respectively in this regression and subsequent analysis. Table 2.3 presents the estimated price coefficient and intercept of the demand curves. The slope of demand curve (absolute value of the price coefficient) increases with the realization probability, consistent with the prediction of our theory.

We further examine how individual-level purchase decisions are jointly influenced by price and realization probability. Column (1) of Table 2.4 shows that the likelihood of purchase decreases with price, as expected. It also decreases with realization probability, an effect we will comment on later. In column (2), we add the interaction term of price and realization probability, and this interaction term has a significantly negative coefficient. This means that users are more price sensitive with higher realization probabilities, consistent with the theory prediction. In column (3), we further control for user characteristics, namely, Log-Diamond and VIP. Having more diamonds and having lower VIP status are associated with higher purchase rates. Again, users become more price sensitive as realization probability increases.

A comment on the level of demand is in order. According to our theory model, for any given price, the level of demand decreases with realization probability only when prices are higher than users’ average product valuation. As a further test of the theory, in the post-choice survey we ask users to rate how they perceive the price of the product on a scale from 1 (very low) to 5 (very high). Indeed, the answers confirm that users view the price as relatively high; the mean answer is 3.99, significantly higher than the neutral level of 3 (t = 52.83, p < 0.001).

So far, data support our theory prediction that the demand curve becomes steeper as the realization probability increases. Next we examine whether the change in the slope of the demand curve is driven by the decision effort mechanism we propose. We need a measure of users’ decision effort and examine how it changes with price and realization probability. Measuring decision effort is difficult (Bettman et al. 1990), and we try to do so using two measures. First, our experiment setting allows us to gauge how much a user has learned about the product. More specifically, in the
post-choice survey, we ask each user to answer “which of the following soccer players was not included in the player package.”

If a user has carefully thought about her valuation of the player package, presumably she should know its content. We let the effort measure equal 1 if the user provides the correct answer (there is only one correct answer), and 0 otherwise. As a second proxy of decision effort, we draw upon the classic measure of decision time (Wilcox 1993). We record decision time as the number of seconds it takes from the point the user first arrives at the choice task page to the point she makes a choice. The decision time variable is highly right-skewed with some extremely large values, hence we take a log transformation of it for subsequent analysis. Table 2.5 reports the summary statistics of these effort measures.

As a direct mechanism test, we regress the two measures of decision effort on realization probability and price. Table 2.6 presents the result. For both measures of effort, users’ effort input increases with realization probability, consistent with Proposition 1. Effort also decreases with price, although the effect is insignificant. The negative effect of price on effort echoes the survey result that users perceive the price of the player package as relatively high. As price increases from an already-high level, not to buy becomes a clearer decision regardless of a user’s true product valuation, which makes effort less needed. This result is again consistent with Proposition 1.

2.6 Evaluating the Demand Estimation Method

In this section, we use data from the field experiment to evaluate the proposed demand estimation method. The core of the method is a structural model of consumer product choice based on the decision effort mechanism we propose. We estimate the structural model drawing on choice data from the 1/2-probability and 1/30-probability groups, leaving data from the 1-probability group as the holdout sample. We then use the structural estimates to forecast demand in actual purchase settings (i.e., settings where realization probability equals 1), and compare the forecast with demand in the holdout sample. To assess the value of having a theory-based model, we also compare
the structural forecast with simple extrapolation of demand from the 1/2-probability and 1/30 probability groups.

2.6.1 A Structural Model of Consumer Product Choice

The structural model captures the same behavioral process as the theory model of Section 2.3 but operationalizes it to match the empirical context. For a conservative evaluation of the proposed demand estimation method, we strive to keep the structural model parsimonious.

We let user $i$'s true valuation of the product be

$$v_i = b_0 + b_1 \log(\text{Diamond}_i) + b_2 \text{VIP}_i + \epsilon_{vi}, \quad (2.4)$$

where $\epsilon_{vi}$ represents the unobserved heterogeneity in consumers' true product valuation, which follows a normal distribution $N(0, \sigma^2_v)$. Recall that $\log(\text{Diamond}_i) = \log(\text{Diamond}_i + 1)$, where $\text{Diamond}_i$ is the number of diamonds user $i$ has at the time of the experiment. $\text{VIP}_i$ denotes the VIP level of user $i$ at the time of the experiment, which is determined by how much this user has spent in the game. For the ease of interpreting the parameter estimates, we scale both $\log(\text{Diamond}_i)$ and $\text{VIP}_i$ to $[0, 1]$ by dividing each variable by its maximum value. We conjecture that a user with more diamonds at hand is likely to have a higher willingness-to-pay for the product. The sign of VIP is a priori ambiguous. A user who has spent a lot may be more likely to spend on the new product out of habit, or less likely to spend because she already owns high-quality players contained in the player package.

User $i$'s prior belief about her product valuation, $\mu_{vi}$, follows the normal distribution $N(v_i, \sigma^2_{vi})$, where the prior uncertainty $\sigma_{vi}$ is operationalized as

$$\sigma_{vi} = \exp(a_0 + a_1 \text{VIP}_i). \quad (2.5)$$

We use the exponential function here to guarantee that $\sigma^2_{vi}$ is positive. We expect VIP to have a negative coefficient because, other things being equal, more spending
arguably means greater experience with the game and hence less uncertainty about product valuation.

Knowing her prior mean valuation of the product $\mu_{0i}$ and her prior uncertainty $\sigma_{0i}$, user $i$ can derive her optimal level of effort in the same way as in the theory model:

$$t_i = \min \left\{ \frac{r_i}{c_i} \left( \mathbb{E}[(v_i - p_i)^+] - (\mu_{0i} - p_i)^+ \right), 1 \right\}$$

(2.6)

where the expectation is taken over consumer $i$'s belief that $v_i \sim N(\mu_{0i}, \sigma_{0i}^2)$. $p_i$ and $r_i$ are the price and realization probability that user $i$ is randomly assigned in the experiment. We restrict effort $t_i$ to be no larger than 1 because it is defined as the probability that the consumer will learn her true valuation (see Section 2.3). We further operationalize user $i$'s effort cost $c_i$ as

$$c_i = \exp(c_1 + c_2 e_{ci})$$

(2.7)

where $e_{ci} \sim N(0,1)$. The exponential transformation guarantees that effort cost is positive. The $e_{ci}$ term allows effort cost to be heterogeneous among consumers.

Given her effort level $t_i$, with probability $t_i$, user $i$ learns her true product valuation $v_i$ and should buy the product if $v_i \geq p_i$. With probability $1 - t_i$, user $i$ retains her prior belief and should buy if $\mu_{0i} \geq p_i$. We assume that users have a response error when making purchase decisions, and the response error follows i.i.d. standard Type I extreme value distribution. It follows that user $i$’s probability of choosing “willing to buy” is given by the standard logit formula:

$$\Pr(\text{Buy}_i = 1) = \frac{\exp(v_i - p_i)}{1 + \exp(v_i - p_i)} + (1 - t_i) \frac{\exp(\mu_{0i} - p_i)}{1 + \exp(\mu_{0i} - p_i)}.$$  

(2.8)

The log-likelihood function of the observed purchase decision data is

$$LL = \sum_{i=1}^{N} \left[ 1(\text{Buy}_i = 1) \log \Pr(\text{Buy}_i = 1) + 1(\text{Buy}_i = 0) \log (1 - \Pr(\text{Buy}_i = 1)) \right].$$

(2.9)

The above formulation of the log-likelihood function does not rely on actual data
on consumer effort choices. Instead, it calculates effort choices based on model parameters following the process described in the theory model. Recall that we do have measures of effort from the field experiment. We could in theory incorporate these measures to derive additional moments for the estimation. Again, for a conservative evaluation of the proposed demand estimation method, we deliberatively avoid relying on effort data for model calibration. In fact, we would like the model to forecast well in the absence of effort data, which will lower the data requirement and broaden the applicability of the proposed demand estimation method.

2.6.2 Estimation Procedure

The structural model is estimated using the simulated maximum-likelihood estimation approach (Train 2009). For a given set of parameter values, we calculate the purchase probability of each user averaged over a large number of pre-simulated random draws, and then calculate the log-likelihood by summing up the log-likelihood of each user. The estimated parameter values are found by maximizing the simulated log-likelihood. The standard error is estimated using the inverse of Hessian matrix at the estimated parameter values. We present the detailed estimation procedure in the Appendix.

We use data from conditions where realization probability equals 1/30 or 1/2 to estimate the model parameters. We leave the 1-probability condition as the hold-out sample to assess the forecast ability of the proposed demand estimation method. We do not use data from the 0-probability condition in estimation for two reasons. First, our theory predicts that the consumer can make any choice decision in this purely hypothetic setting. Thus we need to make further tie-breaking assumptions to interpret data from this condition. For instance, we could estimate an additional parameter that captures consumers' tendency to act on their true beliefs when indifferent. The identification of this parameter, however, still has to rely on information from the 1/2-probability and 1/30-probability groups. Second, we include the 0-probability group in the field experiment to assess how hypothetical surveys perform compared with other demand estimation methods within the same empirical context. Application of the proposed demand estimation method, however, does not require data
from the 0-probability group. We exclude this group from the estimation to keep the method "lean" in terms of data requirement. Nevertheless, we verify that including the 0-probability group does not change the estimation results significantly.

2.6.3 Identification

The parameters we need to estimate are the constant and coefficients in users' true valuation \( (b_0, b_1, b_2) \), the constant and coefficient in users' prior uncertainty \( (a_0, a_1) \), the parameters determining users' effort cost \( (c_1, c_2) \), and the standard deviation of users' unobserved heterogeneity in true valuation \( \sigma_v \). \( b_0 \) is identified from the overall level of demand. \( b_1, b_2 \) are identified from the exogenous variation in users' characteristics (Log-Diamond and VIP) and their difference in purchase tendency. \( (a_0, a_1) \) and \( (c_1, c_2) \) together determine users' optimal effort and thus determine the difference in intercepts and slopes of demand curves under different realization probabilities. \( (a_0, a_1) \) can be separately identified from \( (c_1, c_2) \) because it not only enters into the formulation of optimal effort, but also determines the variance of prior beliefs and hence determines the slope of the demand curve when no effort is expended. \( a_1 \) is separately identified from how the gap in demand curves differs for users with different VIP levels. Since every user only makes one purchase decision in our data, the unobserved heterogeneity \( \sigma_v \) cannot be identified using the systematic difference in users' behavior. The estimated value actually measures the part of heterogeneity in valuation that cannot be captured by \( b_1 \text{Log-Diamond}_i + b_2 \text{VIP}_i \) and is identified from the slope of the demand curve.

2.6.4 Estimation Results

Table 2.7 reports the parameter estimates and their standard errors. Users' true valuation of the product, not surprisingly, increases with the amount of currency they own in the game \( (b_1 > 0, p < 0.001) \). Users' true valuation of the product also decreases with the VIP level \( (b_2 < 0, p = 0.08) \). One explanation is that users with higher VIP levels tend to have spent more in the game and, as a result, are more likely
to have staffed their teams with high-quality players already, so that the new player package we sell is of less value to them. In addition to these observed variations, there is significant unobserved heterogeneity in users' true valuation ($\sigma_v > 0, p < 0.10$). The magnitude of this unobserved heterogeneity is nontrivial given that Log-Diamond and VIP are both normalized to $[0, 1]$ in the estimation. Moving on to prior uncertainty, as expected, users with higher VIP levels are more certain about their valuation of the product ($a_1 < 0, p = 0.06$). Finally, the effort cost parameter $c_1$ is positive and significant ($p < 0.01$), but the heterogeneity term $c_2$ is almost zero. These results suggest that decision effort is costly, and similarly costly to all users in this empirical context.

To put the estimation results in context, we calculate each user’s optimal effort level ($t_i^*$) and total cost of effort ($c_i t_i^2 / 2$) based on the parameter estimates. Table 2.8 presents the mean and standard deviation of these measures by experiment condition. Both the optimal effort level and effort cost increase with realization probability. They equal zero in the hypothetical condition by definition. In the actual choice condition with realization probability equal to 1, users on average spend an effort of 0.447 out of a normalized range of 0 to 1. This translates into an average effort cost of $12.77 in the actual choice condition. Recall that the average price of the product in the experiment is $24. Roughly speaking, users on average incur a cost equivalent to 53% of product price to learn their product valuation in the actual choice condition. This is a sizable cost. As such, when their product choice is only realized with 50% probability, users are only willing to spend $4.15 of effort to learn their product valuation; when the realization probability is further reduced to $1/30$, users end up making merely 2 cents’ worth of effort.

### 2.6.5 Forecasting Demand in Real Purchase Settings

Based on the parameter estimates, we simulate the purchase decision of each individual assuming realization probability equals 1 in the structural model (see the Appendix for details of the simulation). The simulation results form our forecast of demand in real purchase settings. We compare the forecast against actual demand in
the hold-out sample, that is, the 1-probability group we have set aside. To put the forecast in context, we also compare it with actual demand in the 1/30-probability and 1/2-probability groups. For the ease of visualization, we fit a linear demand curve based on the forecast and for each of these probability groups.

Figure 2-6 presents the comparison. Consistent with prior findings from the literature, the hypothetical approach (i.e., the 0-probability group) performs poorly; it overestimates demand considerably and it underestimates the degree of price sensitivity. Incentive alignment (i.e., the 1/30-probability and 1/2-probability groups) improves forecast accuracy, especially if realization probability is higher (1/2 as opposed to 1/30). The structural forecast generates a demand curve the closest to the actual demand curve of the hold-out sample.

A natural question at this point is whether one can forecast demand using simple extrapolation methods instead of a complex structural model – one can use data from the two interim probability groups and extrapolate to the case where realization probability equals 1. To answer this question, we examine two extrapolation methods. The first method, linear extrapolation, estimates an individual-level linear regression model of purchase decisions as a function of price, realization probability (1/30 or 1/2), their interaction terms, as well as observed user characteristics (Log-Diamond and VIP). The estimates then allow for extrapolation of purchase decisions to the case of realization probability being 1. The second method is based on the same idea but replaces the linear model with a logit model to capture the binary nature of the purchase decision.

We plot the fitted demand curve based on linear extrapolation in Figure 2-6. The logit version is very close and is omitted from the figure for the clarity of presentation. This fitted demand curve is closer to actual demand than the raw demand curves in the two interim probability groups. However, it performs notably worse than structural extrapolation. This is true although these simple extrapolation methods use exactly the same data as the structural forecast. Structural forecast performs better here because it uses the data in a better way by imposing a theoretically sound and empirically valid behavioral process.
We formally quantify and compare the forecast performance of the various demand estimation methods presented in Figure 2-6. The first column of Table 2.9 reports estimated price sensitivity, the metric we have focused on throughout the chapter. The hypothetical method performs the worst. It generates a price sensitivity estimate nearly 70% lower than its actual value in real purchase settings. Incentive alignment with realization probabilities of 1/30 and 1/2 perform better, but still produce noticeable forecast errors. The forecast error reduces to around 20% for linear and logit extrapolations, and is only 4.62% for the structural forecast.

Besides price sensitivity, Figure 2-6 suggests that the various forecast methods may have overestimated the level of demand. We perform a likelihood ratio (LR) test to determine the overall fit of forecast demand to actual demand. For each forecast method \( k \in \{\text{Prob} = 0, \text{Prob} = 1/30, \text{Prob} = 1/2, \text{Linear, Logit, Structural}\} \), its likelihood ratio is calculated as \( LR_k = -2[LL_{\text{Pooled}} - (LL_{\text{Actual}} + LL_k)] \), where \( LL \) represents the log-likelihood of a linear demand curve based on observed purchases or simulated purchase probability. The likelihood ratio follows a chi-square distribution with degrees of freedom equal to the difference in the number of free parameters, which is 2 in our case. The second column of Table 2.9 reports the likelihood ratio of each forecast method relative to actual demand. We cannot reject the null hypothesis that the structural forecast coincides with actual demand, whereas all the other methods significantly deviate from actual demand at the \( p < 0.01 \) level.

To see the practical value of the proposed demand estimation method, we calculate the optimal price implied by the actual demand and by the various forecast methods, respectively. Suppose the fitted demand curve is \( D(p) = \alpha_0 + \alpha_1p \) and the marginal cost of production is \( mc \), then the optimal price is \( p^* = \frac{mc}{\alpha_0 + \alpha_1} \). The third column of Table 2.9 presents the optimal price implied by the coefficients of each demand curve, assuming a marginal cost of 0 (which is a reasonable assumption considering the digital nature of the product we sell in the field experiment). Furthermore, by plugging the optimal price recommended by each method into the actual demand curve, we can calculate the expected demand and thus the expected profit if that price is charged. Comparing the expected profit to the optimal profit in the actual demand
condition, we obtain the percentage profit loss associated with each method. The last column of Table 2.9 presents the results. By recommending an excessively high price, the hypothetical method leads to no demand and a 100% profit loss in this particular empirical setting. Incentive alignment performs better. Simple extrapolations based on incentive-alignment data introduces further improvement, with logit extrapolation reducing the profit loss to 5.77%. However, structural extrapolation takes the forecast accuracy to yet another level. It cuts the profit loss to 0.58%, which is about one tenth of the loss under logit extrapolation.

To summarize, our proposed demand estimation method forecasts actual demand reasonably well. It forecasts actual demand significantly better than hypothetic surveys, and incentive-aligned choice experiments of moderate to small realization probabilities. Moreover, it forecasts actual demand significantly better than simple extrapolations of these incentive-aligned choice experiments to real purchase settings. We have strived to keep the structural model parsimonious for this first test of the proposed demand estimation method. The method's forecast accuracy may further improve if we enrich the structural model by, for instance, allowing for more sources of consumer heterogeneity.

Finally, it is worth noting that the proposed demand estimation method represents significant savings in market research costs compared with test marketing. For a conservative assessment, let us abstract away from the higher logistic overhead of test marketing and focus solely on the opportunity cost of selling products at suboptimal prices. Suppose a sample size of $n$ purchase decisions is required. To run test marketing, the company must prepare to sell $n$ products at suboptimal prices. To implement the proposed demand estimation method, let us assume the company gathers $n/2$ observations from the 1/30-probability and 1/2-probability groups each. It follows that the company needs to prepare $n/60$ products for the 1/30-probability group and $n/4$ product for the 1/2-probability group. This translates into a 73% savings in products required compared with test marketing. The company may be able to save even more by optimizing the allocation of sample size across probability conditions, and by choosing the probability values wisely.
2.7 Concluding Remarks

In this chapter, we proposed a theory-based, cost-effective method to estimate product demand prior to launch. The proposed method draws on data from incentive aligned choice experiments, but imposes structure on the data via a decision effort mechanism. This method allows us to rely on moderate to small realization probabilities to form reasonably accurate forecast of demand in actual purchase settings.

There are several ways to extend this research. First, the current study chooses realization probabilities somewhat arbitrarily for a first test of the proposed method. As mentioned in the previous section, it will be interesting to investigate the optimal choice of realization probabilities, as well as the size of each probability group. Second, we focus on price as the only product attribute for a clean illustration of our proposed method. It will be rewarding to extend the method to settings of multi-attribute products. Last but not least, formally modeling the role of decision effort on choices may shed new light on the the design of choice experiments in other contexts.
2.A Appendix

2.A.1 Figures

Figure 2-1: Intended Contribution of the Chapter

External validity

Our proposed method

Incentive-aligned approaches

Test marketing

Hypothetical approaches

Cost
Figure 2-2: Illustration of Theory Prediction

Notes. Assuming prior valuation and perception errors are uniformly distributed.

Figure 2-3: Screenshot of the Choice Task

You have the opportunity to buy a new product. If you choose "willing to buy," the system will run a lottery, and with 1/30 chance you will get this product at the price below. If the lottery fails or you choose "not willing to buy," no transaction will happen.
Figure 2-4: Content of the Player Package
Figure 2-5: Aggregate-Level Experiment Result
Figure 2-6: Compare Actual and Forecast Demand Curves
### Table 2.1: Number of Users by Realization Probability and by Price

<table>
<thead>
<tr>
<th>Probability</th>
<th>Entered the Experiment</th>
<th>Made the Choice</th>
<th>Completed the Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1355</td>
<td>1095</td>
<td>882</td>
</tr>
<tr>
<td>1/30</td>
<td>1355</td>
<td>920</td>
<td>708</td>
</tr>
<tr>
<td>1/2</td>
<td>1355</td>
<td>922</td>
<td>723</td>
</tr>
<tr>
<td>1</td>
<td>1355</td>
<td>895</td>
<td>671</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Price</th>
<th>Entered the Experiment</th>
<th>Made the Choice</th>
<th>Completed the Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>1084</td>
<td>774</td>
<td>599</td>
</tr>
<tr>
<td>2000</td>
<td>1084</td>
<td>757</td>
<td>575</td>
</tr>
<tr>
<td>2400</td>
<td>1084</td>
<td>764</td>
<td>589</td>
</tr>
<tr>
<td>2800</td>
<td>1084</td>
<td>761</td>
<td>603</td>
</tr>
<tr>
<td>3200</td>
<td>1084</td>
<td>776</td>
<td>618</td>
</tr>
</tbody>
</table>

### Table 2.2: Summary Statistics of User Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>3134.44</td>
<td>5498.09</td>
<td>1614.00</td>
<td>0</td>
<td>150969</td>
<td>3832</td>
</tr>
<tr>
<td>Log-Diamond</td>
<td>7.09</td>
<td>1.64</td>
<td>7.39</td>
<td>0</td>
<td>12</td>
<td>3832</td>
</tr>
<tr>
<td>VIP</td>
<td>3.00</td>
<td>3.10</td>
<td>2.00</td>
<td>0</td>
<td>15</td>
<td>3832</td>
</tr>
</tbody>
</table>

Notes. The sample consists of all users who completed the choice task.
### Table 2.3: Individual-Level Experiment Result

<table>
<thead>
<tr>
<th></th>
<th>Purchase (Prob=0)</th>
<th>Purchase (Prob=1/30)</th>
<th>Purchase (Prob=1/2)</th>
<th>Purchase (Prob=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-0.0197*</td>
<td>-0.0305***</td>
<td>-0.0413***</td>
<td>-0.0655***</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0113)</td>
<td>(0.0115)</td>
<td>(0.0110)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.753***</td>
<td>0.623***</td>
<td>0.677***</td>
<td>0.725***</td>
</tr>
<tr>
<td></td>
<td>(0.0626)</td>
<td>(0.0700)</td>
<td>(0.0713)</td>
<td>(0.0698)</td>
</tr>
<tr>
<td>N</td>
<td>1095</td>
<td>920</td>
<td>922</td>
<td>895</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.002</td>
<td>0.007</td>
<td>0.013</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### Table 2.4: Price Sensitivity Increases with Realization Probability

<table>
<thead>
<tr>
<th></th>
<th>(1) Purchase</th>
<th>(2) Purchase</th>
<th>(3) Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-0.0372***</td>
<td>-0.0219***</td>
<td>-0.0218***</td>
</tr>
<tr>
<td></td>
<td>(0.00556)</td>
<td>(0.00753)</td>
<td>(0.00752)</td>
</tr>
<tr>
<td>Probability</td>
<td>-0.226***</td>
<td>0.0304</td>
<td>0.00998</td>
</tr>
<tr>
<td></td>
<td>(0.0191)</td>
<td>(0.0844)</td>
<td>(0.0838)</td>
</tr>
<tr>
<td>Price x Probability</td>
<td>-0.0427***</td>
<td>-0.0391***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0134)</td>
<td></td>
</tr>
<tr>
<td>Log-Diamond</td>
<td></td>
<td></td>
<td>0.0217***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00499)</td>
</tr>
<tr>
<td>VIP</td>
<td>-0.0159***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00255)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.773***</td>
<td>0.681***</td>
<td>0.574***</td>
</tr>
<tr>
<td></td>
<td>(0.0349)</td>
<td>(0.0464)</td>
<td>(0.0587)</td>
</tr>
<tr>
<td>$N$</td>
<td>3832</td>
<td>3832</td>
<td>3832</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.044</td>
<td>0.046</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### Table 2.5: Summary Statistics of Effort Measures

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having the Correct Answer</td>
<td>0.55</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>2984</td>
</tr>
<tr>
<td>Log Decision Time</td>
<td>2.91</td>
<td>2.65</td>
<td>-1</td>
<td>10</td>
<td>3832</td>
</tr>
</tbody>
</table>

The variable "Having the Correct Answer" is recorded for all users who completed the survey. Decision time is recorded for all users who completed the choice task.
Table 2.6: Effort Increases with Realization Probability

<table>
<thead>
<tr>
<th>Probability (1)</th>
<th>Effort as Correct Answer 0.0550** (0.0227)</th>
<th>Effort as Decision Time 0.274** (0.109)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-0.00314 (0.00638)</td>
<td>-0.0375 (0.0303)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.552*** (0.0402)</td>
<td>3.039*** (0.191)</td>
</tr>
</tbody>
</table>

N 2984 3832
adj. $R^2$ 0.001 0.002

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Estimation Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>-3.522***</td>
<td>1.196</td>
</tr>
<tr>
<td>$b_1$</td>
<td>14.387***</td>
<td>1.824</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-8.812*</td>
<td>5.036</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.645*</td>
<td>0.370</td>
</tr>
<tr>
<td>$a_0$</td>
<td>4.872***</td>
<td>0.343</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-3.109*</td>
<td>1.672</td>
</tr>
<tr>
<td>$c_1$</td>
<td>3.310***</td>
<td>1.026</td>
</tr>
<tr>
<td>$c_2$</td>
<td>1.534e-07</td>
<td>0.775</td>
</tr>
</tbody>
</table>

N 1842
Log-Likelihood -1238.69

Notes. The sample consists of conditions in which realization probability equals 1/30 or 1/2. Prices are normalized to {4, 5, 6, 7, 8}. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Estimated Effort Level and Effort Cost

<table>
<thead>
<tr>
<th>Estimated Effort Level $(t^*_i)$</th>
<th>Estimated Effort Cost $(c_1t^*_i/2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Prob=0</td>
<td>0</td>
</tr>
<tr>
<td>Prob=1/30</td>
<td>0.017</td>
</tr>
<tr>
<td>Prob=1/2</td>
<td>0.250</td>
</tr>
<tr>
<td>Prob=1</td>
<td>0.447</td>
</tr>
</tbody>
</table>

Notes. Effort level is normalized between 0 and 1.
Table 2.9: Performance Comparison across Demand Estimation Methods

<table>
<thead>
<tr>
<th>Prob</th>
<th>Price Sensitivity</th>
<th>Likelihood Ratio (vs. Actual)</th>
<th>Optimal Price</th>
<th>Profit Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob=0</td>
<td>0.0197</td>
<td>197.43***</td>
<td>$76.58</td>
<td>100%</td>
</tr>
<tr>
<td>Prob=1/30</td>
<td>0.0305</td>
<td>28.52***</td>
<td>$40.87</td>
<td>71.84%</td>
</tr>
<tr>
<td>Prob=1/2</td>
<td>0.0413</td>
<td>21.77***</td>
<td>$32.72</td>
<td>22.96%</td>
</tr>
<tr>
<td>Linear</td>
<td>0.0512</td>
<td>16.95***</td>
<td>$28.45</td>
<td>8.19%</td>
</tr>
<tr>
<td>Logit</td>
<td>0.0543</td>
<td>15.87***</td>
<td>$27.43</td>
<td>5.77%</td>
</tr>
<tr>
<td>Structural</td>
<td>0.0625</td>
<td>2.93</td>
<td>$23.81</td>
<td>0.58%</td>
</tr>
<tr>
<td>Prob=1 (Actual)</td>
<td>0.0655</td>
<td>0</td>
<td>$22.12</td>
<td>0</td>
</tr>
</tbody>
</table>

* $p < 0.10$, i.e., LR $> q_2^2(0.9, 2) = 4.6052$;
** $p < 0.05$, i.e., LR $> q_2^2(0.95, 2) = 5.9915$;
*** $p < 0.01$, i.e., LR $> q_2^2(0.99, 2) = 9.2103$. 

2.A.3 Proof of Proposition 1

Consumer \(i\) observes \(\mu_{oi}, g, p,\) and \(r\) when choosing her optimal effort level. She also knows that \(v_i = \mu_{oi} - e_i\), and thus \(v_i - p \geq 0\) is equivalent to \(e_i \leq \mu_{oi} - p\). Rearranging Equation (2.2), the consumer’s optimal effort level is

\[
t^*(\mu_{oi}; p, r) = \frac{r}{c} \left( \int_{-\infty}^{\mu_{oi} - p} (\mu_{oi} - e_i - p) g(e_i) de_i - (\mu_{oi} - p)^+ \right)
\]

\[
= \begin{cases}
\frac{r}{c} \left( \int_{-\infty}^{\mu_{oi} - p} (\mu_{oi} - e_i - p) g(e_i) de_i \right) & \text{if } \mu_{oi} - p < 0,
\frac{r}{c} \left( \int_{\mu_{oi} - p}^{\infty} [e_i - (\mu_{oi} - p)] g(e_i) de_i \right) & \text{if } \mu_{oi} - p \geq 0,
\end{cases}
\tag{2.A10}
\]

where the second case in (2.A10) is derived from the fact that \(\mu_{oi} - p = \int_{-\infty}^{\infty} (\mu_{oi} - p - e_i) g(e_i) de_i\) which holds because \(\int_{-\infty}^{\infty} g(e_i) de_i = 1\) by definition and \(\int_{-\infty}^{\infty} e_i g(e_i) de_i = 0\) by assumption.

First, consider the case of \(\mu_{oi} - p < 0\). When \(e_i < \mu_{oi} - p\), the first term of the integrand in (2.A10), \(\mu_{oi} - e_i - p\), is positive. Because \(g(\cdot)\) is continuous, as long as \(\mu_{oi} - p\) is strictly within the support of \(g(\cdot)\), there exists \(e_i \in (-\infty, \mu_{oi} - p)\) such that \(g(e_i) > 0\) and that the integral in (2.A10) is positive, which implies that \(t^*(\mu_{oi}; p, r)\) increases with \(r\).

Meanwhile, we obtain

\[
\frac{\partial t^*(\mu_{oi}; p, r)}{\partial (\mu_{oi} - p)} = \frac{r}{c} \int_{-\infty}^{\mu_{oi} - p} g(e_i) de_i
\]

and

\[
\frac{\partial^2 t^*(\mu_{oi}; p, r)}{\partial (\mu_{oi} - p) \partial r} = \frac{1}{c} \int_{-\infty}^{\mu_{oi} - p} g(e_i) de_i.
\]

Both terms are positive as long as \(\mu_{oi} - p\) is strictly within the support of \(g(\cdot)\). This means \(t^*(\mu_{oi}; p, r)\) decreases with \(|\mu_{oi} - p|\), and the effect is is amplified when \(r\) increases, as long as \(\mu_{oi} - p\) is strictly within the support of \(g(\cdot)\).

Second, consider the remaining case of \(\mu_{oi} - p \geq 0\). When \(e_i > \mu_{oi} - p\), the first term of the integrand in (2.A10), \(e_i - (\mu_{oi} - p)\), is positive. Because \(g(\cdot)\) is continuous, as long as \(\mu_{oi} - p\) is strictly within the support of \(g(\cdot)\), there exists \(e_i \in (\mu_{oi} - p, \infty)\)
such that \( g(e_i) > 0 \) and that the integral in (2.A10) is positive, which implies that \( t^\ast(\mu_0; p, r) \) increases with \( r \).

Meanwhile, we obtain

\[
\frac{\partial t^\ast(\mu_0; p, r)}{\partial (\mu_0 - p)} = \frac{1}{c} \int_{\mu_0 - p}^{\infty} (-g(e_i))de_i
\]

and

\[
\frac{\partial^2 t^\ast(\mu_0; p, r)}{\partial (\mu_0 - p)\partial r} = \frac{1}{c} \int_{\mu_0 - p}^{\infty} (-g(e_i))de_i.
\]

Both terms are negative as long as \( \mu_0 - p \) is strictly within the support of \( g(\cdot) \). This means \( t^\ast(\mu_0; p, r) \) decreases with \( |\mu_0 - p| \), and the effect is amplified when \( r \) increases, as long as \( \mu_0 - p \) is strictly within the support of \( g(\cdot) \).

### 2.A.4 Proof of Proposition 2

Based on equation (2.3) and (2.A10)

\[
D(p, r) = \int_p^{\infty} \int_{p-v_i}^{\infty} g(e_i)de_i f(v_i)dv_i + \int_p^{\infty} \int_{-\infty}^{p-v_i} t^\ast(v_i + e_i; p, r)de_i f(v_i)dv_i + \\
\int_{-\infty}^{p} \int_{p-v_i}^{\infty} (1 - t^\ast(v_i + e_i; p, r))de_i f(v_i)dv_i
\]

\[
= \int_p^{\infty} \int_{p-v_i}^{\infty} g(e_i)de_i f(v_i)dv_i +
\int_p^{\infty} \int_{-\infty}^{p-v_i} \frac{r}{c} \int_{-\infty}^{v_i+e_i-p} (v_i + e_i - \tilde{e}_i - p)g(\tilde{e}_i)\tilde{d}_i f(v_i)dv_i +
\int_{-\infty}^{p} \int_{p-v_i}^{\infty} \left(1 - \frac{r}{c} \int_{v_i+e_i-p}^{\infty} (\tilde{e}_i - (v_i + e_i - p))g(\tilde{e}_i)\tilde{d}_i f(v_i)\right)dv_i\text{d}q.A11
\]

Notice that in the second and the third integrals, the first inner layer is to calculate \( t^\ast \) and the integral element is \( \tilde{e}_i \), whereas the second inner layer’s integral element is \( e_i \) and it determines the value of \( \mu_0 \).

We first calculate \( \frac{\partial D(p, r)}{\partial r} \) to investigate how \( D(p, r) \) changes with \( r \),
Based on equation (2.A11),

\[
\frac{\partial D(p,r)}{\partial r} = \frac{1}{c} \int_{-\infty}^{\infty} \int_{p-v_i}^{v_i+\epsilon_i-p} \int_{-\infty}^{\infty} (v_i + \epsilon_i - \tilde{\epsilon}_i - p)g(\tilde{\epsilon}_i)d\tilde{\epsilon}_ig(e_i)f(v_i)de_idv_i -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{p-v_i}^{v_i+\epsilon_i-p} \int_{-\infty}^{\infty} (\tilde{\epsilon}_i - (v_i + \epsilon_i - p))g(\tilde{\epsilon}_i)d\tilde{\epsilon}_ig(e_i)f(v_i)de_idf(2.A12)
\]

\[
= \frac{1}{c} \int_{p-v_i}^{v_i+\epsilon_i-p} \int_{-\infty}^{\infty} \int_{0}^{\infty} xg(v_i + \epsilon_i - p - x) dxg(e_i)de_if(v_i)dv_i -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{p-v_i}^{v_i+\epsilon_i-p} \int_{0}^{\infty} xg(v_i + \epsilon_i - p + x) dxg(e_i)de_if(v_i)dv_i
\]  

(2.A13)

\[
= \frac{1}{c} \int_{p}^{0} \int_{-\infty}^{\infty} \int_{0}^{\infty} xg(y - x) dxg(y + p - v_i)dyf(v_i)dv_i -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y + p - v_i)dyf(v_i)dv_i
\]  

(2.A14)

\[
= \frac{1}{c} \int_{p}^{0} \int_{-\infty}^{\infty} \int_{0}^{\infty} xg(-y - x) dxg(-y - p - v_i)dyf(v_i)dv_i -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y - p - v_i)dyf(v_i)dv_i
\]  

(2.A15)

\[
= \frac{1}{c} \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(-y - x) dxg(-y - p)dyf(z + p)dz -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y - z)dyf(z + p)dz
\]  

(2.A16)

\[
= \frac{1}{c} \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(-y - x) dxg(-y - p)dyf(z + p)dz -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y + z)dyf(-z + p)dz
\]  

(2.A17)

\[
= \frac{1}{c} \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y + z)dyf(z + p)dz -
\]

\[
\frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xg(y + x) dxg(y + z)dyf(-z + p)dz
\]  

(2.A18)

From (2.A12) to (2.A13), we substitute the integral element to \( x = v_i + \epsilon_i - p - \tilde{\epsilon}_i \) in the first part and substitute the integral element to \( x = \tilde{\epsilon}_i - (v_i + \epsilon_i - p) \) in the second part. From (2.A13) to (2.A14), we substitute the integral element to \( y = \epsilon_i - (p - v_i) \). Then we replace \( y \) with \(-y\) in the first part and get (2.A15). From (2.A15) to (2.A16) we substitute the integral element to \( z = v_i - p \), and then get (2.A17) by replacing \( z \) with \(-z\). Based on our assumption that \( g(e) = g(-e), \forall e \), we further get 2.A18.
Define \( H(z) = \frac{1}{c} \int_0^\infty \int_0^\infty xg(y+x)dxg(y+z)dy \). Then

\[
\frac{\partial D(p, r)}{\partial r} = \int_0^\infty H(z)[f(p+z) - f(p-z)]dz.
\] (2.A19)

We first consider the case of \( p > \mu_v \). Since \( p+z-\mu_v > p-z-\mu_v \) and \( p+z-\mu_v > \mu_v - p+z \) for any \( z > 0 \), and hence \( |p+z-\mu_v| > |p-z-\mu_v| \). Recall that we assume \( f(\cdot) \) has a unique mode \( \mu_v \), and \( f(\cdot) \) is weakly increasing on \(( -\infty, \mu_v )\) and weakly decreasing on \([\mu_v, \infty)\). Given that \( p+z \) is further away from the mode \( \mu_v \), we have \( f(p+z) - f(p-z) \leq 0 \). Noticing that \( H(z) \geq 0 \), we get that \( \frac{\partial D(p, r)}{\partial r} = \int_0^\infty H(z)[f(p+z) - f(p-z)]dz \leq 0 \). Since \( f(\cdot) \) cannot be constant throughout the real line\(^5\), then for any \( p \in (\mu_v, \infty) \), there must exist \( z > 0 \) such that \( f(p+z) - f(p-z) < 0 \). Denote \( Z^-(p) = \{z > 0 : f(p+z) - f(p-z) < 0\} \), \( S_g = \{z > 0 : g(z) > 0\} \). If the (Lebesgue) measure of \( Z^-(p) \cap S_g \) is greater than 0, then the integral \( \int_0^\infty H(z)[f(p+z) - f(p-z)]dz \) is strictly negative.

Similarly, we can prove that when \( p < \mu_v \), \( f(p+z) - f(p-z) \geq 0 \), so \( \frac{\partial D(p, r)}{\partial r} = \int_0^\infty H(z)[f(p+z) - f(p-z)]dz \geq 0 \), and it is positive when \( p \) satisfies the condition that \( Z^+(p) \cap S_g \) has a positive (Lebesgue) measure, where \( Z^+(p) = \{z > 0 : f(p+z) - f(p-z) < 0\} \).

Now we calculate \( \frac{\partial}{\partial r} \left( \frac{\partial D(p, r)}{\partial p} \bigg|_{p=\mu_v} \right) \) to see how the central slope of demand curve changes with \( r \). It is easy to see that \( \frac{\partial}{\partial r} \left( \frac{\partial D(p, r)}{\partial p} \bigg|_{p=\mu_v} \right) = \frac{\partial^2 D(p, r)}{\partial p \partial r} \bigg|_{p=\mu_v} \). According to

\(^5\)If \( f(\cdot) \) is constant throughout the real line, \( \int_{-\infty}^\infty f(v)dv = 0 \) or \( \pm \infty \), which conflicts with \( \int_{-\infty}^\infty f(v)dv = 1 \)
and the theorem of integration by parts, we have

\[
\frac{\partial^2 D(p, r)}{\partial p \partial r} = \frac{\partial}{\partial p} \int_0^\infty H(z)[f(p+z) - f(p-z)]dz
\]

\[
= \int_0^\infty H(z)dz[f(p+z) + f(p-z)]
\]

\[
= H(z)[f(p+z) + f(p-z)]|_{z=0} - \int_0^\infty [f(p+z) + f(p-z)]dH(z)
\]

\[
= -2H(0)f(p) - \int_0^\infty [f(p+z) + f(p-z)]dH(z)
\]

\[
= 2f(p) \int_0^\infty dH(z) - \int_0^\infty [f(p+z) + f(p-z)]dH(z)
\]

\[
= \int_0^\infty [2f(p) - f(p+z) - f(p-z)]dH(z)
\]

(2.A20)

(2.A21)

(2.A22)

The reasoning from (2.A20) to (2.A21) is as follows. By definition of p.d.f, \( \int_{-\infty}^\infty g(z)dz = 1 \), so we must have \( \lim_{z \to \pm \infty} g(z) = 0 \). Then \( \lim_{z \to \infty} g(y+z) = 0 \) for any \( y > 0 \), and therefore \( \lim_{z \to \infty} H(z) = 0 \). Similarly, \( \lim_{z \to \pm \infty} f(z) = 0 \), and thus \( \lim_{z \to \infty} f(p \pm z) = 0 \).

By (2.A22), \( \frac{\partial}{\partial r} \left( \frac{\partial D(p, r)}{\partial p} |_{p=\mu} \right) = \int_0^\infty [2f(\mu) - f(\mu + z) - f(\mu - z)]dH(z) \). Recall that \( g(\cdot) \) is assumed to be symmetric around 0 and is weakly decreasing and non-constant on \((0, \infty)\). Then according to the definition of \( H(z) \), \( H(z) \) is weakly decreasing and non-constant on \((0, \infty)\). Given our assumption that \( f(\cdot) \) is weakly increasing on \((-\infty, \mu)\) and weakly decreasing on \((\mu, \infty)\), \( 2f(\mu) - f(\mu + z) - f(\mu - z) \geq 0 \). Then \( \frac{\partial}{\partial r} \left( \frac{\partial D(p, r)}{\partial p} |_{p=\mu} \right) \leq 0 \), and it is guaranteed to be negative when \( Z(\mu) = \{z > 0 : f(\mu + z) + f(\mu - z) < 2f(\mu) \} \) has a non-empty intersection with the set of \( z \) that \( H(z) \) is strictly decreasing, which is the same as the set of \( z \) that \( g(z) \) is strictly decreasing.

2.A.5 Details of Structural Estimation and Extrapolation

We first draw three \( N \times K \) matrices of random numbers that are independent and identically distributed, following the standard normal distribution, where \( N \) is the number of individuals, and \( K = 100 \) is the number of iterations we will perform to simulate
the average purchase probability of each individual. The three matrices are denoted as $e_1, e_2, e_3$. The draws are actually quasi-random: we generate a two-dimensional Halton set with three columns, each column of which are evenly distributed numbers on $[0, 1]$, take the first $N \times K$ elements of each column, and then converting the numbers to standard normal distribution by taking the inverse of normal c.d.f. of them. Since Halton set is more evenly distributed on $[0, 1]$ compared to direct random draws of the uniform distribution on $[0, 1]$, the random draws created in this way leads to better convergence performance compared to direct random draws and requires less number of draws (Train 2009).

We also generate two $1 \times T$ vectors, which are the Gauss-Hermite quadrature nodes and weights over $[-1, 1]$, where $T = 1000$. They will be used to calculate the expectation of a function of a normally distributed random variable.

Given these draws, we can calculate the average purchase probability of each individual under a given set of parameters, and then calculate the log likelihood of the observed data. The objective function is the sum of log likelihood of the observed data.

Given a set of parameter values $(b_0, b_1, b_2, a_0, a_1, c_1, c_2, \sigma_v)$, we perform $K$ iterations of calculation. Within each iteration, the steps are as follows.

1. Simulate each individual’s true valuation $v_i = b_0 + b_1 \text{Log-Diamond}_i + b_2 \text{VIP}_i + \sigma_v e_{1ik}$, for $i = 1, ..., N$, where $e_1$ is an $N \times K$ matrix of i.i.d. standard normal draws which we have created at first, and $e_{1ik}$ is the element $(i, k)$ of it.

2. Calculate each individual’s prior uncertainty $\sigma_{0i} = \exp (a_0 + a_1 \text{VIP}_i)$, for $i = 1, ..., N$.

3. Simulate each individual’s prior belief $\mu_{0i} = v_i + \sigma_{0i} e_{2ik}$, for $i = 1, ..., N$, where $e_2$ is also an $N \times K$ matrix of i.i.d. standard normal draws. Thus we have $\mu_{0i} \sim N(v_i, \sigma_{0i}^2)$.

4. Calculate $E[(v_i - p_i)^+]$, $i = 1, ..., N$, where $p_i$ is the price assigned to $i$. The expectation is taken over each individual $i$’s belief about the distribution of $v_i$, .
which is $N(\mu_0i, \sigma_{0i}^2)$. To get better convergence performance, we use Gauss-Hermite quadrature method. That is, if a random variable $Y \sim N(\mu, \sigma^2)$, 
$$\mathbb{E}[f(Y)] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{T} w_j f(\mu + \sqrt{2}\sigma x_j),$$
where $x_j, w_j$ are the Gauss-Hermite quadrature nodes and weights over $[-1, 1]$. In our case, $\mathbb{E}[(v_i - p_i)^+] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{T} w_j \cdot (\mu_0i + \sqrt{2}\sigma_0i x_j - p_i)^+.$

5. Simulate each individual's effort cost $c_i = \exp(c_1 + c_2e_{3ik})$, where $e_3$ is the standard normal matrix that we have drawn.

6. Calculate each individual's effort level

$$t_i = \min \left\{ \frac{r_i}{c_i} \left( \mathbb{E}[(v_i - p_i)^+] - (\mu_0i - p_i)^+ \right), 1 \right\},$$

where $r_i$ is the realization probability assigned to $i$. $c_i, \mathbb{E}[(v_i - p_i)^+], \mu_0i$ have been simulated in previous steps.

7. Calculate each individual's purchase probability $Pr_k(Buy_i = 1) = t_i \frac{\exp(v_i - p_i)}{1 + \exp(v_i - p_i)} + (1 - t_i) \frac{\exp(\mu_0i - p_i)}{1 + \exp(\mu_0i - p_i)}$.

After the $K$ iterations, for each individual, we average over the iterations to get the individual's purchase probability $Pr(Buy_i = 1) = \frac{1}{K} \sum_{k=1}^{K} Pr_k(Buy_i = 1)$, and then calculate the sum of log likelihood $LL = \sum_{i=1}^{N} \left[ 1(Buy_i = 1) \log Pr(Buy_i = 1) + 1(Buy_i = 0) \log (1 - Pr(Buy_i = 1)) \right]$.

Being able to calculate the simulated log likelihood given a set of parameter values, we search over the parameter space and find the set of parameter values that maximizes the simulated log likelihood. We restrict the value of $\sigma_0$ and $c_2$ to be positive, since they represent the standard deviations of a normal distribution and a log-normal distribution. Only the data of $r = 1/30, 1/2$ conditions are used to perform the estimation.

Given the parameter estimates, we follow the same steps as described above to calculate the purchase probability of each individual in the $r = 1$ condition: first draw random numbers, and then go over $K$ iterations to simulate each individual's
purchase probability under the estimated parameter values. Lastly, we aggregate the purchase probability and plot the demand curve of the structural forecast.
Chapter 3

Rational Spamming

3.1 Introduction

With the emergence of new communication technologies and platforms, disseminating information to the public has never been more convenient. Firms in particular are able to advertise to consumers in new, more economical ways. For example, firms can create their official accounts on a social media platform, and advertise by posting product-related messages at negligible marginal costs. However, a consequence of such convenience of communication is that consumers are exposed to far more information than they can process. In 2015, every internet user receives 122 emails on average per day,¹ and over 500 million tweets are sent on Twitter per day, among which 71% are ignored and only 23% generate a reply.² In the age of information explosion, a major challenge firms face in advertising is the limited attention of consumers. In this chapter, we use the advertising of TV shows on a tweeting website as a concrete setting to investigate how consumers respond to excessive advertising and what competing firms’ optimal advertising strategy should be given the limited attention of consumers.

We focus on Weibo, the most popular tweeting website in China. It has similar features as Twitter, and has attracted numerous firms to open business accounts on

²http://www.digitalinformationworld.com/2015/01/twitter-marketing-stats-and-facts-you-should-know.html
the platform. In particular, most media companies in China, such as TV channels, newspapers, and news websites, have opened accounts on Weibo to advertise their products. We track the Weibo accounts of 49 weekly TV shows that were broadcasted across China from April to October 2015. We find that these company accounts post tweets intensively, and there is no big difference in posting intensity across shows with different levels of popularity. We further look at the change in the number of followers. We find that the number of followers goes downward when the posting intensity gets too high. Since more tweets tend to attract more new followers, it is reasonable to infer that existing followers choose to unfollow the accounts that post too many tweets.

Motivated by these observations, we develop an analytical model to understand why consumers with limited attention choose to unfollow a firm and, more importantly, why firms choose to advertise intensively knowing the limited attention of consumers.

In the model, we treat tweets as a form of informative advertising for a product. Following the literature of informative advertising (Erdem and Keane 1996, Erdem et al. 2008, etc.), we assume that advertising exposure gives a consumer noisy signals about product quality, and the consumer uses these signals to update her belief about product quality in a Bayesian manner. As the consumer gets more information about the product quality, her belief becomes more precise. The consumer will have a greater probability of choosing the product with higher true quality, which as a result increases the consumer’s expected utility from consuming the chosen product. However, the consumer’s reading capacity is restricted by her limited attention. If a firm is posting intensively and the consumer already knows about the product quite precisely, the consumer would rationally unfollow the firm. In this way, she can leave more attention to advertisements on product she knows less about and make a more informed choice of which show to watch eventually.

Understanding the mechanism behind consumers’ unfollowing decisions leads to a counterintuitive result—our model indicates that posting intensively can be an optimal strategy for firms. Firms are competing for consumer’s limited attention. If
a firm posts more advertisements, the belief about its show quality may experience a larger change because consumers can receive more information about the product to update their belief (we call this effect the “dispersion effect”); at the same time, information signals from the competitor are crowded out, so the mean belief about the competitor’s product quality is likely to change less (we call this effect the “crowd-out effect”). The firm perceived as having higher quality wants to keep its advantage and hence prefers smaller changes in beliefs about itself and about its competitor. Therefore, this firm would like to exploit the crowd-out effect but avoid the dispersion effect. The firm perceived as having lower quality would like to reverse the current situation, hence it favors the dispersion effect but dislikes the crowd-out effect. Advertising intensity will also affect consumers’ unfollowing decision. Both firms need to make a trade-off among the three effects when deciding their posting intensities. When the better-perceived firm has a more precise prior belief among consumers than its worse-perceived competitor, the benefit from the crowd-out effect can dominate the damage of the dispersion effect when consumers’ attention is very limited, and the better-perceived firm will post as many advertisements as possible. The worse-perceived firm would post intensively in any case because its priority is to improve its perception among consumers through the dispersion effect. Consumers’ unfollowing does not prevent the better-perceived firm from advertising intensively, because the better-perceived firm’s objective is to minimize the information conveyed to consumers, which is completely opposite to consumers’ interest. Thus reducing advertising intensity to the level that can retain followers will increase the amount of information conveyed by advertisements and make the better-perceived firm worse off. The worse-perceived firm’s objective is in line with consumers’ interest and thus consumers will not unfollow the worse-perceived firm in equilibrium.

There is a growing literature on limited attention. Previous works in psychology provide evidence that people have limited attention in receiving and processing information (see Eppler and Mengis 2004, DellaVigna 2009 for a review). Van Zandt (2004) points out that as the costs of generating and transmitting information fall, the main bottleneck in communication becomes the limited attention of human receivers.
She shows that mechanisms that increase the cost of sending messages will shift the task of screening messages from the consumers to firms, who know the contents better, a shift that will benefit both parties. Anderson and De Palma (2009) further consider message examination as an endogenous decision of consumers, and confirm that a higher cost of sending messages will increase the number of messages examined by consumers because the expected average surplus rises. Built on these findings, Anderson and De Palma (2012) model multiple sectors competing for consumers' attention, with competition in price within each section. Iyer and Katona (2015) analyze the incentives to be a sender or a receiver for non-commercial users in social networks given that people derive utility from being listened to but receivers have limited attention. Zhu and Dukes (2015) study competing firms' strategy in choosing the product attribute to make prominent when consumers have limited attention, and find that firms emphasizing the same attribute may actually increase product differentiation. Our work shares a similar context with Anderson and De Palma (2009, 2012) since we all investigate firms' advertising strategies in response to the limited attention of consumers. The difference is that we consider a real-life situation where firms do not pay for consumers' attention, but face the risk of losing or damaging its quality image among attention-constrained consumers. Furthermore, we derive the surprising result that posting intensively can be an optimal strategy for firms in spite of consumers' limited attention.

Our work is closely related to the literature on informative advertising and competitive informative advertising. Anderson and Renault (2006) analyze a monopoly firm's choice of advertising content and find that the firm prefers to convey only limited information if possible. Our model shares some similar assumptions with the paper: consumers having heterogeneous tastes, firms having no private information on consumers' match value, identical priors on consumers' match value, and no cost for informative advertising. The difference is that we consider a duopoly setting, where competing firms can decide the amount of ads but consumers have limited attention. Grossman and Shapiro (1984) investigate the role of informative advertising in a competitive market with horizontal product differentiation and consumer het-
erogeneity. They find that although informative advertising improves the matching between products and consumers and increases demand elasticity faced by each firm, the market-determined levels of advertising are higher than socially optimal. Informative advertising is not free in their context, and in fact a decrease in advertising cost may reduce profits by increasing the severity of price competition. Our context is different in that we are considering an informational good without an explicit price, but consumers can only choose one to consume due to the restriction of time.

Finally, this chapter also contributes to the literature on social media advertising. Stephen and Galak (2012) examine how traditional earned media (e.g., press mentions) and social earned media (e.g., blogs and online community posts) affect sales. They find that, accounting for event frequency, social earned media has a larger effect on sales in aggregate. Toubia and Stephen (2013) point out that as Twitter gets mature, non-commercial users will probably reduce their posting and Twitter may become a media channel where firms broadcast content to consumers. Existing works have shown that social media websites like Twitter and Weibo can be an effective marketing platform. Gong et al. (2014) conduct a field experiment on Weibo, showing that posting tweets about TV shows and recruiting influentials to repost these tweets increase show viewing significantly. Liu et al. (2014) find that the contents of tweets related to TV shows can make more accurate predictions on show viewing compared to other online data, which at the same time provides evidence that Twitter might be an effective marketing platform for the spread of information. Built on these studies, our work captures two features of social media advertising (especially tweeting)—low advertising costs and limited consumer attention. We also take a step further to explore competing firms’ advertising strategies in an equilibrium setting.

The rest of the chapter is organized as follows. In §3.2, we describe the motivating evidence gathered on Weibo. In §3.3, we present the model, analysis, and results. In §3.4, we discuss the limitations of the model and directions for future research.
3.2 Background and Motivating Phenomena

3.1 Background and Data

Weibo.com is a leading microblogging website in China. Launched by Sina Corporation in August 2009, it has been one of the most popular social media platforms in China. As of June 2015, the number of monthly active users has been 212 million.³

The functions and features of Weibo are very similar to Twitter. A registered user can post a “tweet”, which is a message with no more than 140 Chinese characters and may include images, short videos, URLs, or a pair of hashtags that indicate the topic of the tweet. A user can “follow” other users to subscribe their tweets so that their future tweets will appear on her homepage automatically, and she can “unfollow” these users later on to cancel the subscription. Notice that following is a unidirectional relationship: the followee’s tweets will appear on the follower’s homepage, but not vice versa. A user can “like” a tweet by clicking the “like” button. She can also give a comment under a tweet, or repost a tweet so that it will appear on her timeline as a “retweet”. For each user, the tweets posted by those she follows will appear on her homepage according to a reverse chronological order, that is, the most recent tweets will be displayed at the top of the page, and she needs to scroll down the screen to see the earlier posts.

The unidirectional relationship and the huge user base make Weibo an ideal platform for wide and fast transmission of information, and hence more and more celebrities, organizations, and companies have opened accounts on Weibo, using it as a marketing tool for themselves or their products. Weibo verifies the authenticity of these accounts by adding a mark “V” under the accounts’ profile photos, certifying that they are verified accounts.

We collected data from Weibo to observe how weekly TV shows advertise on this popular social media platform and how consumers respond. There are more than 30 satellite TV channels in the mainland of China, including national and provincial ones, all of which can broadcast their shows to the audience throughout the country.

Almost all of these channels have at least one weekly TV show, aired at prime time. Most of them are reality shows or entertainment shows. Some of them have been very popular and successful, attracting millions of audience. With highly overlapping broadcasting times and similar topics, these TV shows are competing severely. As a major social media platform, Weibo is playing an important role in these TV channels' social media marketing campaign. Usually, a TV channel opens one account for each of its weekly TV show, posting tweets containing information related to the show.

We track 49 accounts, each representing a weekly TV show on a satellite TV channel\(^4\). We collect data from Weibo API\(^5\) from April 2015 to October 2015 on an hourly base. The data we scraped include the number of followers of the account at the scraping time, and all the tweets the account posted during the hour. For each tweet, we know the following characteristics: when it was created, how many times it has been retweeted, commented, or liked, the content of tweets, whether it contains a picture, whether it has a hashtaged topic, whether it is a retweet, the content and poster of the original tweet if it is a retweet.

In order to get supplemental data that reflects the popularity of each show, we scrape data from Baidu Search Index using the name of each show as a keyword\(^6\). Baidu is the leading search engine in China. Baidu Search Index gives the public interest trend on Baidu for each keyword, and the index is comparable across different keywords. The index we scraped is on a daily base and cover the same time range as the Weibo data.

\(^4\)There is a big difference in the levels of management across channels. Not all the channels are actively engaged in managing their Weibo accounts. We select the top 10 channels in China. The rank of channels comes from their rating in the last five years, see http://www.tvtv.hk/archives/2445.html. These 49 accounts are all the active accounts we can find for in-season TV shows broadcasted by the 10 channels in the observation window. All the shows we track are reality shows or entertainment shows. For each show, different episodes are in the same form but the contents can differ, so the information value of tweets does not decrease fast as time goes on.

\(^5\)http://open.weibo.com/wiki/%E5%BE%AE%E5%8D%9AAP

\(^6\)http://index.baidu.com/
3.2 Summary Statistics and Motivating Phenomena

Table 3.1 gives the summary statistics of each account in the six months, including the average number of followers throughout the period, the total number of tweets each account posts in the period, the maximum hourly number of posts, and the average daily Baidu search index. We can see that these accounts are overall quite popular, with a mean number of followers more than 0.8 million, and the most popular account has more than eight million followers. Their posting activeness varies, but there are many accounts post intensively in certain periods: the maximum hourly number of posts is about 15 tweets on average (i.e., about one tweet in every 4 minutes), and the most aggressive account posted 41 tweets in an hour.

There are three main features emerging from the data.

First, these accounts post intensively in some periods. As a result, there are too many tweets for users to read.

These official accounts have different posting behavior at different times of a day. We classify a day into three sections: 12AM-8AM as “night” during which most people are sleeping, 8AM-7PM as “daytime” during which most people are working, and 7PM-12AM as “Prime Time” during which most people are relaxing and may have more time to watch TV shows or browse Weibo. In each section, the proportions of observations with a positive number of postings are 0.68%, 24.58%, 20.60%, respectively. (One observation is a show-hour combination.) For those observations with a positive number of postings, we summarize the mean, standard deviation, 75%-quantile, 95%-quantile, and maximum of the number of postings in each section (see Table 3.2). The distribution of hourly number of postings during prime time is highly right-skewed: the average is about 2.5, whereas the top 5% are greater than 10. Posting more than 10 tweets in an hour means every six minutes a tweet is posted.

We want to demonstrate that the intensive posting leads to too many tweets for...

---

7Official definition of “Prime Time” in China is 7PM-10PM. However, we notice that in many satellite TV channels, they are broadcasting daily TV series before 10PM, and broadcasting popular weekly reality shows or entertainment shows after 10PM since the targeting audience are younger people who usually go to bed later. Since many of our observed shows belong to this category and the majority of Weibo users are younger people, we extend the “Prime Time” to 7PM-12AM.
users to read. We recruit 286 Weibo users and run an online survey about their using behavior. (A translation of survey questions and a summary of results are available in Appendix.) 12.95% of them browse Weibo less than once a day, 51.55% of them browse Weibo one to three times a day, 18.65% of them browse Weibo four to six times a day, and 16.84% of them browse more than six times a day. We can see that most of these users browse Weibo on a regular basis. We ask them to estimate how many verified accounts they are following. 73.57% of the subjects estimate that they follow more than 10 verified accounts, and among them nearly 13 follow more than 30 verified accounts. If we assume that all verified accounts have similar posting intensity as the TV show accounts we track, a user who follows 10 verified accounts has approximately $1.56 \times 10 \times 24 \approx 374$ tweets from verified accounts per day. If she checks Weibo three times per day, then she has at round 124 tweets to read every time, not to mention tweets from non-verified accounts.

To further demonstrate this point, we ask these Weibo users whether they can finish reading all the new tweets each time. Only 6.22% of them say that they can always read all the new tweets each time they browse Weibo. 8.29% say that they can never finish reading all the new tweets, 25.65% say that they usually cannot finish reading all the new tweets, 32.12% can sometimes finish reading all the new tweets, and 27.72% can usually finish reading all the new tweets. 46.85% of the subjects agree or strongly agree that some verified accounts post too intensively, preventing them from reading tweets from other accounts.

Second, posting intensities are not significantly positively correlated with popularity of the shows and accounts.

We measure the posting intensity of each account using the average hourly number of tweets posted in active periods. By active periods, we mean the hours that the account posts at least one tweet. We only count active periods because overall, accounts are actively posting tweets in only about 16% of time. The average hourly number of tweets posted in active periods is approximately $1.56 \times 10 \times 24 \approx 374$.

---

8 We ask about “verified accounts” instead of merely TV show accounts so that more subjects have related experience. As we have mentioned, TV show accounts are a type of verified accounts. Verified accounts have similar purpose of posting: they are mostly posting tweets to advertise the works, products, or services of the celebrities, organizations or companies that they represent. Their posting frequencies are also comparable.
posts over all the periods will be more of a measure of how active the account is rather than about how intensively the account posts when it is active. We use the average daily (Baidu) search index to measure the popularity of shows, and use the average number of followers throughout the observation window to approximate the popularity of accounts. We calculate the correlation between the posting intensity of each account with the two measures of popularity. They are not significantly positively correlated: \( \text{Corr(Hourly Posting Intensity, Avg. Search Index)} = 0.054 \) (\( p=0.712 \)), and \( \text{Corr(Hourly Posting Intensity, Avg. Number of Followers)} = -0.131 \) (\( p=0.371 \)). Figure 3-1 shows the scatter plots of accounts’ average posting intensities over (a) Log(Avg. Daily Search Index) and (b) Log(Avg. Number of Followers). We can see that the insignificance of correlation should not be due to the small number of observations.

Third, advertising too intensively may drive followers away.

To observe the effect of posting intensity on the number of followers, for each account, we regress its number of followers at the end of each time period on the lag number of followers, the number of tweets posted in the period, and time-varying show attributes as well as tweet attributes:

\[
y_{jt} = \alpha y_{jt-1} + \beta_1 n_{jt} + \beta_2 f(n_{jt}) + \gamma_1 S_{jt} + \gamma_2 X_{jt} + \eta_{jt},
\]

(3.1)

where \( y_{jt} \) denotes the number of followers of account \( j \) at the end of time \( t \), \( n_{jt} \) denotes the posting intensity of account \( j \) during time \( t \) and \( f(n_{jt}) \) is a function of \( n_{jt} \). \( S_{jt} \) represents the time-varying show attributes and \( X_{jt} \) is for time-varying tweet attributes, \( j = 1, ..., 49, t = 2, ..., T \). Time-varying show attributes include the following variables: \textit{Daily Search Index}, daily Baidu search index of the show name; \textit{In Season Dummy}, whether show \( j \) is in its broadcasting season at time \( t \); \textit{Show Day Dummy}, whether show \( j \) is broadcasted on the day which time \( t \) belongs to; \textit{Show Time Dummy}, whether show \( j \) is broadcasted during time \( t \). Time-varying

\footnote{We also calculate the average hourly number of posts over all the periods. It is positively correlated with each account’s average search index (Corr(Hourly Posting Intensity, Avg. Search Index) = 0.261, \( p=0.070 \)), but not significant for the average number of followers (Corr(Hourly Posting Intensity, Avg. Number of Followers) = 0.178, \( p=0.220 \)). It means that popular shows’ accounts are overall more active than unpopular ones’, but they are not posting more intensively than the unpopular ones in active periods.}
tweet attributes are: *Avg. # Comments*, the average number of comments under the tweets posted by account *j* during time *t*; *Avg. # Likes*, the average number of likes under the tweets posted by account *j* during time *t*; *% of Tweets Containing Pictures*, the percentage of tweets posted by *j* during time *t* that contain pictures; *% of Tweets Containing Hashtag Topics*, the percentage of tweets posted by *j* during time *t* that contain hashtag topics; *% of Tweets that are Reposts*, the percentage of tweets posted by *j* during time *t* that are reposts of previous tweets; *Hour # Being Retweeted*, the number of times that the tweets posted by account *j* during time *t* has been retweeted.

Since we want to observe the effect of too many tweets, we let \( f(n_{jt}) = (n_{jt} - n^*)^+ \), where \( n^* \) is a threshold used to define “too-many”.

We first treat each account-hour combination as an observation and run an OLS regression according to equation (3.1) (Table 3.3, Column (1)), defining \( n^* = 6 \), which is the 90-percentile of hourly number of posts during prime time.\(^{11}\) Time-varying tweet attributes are the aggregate characteristics of tweets posted in the past hour by the account. We see that \( (n_{jt} - 6)^+ \)'s coefficient, \( \beta_2 = -32.16 \), is significantly negative and has a larger magnitude than the coefficient of \( n_{jt} \), \( \beta_1 = 30.00 \). We can interpret it as when \( n_{jt} \) is above 6, a further increase in \( n_{jt} \) might have a negative impact on account *j*’s number of followers.

However, across different shows (accounts), there can be substantial heterogeneity in the dynamics of followers, which are hard to be fully captured by the data we have. The show-specific attribute is very likely to be correlated with \( n_{jt} \) and \( X_{jt} \). For example, some shows have better management of the account in the sense that they post relatively more tweets with high quality. Besides the effect of the number

\(^{10}\) The number of comments under each tweet increases over time. We scrape data from Weibo on an hourly base, and each time we can observe the last 500 tweets posted by the accounts we are tracking. Hence each tweet appears in our dataset for multiple times, and for each tweet we regard the last record as its final record. We find that in the last few records for each tweet, the numbers of comments and likes become relatively stable, so it is reasonable to use the numbers in the last record to measure the performance of each tweet.

\(^{11}\) We try different cutting threshold. For \( n^* = 3, 4, 5 \), the coefficient of \( (n_{jt} - n^*)^+ \) is also significantly negative in the OLS and fixed effects model, but its magnitude is not larger than the coefficient of \( n_{jt} \). 6 is smallest number that leads to a larger magnitude of the coefficient of \( (n_{jt} - n^*)^+ \).
of tweets, this reputation helps them attract new followers. Such kind of unobserved heterogeneity is embedded in the error term $\eta_{jt}$ in the regression (3.1). Hence the terms $n_{jt}$, $f(n_{jt})$, $X_{jt}$ are likely to be correlated with the error term. We decompose the error term $\eta_{jt}$ as $\eta_{jt} = u_j + v_{jt}$, where $u_j$ denotes show-specific heterogeneity, and $v_{jt}$ is the idiosyncratic shock to show $j$ at time $t$. Then equation (3.1) can be rewritten as:

$$y_{jt} = \alpha y_{j,t-1} + \beta_1 n_{jt} + \beta_2 f(n_{jt}) + \gamma_1 s_{jt} + \gamma_2 X_{jt} + u_j + \epsilon_{jt}. \quad (3.2)$$

Equation (3.2) can be estimated using fixed effects model, as fixed effects model uses only within-unit variation in the regressors and allows for correlation between regressors and $u_j$. Strictly speaking, equation (3.2) is a dynamic panel data model since the regressors include the lag of the dependent variable. The estimated $\hat{\alpha}$ is biased because when we estimate (3.2), we need to take first-order-difference and $y_{j,t-1} - y_{j,t-2}$ is correlated with $\epsilon_{jt} - \epsilon_{j,t-1}$ (Nickell 1981). However, this bias is negligible when $T$ is very large compared to $N$. In our dataset, it is reasonable to regard $T$ as very large compared to $N$ as $N = 49$ and $T \approx 2750^{12}$. Therefore, we simply use fixed effects model to estimate (3.2), as reported in Table 3.3, Column (2). In Column (3), we further add time effects in the regression, including hour-of-day effects, day-of-week effects, and in-season, out-of-season week effects which capture how long the show has been in season or out of season. We see that the coefficient of $(n_{jt} - 6)^+$ is significantly negative and has a consistently larger magnitude than the coefficient of $n_{jt}$, lending further evidence that the effect of tweets on the number of followers is downward sloping when the number of tweets goes beyond 6.

When people are browsing Weibo, they may scroll down and read earlier tweets that were posted more than one hour ago. Thus the tweets from earlier periods may also affect the number of followers in the current period. We include the lags of number of tweets up to four hours ago, i.e., $n_{j,t-1}, ..., n_{j,t-4}$, and the corresponding function $f(\cdot)$ of these four lags in the regression. The result is reported in Column (4).

---

12 The exact number of observations can be different for different accounts due to different account opening time and some missing data.
of Table 3.3. We see that the coefficients of \((n_j, t - \tau - 6)^+\) have a larger magnitude than the coefficients of the corresponding \(n_{j,t-\tau}\) for \(\tau = 3, 4\). The results imply that advertising too intensively may have a delayed negative effect on the number of followers.

Furthermore, we try another value for the threshold of intensive advertising by letting \(n^* = 10\), which is the 95-percentile of the hourly number of posts during prime time. We find that the difference between the absolute value of the coefficient of \((n_{jt} - n^*)^+\) and the coefficient of \(n_{jt}\) is becomes even larger (Column (5) of Table 3.3). When adding four lags of \(n_{jt}\) and \((n_{jt} - n^*)^+\), the coefficients of \((n_j, t - \tau - 6)^+\) are all of greater absolute value than the coefficients of \(n_{j,t-\tau}\).

Previous results imply that the effect of tweets can be delayed to some extent. As a robustness check, we observe the effect of tweets in four-hour periods. We classify each day into six four-hour periods: 12AM-4AM, 4AM-8AM, 8AM-12PM, 12PM-4PM, 4PM-8PM, 8PM-12AM. Users’ behavior and these accounts’ activities should be relatively similar and consistent within each four-hour period. We first try \(n^* = 12\), which is the 90-percentile of number of tweets posted in four-hour periods. Similar as the hourly case, we run fixed-effects regression, and then add time effects, lags of number of tweets, and lastly let \(n^* = 24\) which is the 95-percentile of number of tweets posted in four-hour periods. The results mostly replicate what we get in the hourly case.

Our survey result also reinforces this point. 74.08% of the subjects say that they have unfollowed verified accounts. Among them, 58.3% agree or strongly agree that the reasons of unfollowing include that the accounts are posting intensively and prevent them from reading tweets from other accounts.

To summarize, the above results lend evidence that post tweets too intensively can have a negative impact on the number of followers.

Besides this finding, the coefficients of time-varying tweet attributes are also informative. In the hourly case, we see that the number of “likes” has a positive impact on the number of followers, whereas containing pictures in tweets has a negative impact on the number of followers, implying that people may find tweets with pictures more
annoying as they take more space but are less informative. In the four-hour case, the coefficient of the number of times being retweeted becomes significantly positive. This is reasonable since when a tweet has been retweeted more, it means that more people can see the tweet from the account, and thus the account can attract more new followers. The difference in the results of hourly and four-hour cases might be due to the delay in the effect.

Motivated by the phenomena above, we would like to ask: when will a rational consumer choose to unfollow an account? Given that consumers have limited attention, what should be competitive firms' optimal posting strategy? We next build a model in which tweets serve as informative advertising and consumers have limited attention. We give an explanation of why consumers may unfollow an intensively posting account rationally. We also find that posting intensively can be the optimal strategy of firms under the competitive setting, even when they know the possible unfollowing action of consumers.

3.3 Model

There are two firms in the market. Each of them advertises one product on a social media site by posting ads via their official accounts. Corresponding to the empirical setting above, we are basically considering two TV channels, where each of them advertises a TV show on Weibo to its followers by posting tweets. Compared to most traditional advertising such as TV commercials, tweets have two features: (1) they are free; (2) they are usually non-repetitive and provide lots of information about the product. For example, tweets posted by TV channels can contain trailers, stage photos of the upcoming episode, etc. In our survey, 75.65% of subjects are following or used to follow TV show accounts. Among these subjects, 92.12% agree or strongly agree that tweets posted by the TV show accounts contain relevant information about the TV shows, and 90.41% agree or strongly agree that tweets posted by the TV show accounts can help them know the content of the TV shows. Therefore we consider

\[13\] Tweets may also function as consumption goods which complements the product (Becker and Murphy 1993). It is hard to distinguish these two roles empirically or consider them at the same
a model of informative advertising.

The two firms' products are horizontally differentiated. Each consumer will derive a match value $V_j$ from the consumption of product $j$ ($j = 1, 2$). Consumers can have heterogeneous tastes, and therefore $V_j$ could be different for different consumers. However, we assume they share the same prior belief on $V_j$ ex ante. Ads from firm $j$ are noisy but informative signals on $V_j$, with the $k$-th signal $S_{jk} | V_j \sim N(V_j, \omega^2)$, where $S_{jk}$ is independent and identically distributed across $j$ and $k$, and $\omega$ measures the noisiness of the signals. We assume that there is no information asymmetry on $V_j$ between firms and consumers, so all of them share a common prior belief $V_j \sim N(\mu_{0j}, \sigma_{0j}^2)$ for $j = 1, 2$. The idea behind this assumption is that even if a firm knows the content of the product, it does not know consumers' tastes perfectly. This assumption has been made in previous literatures on informative advertising (e.g., Anderson and Renault 2006). Since consumers and firms have identical priors, we normalize consumers to a single representative consumer. Without loss of generality, we assume that $\mu_{01} > \mu_{02}$, i.e., product 1 is believed to have a higher match value than product 2 a priori.

We study a game with the following timeline of events.

First, both firms decide how many ads to post simultaneously, $n_j$ for $j = 1, 2$. We call $n_j$ firm $j$'s posting intensity, and assume that $n_j \in [0, \bar{n}_j]$. $\bar{n}_j$ is assumed to be exogenously given, and in practice can be determined by the amount of available content to post and firms' work capacity, etc.

Second, the consumer observes firms' posting intensities and decides whether to unfollow each firm. Denote the representative consumer's decision on firm $j$ as

---

14Signals $S_{jk}$ are revealed by a common set of ads, but could differ among consumers, because different consumers may have different preference (tastes) over the information revealed by the ads. For example, some consumers like action movies while others like comedy; therefore, an ad revealing the genre of a TV show means differently for different consumers. Consequently, consumers' posterior belief on $V_j$ could be different for different consumers. As shown below, we do not need to consider this ex post difference explicitly, because consumers and firms need to make their decisions before observing the signals.

15The assumption of symmetric information rules out the possibility that firms can potentially signal their product match value via advertising intensity.

16Here we have basically assumed that consumers can observe firms' posting intensities before starting reading ads. This is a reasonable assumption if we consider a typical setting in practice.
If she unfollows firm j’s account (i.e., \( f_j = 0 \)), she will receive zero ads on product j; otherwise she continues following firm j’s account (i.e., \( f_j = 1 \)) and will receive \( n_j \) ads on product j.

Lastly, consumers read the ads, update their posterior beliefs and decide which product to consume.

We assume that consumers have limited attention—a consumer can read no more than \( M > 0 \) ads. Consumers read ads sequentially, not being able to choose which ads to read, and there is no cost of reading ads within the span of attention. When the total number of ads exceed her attention capacity, i.e., \( f_1 n_1 + f_2 n_2 > M \), the consumer will read \( M \) ads and the number of ads she read from each firm is proportional to the posting intensity \( n_j \); otherwise, she will read all ads available to her. This is a reasonable assumption considering a typical social media site, where ads from different accounts are mixed and shown in a reverse-chronological order, so it is very difficult for a consumer to cherrypick.\(^7\) Based on the assumptions, the number of ads the consumer will read from firm j is

\[
R_j(f_1, f_2; n_1, n_2) = \begin{cases} 
\frac{f_j n_j}{f_1 n_1 + f_2 n_2} M & f_1 n_1 + f_2 n_2 > M, \\
\frac{n_j}{f_1 n_1 + f_2 n_2} & f_1 n_1 + f_2 n_2 \leq M,
\end{cases} \quad j = 1, 2. \tag{3.3}
\]

It is worthwhile noticing that the expression of \( R_j \) in equation (3.3) implies the externality of firms’ choice of posting intensities—\( R_j \) depends on not only \( n_j \) but also \( n_{j'} (j' \neq j) \). An increase in the other firm’s posting intensity may negatively impact the readership of firm j’s ads due to the limited attention of consumers. For tractability of the model, we assume that \( \bar{n}_j > M, j = 1, 2 \). Under this assumption, where consumers receive firms’ ads continuously over time. Before a consumer decides whether to forgo all future ads from a firm, he has a good knowledge about the firm’s posting intensity. This assumption can be also justified in our current model, if we allow a consumer to sample a small portion of the ads and infer each firm’s posting intensity, before deciding whether to unfollow the firm.

\(^7\)Weibo provides a function that allows consumers to choose to read tweets from category of accounts, such as celebrity accounts, media accounts, etc. Within each category, consumers still cannot choose which ads to read. Our model also applies if consumers only read tweets from a certain category. Given the short length of ads (tweets), we do not consider the difference between “preview an ad” and “read an ad”, and it is reasonable to assume that each ad takes the same amount of attention.
each firm \( j \) alone can deplete a consumer's attention capacity.

After reading the ads, the consumer's belief about her match value with product \( j \) is updated. We denote the posterior belief as \( V_j|\{S_{jk}\}_{k=1}^{R_j} \sim N(\mu_j, \sigma_j^2) \), where according to Bayes' rule,

\[
\begin{align*}
\mu_j &= \frac{\omega^2}{\omega^2 + R_j \sigma_j^2} \mu_{0j} + \frac{R_j \sigma_j^2}{\omega^2 + R_j \sigma_j^2} \cdot \frac{\sum_{k=1}^{R_j} S_{jk}}{R_j}, \\
\sigma_j^2 &= \frac{\sigma_{0j}^2 \omega^2}{\omega^2 + R_j \sigma_{0j}^2}.
\end{align*}
\]

The posterior beliefs will be independent across product \( j = 1, 2 \), because both the prior beliefs and signals are independent across products.

Based on the posterior belief, the consumer decides which product to consume. We assume that each consumer is risk-neutral, having a unit demand and no outside option. Under this assumption, the consumer will choose to consume the product with the highest posterior expected match value \( \mu_j \), and the consumer's expected utility will be \( \max\{\mu_1, \mu_2\} \).

So far, we have completely outlined the model setup. We will solve the model by backward induction below. We first analyze consumers' unfollowing decisions and then firms' competitive advertising decisions.

### 3.1 Rational Unfollowing

At the time point when deciding whether to unfollow the two firms, a consumer has observed both firms' posting intensities, but has not read any ads. Her objective is to maximize the expectation of the maximum of the posterior expected match values, \( E_0[\max\{\mu_1, \mu_2\}] \), where the expectation \( E_0[\cdot] \) is taken over all possible values of \( \mu_j, j = 1, 2 \) from an ex-ante point of view before the signals realize. At this point, the posterior expected match value, \( \mu_j \), is uncertain and follows a normal distribution
\[ N(\mu_0j, \Sigma_j), \] where

\[
\Sigma_j = \left( \frac{R_j \sigma_{0j}^2}{\omega^2 + R_j \sigma_{0j}^2} \right)^2 \left( \sigma_{0j}^2 + \frac{\omega^2}{R_j} \right)
\]

\[
= \frac{R_j \sigma_{0j}^4}{\omega^2 + R_j \sigma_{0j}^2}. \tag{3.6}
\]

\(\Sigma_j\) measures how much the posterior mean \(\mu_j\) can differ from the prior mean \(\mu_0j\). The uncertainty in the value of \(\mu_j\) comes from two aspects: one is the average variation of signals around the true match value, measured by \(\frac{\omega^2}{R_j}\), and the other is the uncertainty in the true match value, i.e., \(\sigma_{0j}^2\). It is easy to see that \(\Sigma_j\) increases in \(R_j\), i.e., by reading more ads, consumers obtain more information to update the belief, so that the range of potential change in belief gets larger.\(^{18}\)

Based on the distribution of \(\mu_j\), we can calculate \(E_0[\max\{\mu_1, \mu_2\}]\) by applying the formula in Nadarajah and Kotz (2008).

\[
E_0[\max\{\mu_1, \mu_2\}] = \mu_{01} \Phi \left( \frac{\mu_{01} - \mu_{02}}{\sqrt{\Sigma_1 + \Sigma_2}} \right) + \mu_{02} \Phi \left( \frac{\mu_{02} - \mu_{01}}{\sqrt{\Sigma_1 + \Sigma_2}} \right)
\]

\[
= + \sqrt{\Sigma_1 + \Sigma_2} \Phi \left( \frac{\mu_{01} - \mu_{02}}{\sqrt{\Sigma_1 + \Sigma_2}} \right). \tag{3.7}
\]

where \(\Phi(\cdot)\) and \(\phi(\cdot)\) represent the CDF and PDF of the standard normal distribution respectively. According to equation (3.7), the derivative of \(E_0[\max\{\mu_1, \mu_2\}]\) with respect to \(\frac{\Sigma_1 + \Sigma_2}{\sqrt{\Sigma_1 + \Sigma_2}}\) is equal to \(\phi \left( \frac{\mu_{01} - \mu_{02}}{\sqrt{\Sigma_1 + \Sigma_2}} \right) > 0\), so \(E_0[\max\{\mu_1, \mu_2\}]\) strictly increases with \(\Sigma_1 + \Sigma_2\). This is intuitive—as \(\Sigma_j\) increases, the potential value of \(\mu_j\) gets more dispersed, and the expectation of the maximum of \(\mu_1\) and \(\mu_2\) gets greater. Moreover, we have shown that \(\Sigma_j\) increases with \(R_j\) according to equation (3.6). Therefore, holding others constant, \(E_0[\max\{\mu_1, \mu_2\}]\) will also increase with \(R_j\), which means that as a consumer reads more ads, she gets more information to make a better choice, and thus expects a higher expected match value.

\(^{18}\)One may notice that \(\Sigma_j\) also increases with the variance of prior belief, \(\sigma_{0j}^2\), for two reasons: first, the consumer will put more weights on the advertising signals when the prior belief is not precise; second, a higher \(\sigma_{0j}^2\) also means that the uncertainty in the underlying true value \(V_j\) is larger. If the consumer reads no ads with \(R_j = 0\), then \(\mu_j = \mu_0j\) for sure, so \(\Sigma_j = 0\); if the consumer reads infinite number of ads with \(R_j \to \infty\), \(\mu_j\) will converge to \(V_j\) ex post, with the range of potential change equal to the prior uncertainty, i.e. \(\Sigma_j = \sigma_{0j}\).
We notice that \( E_0[\max\{\mu_1, \mu_2\}] \) depends on \( R_1 \) and \( R_2 \) (and thus \( f_1 \) and \( f_2 \)) solely via \( \Sigma_1 + \Sigma_2 \), so the consumer’s objective function can be equivalently simplified as to maximize

\[
\Sigma_S(f_1, f_2; n_1, n_2) = \Sigma_1(R_1(f_1, f_2; n_1, n_2)) + \Sigma_2(R_2(f_1, f_2; n_1, n_2)) = \frac{R_1(f_1, f_2; n_1, n_2)\sigma_{01}^2}{\omega^2 + R_1(f_1, f_2; n_1, n_2)\sigma_{01}^2} + \frac{R_2(f_1, f_2; n_1, n_2)\sigma_{02}^2}{\omega^2 + R_2(f_1, f_2; n_1, n_2)\sigma_{02}^2}, \tag{3.8}
\]

where \( R_j(f_1, f_2; n_1, n_2) \) is given by (3.3). In fact, \( \Sigma_S = \Sigma_1 + \Sigma_2 \) can be viewed as a measure of how much information consumers can get from the ads. The more information consumers get from reading the ads, the better decision they will make when choosing the product.

Before solving the consumer’s optimization problem, we first notice that her objective function in equation (3.8) does not depend on \( \mu_{0j} \). Formally, we have the following proposition.

**Proposition 3** When there are two firms, only the firm with a smaller prior uncertainty \( \sigma_{0j}^2 \) may get unfollowed by consumers.

It implies that even though a priori, a consumer may like product 1 better, i.e., \( \mu_{01} > \mu_{02} \), she will not be more likely to keep following firm 1, ceteris paribus. This is because without following firm 1 to receive its ads, a consumer can still choose to purchase product 1 later; the unfollowing decision is purely driven by information maximization—ads from a bad product can be equally or even more informative as those from a good product for the sake of facilitating choice making. In other words, the usefulness of ads depends on how uncertain people are about the product’s value, but does not depend on how well people perceive the product a priori.

Now, let us solve the consumer’s optimization problem: \( \max_{f_1, f_2} \Sigma_S(f_1, f_2; n_1, n_2) \) given firms’ posting intensity \( (n_1, n_2) \). As \( f_j \in \{0, 1\} \), to determine which firm to unfollow, the consumer only needs to compare \( \Sigma_S(1, 1; n_1, n_2) \), \( \Sigma_S(1, 0; n_1, n_2) \), \( \Sigma_S(0, 1; n_1, n_2) \), and \( \Sigma_S(0, 0; n_1, n_2) \). We first notice that

\[
\Sigma_S(f_1, f_2; n_1, n_2) > \Sigma_S(0, 0; n_1, n_2) = 0
\]
for $\forall f_1 + f_2 > 0, \forall n_1 + n_2 > 0$, i.e., it is never optimal for a consumer to unfollow both firms, because having some information is always better than having no information. The following theorem summarizes the comparison among $\Sigma_S(1,1; n_1, n_2)$, $\Sigma_S(1,0; n_1, n_2)$, and $\Sigma_S(0,1; n_1, n_2)$.

**Theorem 1** When $n_1 + n_2 \leq M$, the consumer always follow both accounts, i.e., $f_1^* = f_2^* = 1$.

When $n_1 + n_2 > M$, the consumer's decision is

$$(f_1^*, f_2^*) = \begin{cases} (0,1), & \text{if } \sigma_{01}^2 < \sigma_{02}^2 \text{ and } \sigma_{01}^2 \leq \sigma_{02}^2 \text{ and } n_1 \geq M, \\ (1,0), & \text{if } \sigma_{02}^2 < \sigma_{01}^2 \text{ and } \sigma_{02}^2 \leq \sigma_{01}^2 \text{ and } n_2 \geq M, \\ (1,1), & \text{otherwise}, \end{cases}$$

where $M_0 = \frac{\omega^2(\sigma_{01}^2 - \sigma_{02}^2)}{\sigma_{01}^2 \sigma_{02}^2}$, $\tilde{n}_1(n_2)$ solves

$$M = \frac{-\Delta + \sqrt{\Delta^2 + 4n_1n_2\sigma_{01}^2\sigma_{02}^2(n_2\sigma_{01}^2\sigma_{02}^2 + (\sigma_{01}^2 + \sigma_{02}^2)\omega^2)}}{2n_1n_2\sigma_{01}^2\sigma_{02}^2 [n_2\sigma_{01}^2\sigma_{02}^2 + \omega^2(\sigma_{01}^2 + \sigma_{02}^2)]},$$

and $\tilde{n}_2(n_1)$ solves

$$M = \frac{-\Delta + \sqrt{\Delta^2 + 4n_1n_2\sigma_{01}^2\sigma_{02}^2(n_1\sigma_{01}^2\sigma_{02}^2 + (\sigma_{01}^2 + \sigma_{02}^2)\omega^2)}}{2n_1n_2\sigma_{01}^2\sigma_{02}^2 [n_1\sigma_{01}^2\sigma_{02}^2 + \omega^2(\sigma_{01}^2 + \sigma_{02}^2)]},$$

$$\Delta \equiv \omega^2(n_1\sigma_{01}^4 + n_2\sigma_{02}^4) - n_1n_2\sigma_{01}^2\sigma_{02}^2|\sigma_{02}^2 - \sigma_{01}^2|.$$

Figure 3-2 illustrates the consumer's decision of unfollowing firm 1 or firm 2.
When \( n_1 + n_2 \leq M \), the consumer will not unfollow either firm, because she is able to read all available ads now and unfollowing will only reduce the amount of information she can get. Now let’s observe the case of firm 1 having a smaller prior uncertainty \( (\sigma_{01}^2 < \sigma_{02}^2) \). The consumer knows more about firm 1 than firm 2 \textit{ex ante} and thus information from firm 2 is more valuable than that from firm 1. Given that there are only two firms, it is never optimal for the consumer to only read ads from firm 1, so the consumer will not unfollow firm 2 in this case. When firm 1 post too many ads and crowd-out the ads from firm 2, the consumer may want to unfollow firm 1 so that she can make room for more information from firm 2. When \( n_2 < M \), the consumer will have less ads to read if she unfollows firm 1. The closer \( n_2 \) is to \( M \), the decline in the number of available ads caused by unfollowing is smaller. Therefore, the boundary condition for the consumer to unfollow firm 1 is a decreasing function of \( n_2 \), meaning that the consumer is more likely to unfollow firm 1 when \( n_2 \) increases within in \([0, M]\). When \( n_2 \geq M \), the number of ads the consumer can read will not change (still equal to \( M \)) if she unfollows firm 1. In such case, the consumer will unfollow firm 1 for sure if her attention is very limited \((M \leq M_0)\), and will unfollow firm 1 when firm 1 posts too intensively compared to firm 2 \((n_1 > \frac{M-M_0}{M_0}n_2)\) if her attention is not that limited \((M > M_0)\). The case of firm 2 having a smaller prior uncertainty \((\sigma_{01}^2 > \sigma_{02}^2)\) is symmetric.

### 3.2 Rational Spamming

Now we ask, knowing that consumers have limited attention and may unfollow a firm rationally, what should be firms’ advertising strategy. We find that competitive firms can be strategic in advertising intensively.

The ultimate goal of a firm should be to maximize the number of people who will consume the product. To achieve this goal, the firm needs to inform more people of the product, and also to maximize the probability that an informed consumer would choose its product to consume. The followers of a firm on social media are already informed of the product. Their perception of the product quality can spread out through word-of-mouth and affect the decision of others. Thus for now we do
not consider the role of advertising in informing more consumers, but focus on its function of influencing existing followers' perception of the product quality and thus affecting the probability that an informed consumer will choose the product. We will show that merely the function of influencing people's perception of the product can already lead firms to advertise very intensively. We assume the marginal cost of posting an advertisement is negligible.

When deciding which product to consume, the consumer has formed posterior belief \( V_j | \{ S_j k \}_{k=1}^{k=R_j} \sim (\mu_j, \sigma_j^2) \) and will choose the product with higher \( \mu_j \). Before consumers read ads, the value of \( \mu_j \) follows the distribution \( N(\mu_{0j}, \Sigma_j) \). Therefore, when each firm \( j \) decides \( n_j \), its objective is to maximize

\[
\Pi_j = E_0[1(\mu_j > \mu_{j'})] = Pr(\mu_j > \mu_{j'}) = \Phi(\frac{\mu_{0j} - \mu_{0j'}}{\sqrt{\Sigma_1 + \Sigma_2}}), \tag{3.10}
\]

where \( \Phi(\cdot) \) denotes the CDF of standard normal distribution.

As the value of \( \mu_{01}, \mu_{02} \) are known, both firms' objective function are determined by the value of \( \Sigma_1 + \Sigma_2 \). Notice that \( \Phi(\cdot) \) is an increasing function and we assume \( \mu_{01} - \mu_{02} > 0 \), so firm 1's objective function decreases with \( \Sigma_1 + \Sigma_2 \), whereas firm 2's objective function increases in \( \Sigma_1 + \Sigma_2 \). Therefore, firm 1's objective is equivalent to minimizing \( \Sigma_1 + \Sigma_2 \), whereas firm 2's objective can be transformed as maximizing \( \Sigma_1 + \Sigma_2 \). That is, firm 1 wants to minimize the total amount of information conveyed to consumers to keep its advantage, and firm 2 wants to maximize the information conveyed to consumers to change the current situation.

We need to notice that the two firms' advertising intensities influence \( \Sigma_1 + \Sigma_2 \) not only by affecting the amount of information available to consumers, but also by influencing consumers' unfollowing decisions. We denote the objective function as \( \Sigma_S(n_1, n_2) \equiv \Sigma_S(f_1^*, f_2^*; n_1, n_2) = \Sigma_1(R_1(f_1^*, f_2^*; n_1, n_2)) + \Sigma_2(R_2(f_1^*, f_2^*; n_1, n_2)) \), which is the value of \( \Sigma_1 + \Sigma_2 \) given consumers' unfollowing decision \( (f_1^*, f_2^*) \) under \( (n_1, n_2) \). \( R_j(f_1^*, f_2^*; n_1, n_2) \) is the number of ads that consumers will read from firm \( j \) given their unfollowing decision \( (f_1^*, f_2^*) \) under \( (n_1, n_2) \).

Knowing the two firms' objective, we characterize the two firms' posting strategy.
in equilibrium, taking consumers’ unfollowing decision into account. When a firm is indifferent in a range of advertising intensities, we impose an equilibrium selection criterion that the firm will choose the intensity that is optimal if there is a proportion $\alpha$ of consumers who will never choose to unfollow either firm, where $\alpha$ can be any value in $(0, 1)$. Proposition 4 gives the two firms’ posting intensities and consumers’ corresponding unfollowing decision in equilibrium.

**Proposition 4** Suppose firm 1 is perceived as having a higher quality offering than firm 2 a priori ($\mu_{01} > \mu_{02}$). When the prior belief about firm 1 is more precise ($\sigma_{01}^2 < \sigma_{02}^2$) and consumers have very limited attention ($M \leq \left(1 + \frac{n_1}{n_2}\right)M_0$), both firms will post intensively to the upper bound but rational consumers will unfollow firm 1 in equilibrium. Otherwise, firm 1 will not advertise, firm 2 will take up consumers’ attention, and consumers will keep following both firms in equilibrium.

(See Appendix for proof)

To understand the result of Proposition 4, let’s first consider how the objective function changes with the two firms’ posting intensity if consumers always keep following both firms, i.e., how $\Sigma_S(1, 1; n_1, n_2)$ changes with $n_1, n_2$. Given limited attention of the consumer, for each firm $j$, posting more ads can have two effects on $\Sigma_S(1, 1; n_1, n_2)$. The direct effect is that it makes the consumer read more of its own ads ($R_j$ increases), and thus the range of potential change in the belief of product $j$’s quality becomes larger ($\Sigma_j$ larger). We call this “dispersion effect”. The indirect effect is that when the total posting intensity exceeds consumers’ attention capacity ($n_1 + n_2 \geq M$), more ads from firm $j$ will crowd out the ads from its competitor ($R_{j'}$ decreases), making the potential change in the belief about the competitor smaller ($\Sigma_{j'}$ smaller). We call this “crowd-out effect”. Firm 1 favors crowd-out effect which makes $\Sigma_2$ smaller, but dislikes dispersion effect which makes $\Sigma_1$ larger. Firm 2 likes the dispersion effect, but dislikes the crowd-out effect.

The key driving force of the result in Proposition 4 is that firm 2’s interest is completely aligned with consumers’, but firm 1’s objective is completely opposite to consumers’. Consumers and firm 2 all want to maximize the information from ads,
whereas firm 1 wants to keep its advantage by minimizing the information consumers can get from ads. When consumers find reading ads from firm 1 will get them more information and increase their expected utility, firm 1 would rather not advertise. When consumers find firm 1’s intensive advertising crowds out the ads from firm 2 and reduces the total amount of information they can get, firm 1 would like to advertise intensively. In both cases, the best choice of firm 2 is to advertise intensively so as to maximize the information conveyed by ads. Firm 2 does not worry about crowd-out effect in equilibrium, since firm 1’s will either not advertise or advertise to the level that rational consumers will unfollow firm 1. Consumers’ unfollowing does not prevent firm 1 from advertising intensively, because reducing advertising intensity to the level that can retain followers will increase the amount of information conveyed by ads and make firm 1 worse off. Firm 2’s choice is in line with consumers’ interest and thus consumers will not unfollow firm 2 in equilibrium.

Below is a sketch of the proof to illustrate the idea more closely. We first consider firm 1’s optimal choice of $n_1$ given $n_2$. When $n_1 < M - n_2$, a marginal increase in $n_1$ only has the dispersion effect and thus $\Sigma_S(1, 1; n_1, n_2)$ strictly increases with $n_1$. When $n_1 \geq M - n_2$, both the dispersion effect and the crowd-out effect exist. Notice that $\Sigma_j$ changes with $R_j$ faster when $\sigma^2_{0,j}$ is larger\(^{19}\). That is, when the prior belief about firm $j$ is less precise, the number of ads to read from firm $j$ will have a larger impact on its magnitude of potential change in belief, because consumers will put more weight on signals from ads when updating their belief.

When $\sigma^2_{01} \geq \sigma^2_{02}$, $\Sigma_1$ changes with $R_1$ at a faster or equal rate than does $\Sigma_2$ change with $R_2$. We show that in such case, the overall magnitude of dispersion effect is always greater than that of the crowd-out effect. That is, given any $n_2 > 0^{20}$, $\Sigma_S(1, 1; n_1, n_2)$ either strictly increases with $n_1$ or first increases and then decreases in $n_1$, but for any positive $n_1$, $\Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2)$. Then firm 1’s optimal choice is $n_1(n_2) = 0$ for any positive $n_2$. Firm 1 will not choose any positive ad-

\(^{19}\) $\frac{\partial \Sigma_j}{\partial R_j}$ is an increasing function of $\sigma^2_{0,j}$.

\(^{20}\) If $n_2 = 0$, there is no crowd-out effect of advertising for firm 1, and thus firm 1 will choose $n_1 = 0$. This cannot be an equilibrium since firm 2 has incentive to deviate by raising $n_2$. Therefore we can ignore $n_2 = 0$ in the following analysis.
vertising intensity because that always leads to a larger objective function whatever consumers’ unfollowing decision is: for any $n_1 > 0$, $\sum_S(n_1, n_2) \geq \sum_S(1, 1; n_1, n_2) > \sum_S(1, 1; 0, n_2) = \sum_S(0, n_2)$, where the first inequality comes from the definition that consumers choose $(f_1^*, f_2^*)$ to maximize $\sum_S(f_1, f_2; n_1, n_2)$.

When $\sigma_{01}^2 < \sigma_{02}^2$, $\Sigma_1$ changes with $R_1$ at a slower rate than does $\Sigma_2$ change with $R_2$. In such case, the magnitude of crowd-out effect can exceed that of the dispersion effect when $n_1$ is large enough. Given any $n_2 > 0$, $\sum_S(1, 1; n_1, n_2)$ first increases and then decreases in $n_1$. Denote $n_1(n_2)$ as the value of $n_1 > 0$ that solves $\sum_S(1, 1; n_1, n_2) = \sum_S(1, 1; 0, n_2)$, i.e., it is the threshold of firm 1’s advertising intensity that makes the magnitude of crowd-out effect exceed the dispersion effect. If $n_1(n_2) > n_1$, then $\sum_S(1, 1; n_1, n_2) > \sum_S(1, 1; 0, n_2)$ for any $n_1 \in (0, n_1)$, i.e., the magnitude of crowd-out effect cannot exceed that of the dispersion effect when $n_1 \leq n_1$. Following the same logic as in the last paragraph, firm 1’s optimal choice is $n_1(n_2) = 0$ in such case. If $n_1(n_2) \leq n_1$, then consumers will unfollow firm 1 when firm 1 chooses $n_1 > n_1(n_2)$. Knowing that, firm 1 will be indifferent in choosing any $n_1 \in \{0\} \cup [n_1(n_2), n_1]$ but will not choose $n_1 \in (0, n_1(n_2))$, because $n_1$ in the former set will lead to $\sum_S(n_1, n_2) = \sum_S(0, n_2)$ whereas $n_1$ in the latter set will lead to $\sum_S(n_1, n_2) \geq \sum_S(1, 1; n_1, n_2) > \sum_S(1, 1; 0, n_2) = \sum_S(0, n_2)$. Applying our equilibrium selection rule, firm 1 will choose $n_1(n_2) = n_1$ since it minimizes $\sum_S(1, 1; n_1, n_2)$ in such case.

So far we have shown that given any $n_2$, firm 1’s optimal choice is either $n_1(n_2) = 0$ or $n_1(n_2) = n_1$, and consumers must unfollow firm 1 when firm 1 chooses $n_1 = n_1$. In whichever case, rational consumers in fact only read ads from firm 2. Firm 2 will be indifferent in any $n_2 \in [M, \bar{n}_2]$ since they all make consumers’ attention occupied by firm 2’s ads. For $(f_1, f_2; n_1, n_2) = (0, 1; n_1, n_2)$ with $n_2 \geq M$ to be an equilibrium outcome, we must have $n_1(n_2) = \frac{M - M_0}{M_0} n_2 \leq n_1$. We need this inequality hold for any $n_2 \in [M, \bar{n}_2]$, because otherwise firm 2 can increase $n_2$ to some $n'_2$ such that $n_1(n'_2) > n_1$, which makes consumers keep following both firms under $(\bar{n}_1, n'_2)$ and $\sum_S$ larger.21 Thus the necessary and sufficient condition for $(0, 1; n_1, n_2)$ with $n_2 \in [M, \bar{n}_2]$

\[21\text{When } n_1(n'_2) > n_1, \Sigma_S(1, 1; n_1(n'_2)) > \Sigma_S(1, 1; 0, n'_2), \text{so consumers will keep following both firms. } \Sigma_S(n_1, n'_2) = \Sigma_S(1, 1; n_1, n'_2) > \Sigma_S(0, 1; n_1, n'_2) = \Sigma_S(0, 1; n_1, n_2) = \Sigma_S(\bar{n}_1, n_2).\]
to be an equilibrium is $\sigma_{01}^2 < \sigma_{02}^2$ and $\frac{M-M_0}{M_0} \bar{n}_2 \leq \bar{n}_1$. Applying our equilibrium selection rule, firm 2 will choose $n_2^* = \bar{n}_2$ since it maximizes $\Sigma_S(1, 1; \bar{n}_1, n_2)$ in such condition. When the condition is not satisfied, the equilibria will be $(1, 1; 0, n_2)$ with $n_2 \in [M, \bar{n}_2]$.

Lastly, we analyze the comparative statics. According to Proposition 4, when $\sigma_{01}^2 < \sigma_{02}^2$ and

$$M \leq \left( 1 + \frac{\bar{n}_1}{\bar{n}_2} \right) M_0 = \left( 1 + \frac{\bar{n}_1}{\bar{n}_2} \right) \omega^2 \left| \frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2} \right|,$$

both firms will post to the upper bound $((n_1^*, n_2^*) = (\bar{n}_1, \bar{n}_2))$; otherwise, firm 1 will not post and firm 2 will take up consumers’ attention $(n_1^* = 0, n_2^* \in [M, \bar{n}_2])$. Figure 3-3 illustrates the boundary condition in terms of $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2}$ and $M$.

$1/\sigma_{0j}^2$ measures the precision of prior belief about firm $j$. When $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2} > 0$, the boundary condition of $M$ for both firms to advertise intensively is a linear function of $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2}$ with slope $\left( 1 + \frac{\bar{n}_1}{\bar{n}_2} \right) \omega^2$. This leads to the following implication.

The upper bound of $M$ for firm 1 to advertise intensively increases with $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2}$ when $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2} > 0$. That is, when the prior belief about firm 1 is more precise (equivalent to $\sigma_{01}^2$ smaller), firm 1’s advantage in prior belief becomes more stable, and consumers will put less weight on the signals when updating their belief about firm 1. Hence the risk brought by dispersion effect becomes smaller. Similarly, when the prior belief about firm 2 is less precise (equivalent to $\sigma_{02}^2$ larger), consumers will put more weight on the signals when updating their belief about firm 2. Then the benefit of the crowd-out effect becomes stronger. Therefore, either an increase in $\frac{1}{\sigma_{01}^2}$ or a decrease in $\frac{1}{\sigma_{02}^2}$ will lead to a higher upper bound for $M$, i.e., firm 1 becomes more likely to post intensively.

Given any $\frac{1}{\sigma_{01}^2} - \frac{1}{\sigma_{02}^2} > 0$, the upper bound of $M$ for firm 1 to advertise intensively increases in $\omega^2$. When $\omega^2$ is larger, the signals conveyed by ads are less precise about the true match value and consumers will put less weight on the signals when updating their beliefs, which reduces the risk brought by dispersion effect. Hence firm 1 becomes more likely to advertise intensively.
Given any \( \frac{1}{\sigma_{01}} - \frac{1}{\sigma_{02}} > 0 \), the upper bound of \( M \) for firm 1 to advertise intensively increases in \( \bar{n}_1 \) and decreases in \( \bar{n}_2 \). Recall that when \( \sigma_{01}^2 < \sigma_{02}^2 \), the overall magnitude of crowd-out effect exceeds that of the dispersion effect only when \( n_1/n_2 \) is large enough. That is, when firm 1 is able to post a large enough number of ads relative to firm 2, advertising intensively tends to be more attractive than no advertising for firm 1 since it is able to crowd out a large enough proportion of firm 2's ads.

3.3 Extensions

Endogenous Noisiness of Signals

In our main model, each ad is a noisy signal of the true value \( V_j \), following the distribution \( S_{jk}|V_j \sim N(V_j, \omega^2) \), and the noisiness of signals, \( \omega^2 \), is exogenously given. We extend the model by assuming that each firm \( j \) can have different noisiness of signals \( \omega_j^2 \) and it is endogenously chosen by the firm on \([\omega^2, \infty)\) where \( \omega^2 > 0 \). The lower bound \( \omega^2 \) is positive since we consider a context that firms are not able to fully reveal the content of the product in ads. Ads with \( \omega_j^2 \to \infty \) are uninformative of the product at all.

As noisiness of signals becomes endogenous, we denote

\[
\Sigma_S(n_1, n_2; \omega_1^2, \omega_2^2) = \frac{R_1(f_1^*, f_2^*; n_1, n_2)\sigma_{01}^4}{\omega_1^2 + R_1(f_1^*, f_2^*; n_1, n_2)\sigma_{01}^2} + \frac{R_2(f_1^*, f_2^*; n_1, n_2)\sigma_{02}^4}{\omega_2^2 + R_2(f_1^*, f_2^*; n_1, n_2)\sigma_{02}^2}. \tag{3.12}
\]

Firm 1 minimizes \( \Sigma_S(n_1, n_2; \omega_1^2, \omega_2^2) \) by choosing \( \omega_1^2, n_1 \), and firm 2 maximizes \( \Sigma_S(n_1, n_2; \omega_1^2, \omega_2^2) \) by choosing \( \omega_2^2, n_2 \). The value of \( \omega_j^2 \) affects \( \Sigma_j \) directly, and also affects \( \Sigma_1, \Sigma_2 \) by influencing \( R_1(f_1^*, f_2^*; n_1, n_2), R_2(f_1^*, f_2^*; n_1, n_2) \).

Both firms choose \( \omega_j^2, n_j \) first \((j = 1, 2)\), and then consumers make the unfollowing decision \((f_1, f_2)\). The rest of the game is the same as the main model. We are able to show the following proposition.
Proposition 5 In equilibrium, firm 1 may post ads with any informativeness level \((\omega_1^2 \in [\omega^2, \infty))\) and firm 2 will post the most informative ads \((\omega_2^2 = \omega^2)\).

(Detailed proof is available in Appendix.)

We first show that given any \(\omega_2^2, n_2\), firm 1 will choose either \(n_1(n_2) = 0\) or \(n_1(n_2) = \bar{n}_1\), and is indifferent in choosing any \(\omega_1^2 \in [\omega^2, \infty)\). In the former case, consumers will keep following both firms, and in the latter case, consumers will unfollow firm 1. In either case, firm 2 will be indifferent in choosing any \(n_2 \in [M, \bar{n}_2]\), and

\[
\Sigma_S(n_1^*, n_2^*; \omega_1^2, \omega_2^2) = \frac{M\sigma_2 \omega_1^2}{\omega_1^2 + M\sigma_2},
\]

which is decreasing in \(\omega_2^2\) but not affected by \(\omega_1^2\) at all. Therefore, firm 1 is indifferent in any \(\omega_1^2 \in [\omega^2, \infty)\) and firm 2 will choose \(\omega_2 = \omega^2\) in equilibrium.

If there is an \(\alpha\) proportion of consumers who never unfollow either firm, firm 1 will post uninformative advertisements only \((\omega_1^* \rightarrow \infty)\), because

\[
\Sigma_S(1, 1; n_1, n_2) = \frac{R_1(1, 1; n_1, n_2)\sigma_0^4}{\omega_1^2 + R_1(1, 1; n_1, n_2)\sigma_0^2} + \frac{R_2(1, 1; n_1, n_2)\sigma_0^4}{\omega_2^2 + R_2(1, 1; n_1, n_2)\sigma_0^2}
\]

is decreasing in \(\omega_1^2\) (\(\omega_1^2\) does not affect \(R_j(1, 1; n_1, n_2), j = 1, 2\)).

Intuitively, if consumers do not choose to unfollow firms, firm 1 will choose to send uninformative ads only, as they can crowd out the ads from the competitor and have no dispersion effect at all. Being able to send uninformative ads only, firm 1 has no concern of dispersion effect when deciding advertising intensity. In other words, there will be only crowd-out effect when firm 1 increases its advertising intensity, so firm 1 will always advertise intensively to the upper bound. If consumers do choose to unfollow firms rationally and firm 1 internalize their choice, firm 1 will be indifferent in any informative level, since firm 1’s choice of posting intensity and consumers’ unfollowing will make the value of objective function always equal to \(\Sigma_S(0, n_2)\), i.e., equivalent to firm 1 not advertising. Firm 2 will choose to send the most informative ads, in hope of reversing the current situation. The dispersion effect of ads are strengthened when ads become more informative.

We say that a firm is currently more popular if it has a larger prior mean. Our result implies that a currently less popular firm will post ads that contain more
Back to our empirical setting, we test this implication by comparing the proportion of retweets posted by shows with different levels of popularity. In our dataset, we classify shows into two groups according to their average search indices. We find the proportion of retweets posted by popular shows is nearly twice as high as that of non-popular shows ($M_{\text{pop}}=0.26$, $M_{\text{non-pop}}=0.14$, $t(10719)=20.2$, $p<0.001$). A retweet is a reposting of a tweet that has been posted by others or by itself. A retweet from other accounts is usually not directly related to the content of the show, and a retweet from itself is repetitive information, so a retweet is generally less informative of the show compared to an original tweet. Therefore, the fact that popular shows are much more likely to post retweets is consistent with the implication above. Distinguishing retweets and original tweets is a rough way of identifying the informativeness of tweets. We may also use machine learning methods to identify the informativeness of tweets and test if popular shows do tend to post tweets that are less informative of show content.

**Existence of a Third Firm**

In the main model, we consider a problem of two competing firms. What if there exists a third firm?

Suppose the prior belief about the three firms are $V_j \sim N(\mu_{0j}, \sigma_{0j})^2, j = 1, 2, 3$. Without loss of generalizability, we assume $\mu_{01} > \mu_{02} > \mu_{03}$. Each firm needs to decide the posting intensity $n_j$. The consumer's problem is to maximize

$$E[\max\{\mu_1, \mu_2, \mu_3\}|R_j(f; n), j = 1, 2, 3]$$

by making following/unfollowing decision $f$ given firms' advertising intensities, where $n = (n_1, n_2, n_3)$, $f = (f_1, f_2, f_3)$, and $f_j \in \{0, 1\}$ represents the consumer's decision of unfollowing or following firm $j$. $R_j(f; n)$ can be written in a similar way as equation (3.3). Anticipating consumers' unfollowing decision $f^*$ as a function of advertising intensity $n$, each firm $j$ decides the advertising intensity $n_j$ to maximize its probability of being chose, $Pr(\mu_j > \mu_{j'}, \forall j' \neq j|f^*; n)$. 

89
We are not able to solve the three-firm problem analytically because the expectation of the maximum of three normal random variables and the probability that one normal random variable is greater than the other two cannot be expressed in a closed form. We consider a special case where $\sigma_{01}^2 < \sigma_{02}^2 < \sigma_{03}^2$. This is a very realistic case: a higher prior mean indicates that the firm is more popular and has more information available to consumers before advertising, so the prior uncertainty in its product quality is likely to be smaller.

We run a numerical simulation to show that in this case, there exists an equilibrium in which all three firms advertise intensively when consumers’ attention capacity $M$ smaller than some threshold (or equivalently $\omega^2$ is relatively large$^{22}$). In the numeric example, we let $(\mu_{01}, \mu_{02}, \mu_{03}) = (3, 2, 1)$, $(\sigma_{01}^2, \sigma_{02}^2, \sigma_{03}^2) = (1, 2, 4)$, $M = 8$, $\bar{n}_j = \bar{n} = 14$, $j = 1, 2, 3$, $\omega^2 = 6^{23}$. We first calculate the optimal following decision $f^*$ given any posting intensity combination $n$, and then calculate $Pr(\mu_j > \mu_{j'}, \forall j' \neq j|f^*; n)$ for each $n$ with $n_j \leq \bar{n}$. $E[\max\{\mu_1, \mu_2, \mu_3\}|R_j(f; n), j = 1, 2, 3]$ and $Pr(\mu_j > \mu_{j'}, \forall j' \neq j|f^*; n)$ are calculated by numerical integration.

Figure 3-4 shows how the probability of firm $j$ being chosen, $Pr(\mu_j > \mu_{j'}, \forall j' \neq j|f^*; n)$, changes with $n_j$ when the other two firms advertise intensively to the upper bound $(n_{j'} = \bar{n}, \forall j' \neq j)$. Given $n_2 = n_3 = \bar{n}$, consumers keep following the three firms when firm 1 chooses $n_1 \in [0, \bar{n}]$. When $n_1$ is relatively small, the probability of firm 1 being chosen decreases with $n_1$ since the dispersion effect dominates the crowd-effect; when $n_1$ becomes larger, the probability of firm 1 being chosen increases with $n_1$ since crowd-out effect plays a major role now. Given $n_1 = n_3 = \bar{n}$, consumers keep following all three firms and the probability of firm 2 being chosen increases with $n_2$ for any $n_2 \in [0, \bar{n}]$. Given $n_1 = n_2 = \bar{n}$, when $n_3$ is very small, consumers

---

$^{22}$Recall that in the two-firm problem, the condition for firm 1 advertise intensively is $\sigma_{01}^2 < \sigma_{02}^2$ and $M < M_0(1 + \frac{\bar{n}_1}{\bar{n}_2})$, where $M_0 = \frac{\omega^2|\bar{n}_2| - \sigma_{01}^2}{\sigma_{01}^2, \sigma_{03}^2}$. Now in the three-firm problem, a condition similar as $M < M_0(1 + \frac{\bar{n}_1}{\bar{n}_2})$ is also needed for the magnitude of crowd-out effect being greater than the dispersion effect and thus firm 1 is willing to advertise intensively. $M_0$ now should depend on all three $\sigma_{0j}^2$’s, but should still increase in $\omega^2$, since a greater $\omega^2$ means that consumers will put less weight on signals and thus the concern of dispersion effect becomes less. Therefore, with $M$ fixed, firm 1 is more likely advertise intensively when $\omega^2$ is larger.

$^{23}$When $\bar{n}_1 = \bar{n}_2 = \bar{n}_3$, their exact value does not affect the equilibrium outcome as long as they are larger than $M$. 

90
keep following all three firms, and when \( n_3 \) is in the middle range, consumers will unfollow firm 1 to get more information from firm 3. When \( n_3 \) becomes even larger, consumers will again keep following all three firms since there is firm 3’s amount of information can take an enough portion now. The probability of firm 3 being chosen increases with \( n_3 \) within each interval, and is maximized when \( n_3 = \bar{n} \). In this numerical example, given the other two firms advertising intensively to the upper bound, each firm \( j \)’s probability of being chosen is maximized when it choose the highest advertising intensity. Therefore, all three firms advertising intensively to the upper bound is an equilibrium in this case.

The intuition is as follow. Firm 1 wants to keep the current situation. When posting more ads, it worries about the dispersion effect and likes the crowd-out effect. Similar as in the main model, given it has a smaller prior uncertainty than the other two firms, the crowd-out effect of a marginal increase in its advertising intensity dominates the dispersion effect when its advertising intensity goes beyond a threshold, and the overall magnitude of crowd out effect can exceed that of the dispersion effect when \( \bar{n}_1 \) is large enough (or equivalently \( M \) is relatively small). Similar as in the main model, firm 1 wants to minimize the information consumers can get, which is opposite to consumers’ objective. Thus it can be optimal for firm 1 to choose the highest advertising intensity even though consumers may unfollow it.

Firm 2 has faces a conflict. On the one hand, it wants to change the current situation and surpass firm 1. On the other hand, it wants to keep its advantage compared to firm 3. We find that although the competition with firm 1 and with firm 3 makes firm 2 have conflicting objectives, they surprisingly lead to a consistent behavior of firm 2, i.e., they both make firm 2 post intensively. In its competition with firm 1, the dispersion effect of posting more messages dominates the crowd-out effect as \( \sigma_{02}^2 > \sigma_{01}^2 \), which makes posting intensively appealing to firm 2. In its competition with firm 3, as \( \sigma_{03}^2 > \sigma_{02}^2 \), the overall magnitude of crowd-out effect can be greater than that of the dispersion effect when \( M \) is relatively small.

Lastly, we consider firm 3’s decision. Standing at the lowest position \( a \ prior \), firm 3 prefers dispersion effect and dislikes crowd-out effect. With \( \sigma_{03}^2 \) larger than \( \sigma_{01}^2, \sigma_{02}^2 \),
the magnitude of dispersion effect is greater than that of the crowd-out effect, so firm 3 will advertise intensively as well.

From consumers’ perspective, the same as in two-firm model, consumers tend to unfollow firms that have smaller $\sigma_{0j}^2$, since its prior belief is relatively precise and thus the informative value of more ads is less compared to other firms. The difference is that in the three-firm problem, $\mu_{0j}$ will play a role in consumers’ unfollowing decision. Consumers wants to pick out the firm with the highest $\mu_j$ after reading the messages, so messages from the firms that have the highest and second highest prior means can be more relevant than those from the third firm, since the two firms have a large chance to be selected than the third firm. Therefore, when $\sigma_{0j}^2$’s are the same, firm 3 can be more likely to be unfollowed when it advertises intensively. When $\sigma_{0j}^2$, $\mu_{0j}$ have opposite ranks, the likelihood of unfollowing depends on the relative difference in these two measures. Different from the two-firm model, consumers’ interest is no longer completely opposite to firm 1 in the three-firm model, so consumers may keep following firm 1 even when firm 1 chooses the highest posting intensity.

3.4 Concluding Remarks

In this chapter, we study competitive firms’ advertising strategy given limited attention of consumers, using the advertising of TV shows on a social media platform as an example. The data we collected from a primary Chinese tweeting website provides evidence that firms advertise intensively, although doing so appears to drive followers away. The analytical model we build suggests that consumers with limited attention may choose to unfollow a firm rationally if she already knows enough about her match value with the firm’s product and the firm advertises too intensively. However, advertising intensively can be an optimal strategy for competitive firms, since the firm perceived as having a lesser quality offering wants to change consumers beliefs about its quality, and the firm perceived as having a higher quality offering would like to crowd out ads from the competitor and keep its advantage.

A limitation of our model is that we only consider the role of advertising in
influencing consumers’ beliefs about product quality, but do not consider the effect of advertising in informing more people of the product. We do this for two reasons. First, usually people who follow a firm’s account are already aware of the products (or the brands), so the main effect of posts made by firm accounts is to spread product-related information rather than to arouse awareness. Although a follower may repost the ads and make more people informed of the show, the effect is hard to quantify. Second, the current framework allows us to find a clear relationship between the firms’ objective functions and their posting intensity, and the problem will be intractable if we also incorporate the role of advertising in informing more people of the show. Even if we take that effect into firm, it is likely that the effect will be an additional reason for the firms to advertise intensively.
3.A Appendix

3.A.1 Figures

Figure 3-1: Average Posting Intensity vs. Popularity

(a) Popularity measured by Avg. Daily (b) Popularity measured by Avg. # Followers
Search Index
Figure 3-2: The Consumer’s Optimal Strategy of Unfollowing

(a) If $\sigma_{01}^2 < \sigma_{02}^2$, $M \leq M_0$

(b) If $\sigma_{01}^2 < \sigma_{02}^2$, $M > M_0$

(c) If $\sigma_{01}^2 > \sigma_{02}^2$, $M \leq M_0$

(d) If $\sigma_{01}^2 > \sigma_{02}^2$, $M > M_0$
Figure 3-3: Comparative Statics

Figure 3-4: The Probability of Being Chosen as a Function of Posting Intensity, Given All Other Firms Post to the Upper Bound Intensity

(a) Firm 1  (b) Firm 2  (c) Firm 3
### 3.A.2 Tables

Table 3.1: Summary Statistics of Accounts

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Number of Followers</td>
<td>818624.51</td>
<td>1572013.77</td>
<td>5836</td>
<td>8565933</td>
<td>49</td>
</tr>
<tr>
<td>Total Number of Tweets</td>
<td>798.67</td>
<td>583.76</td>
<td>43</td>
<td>2463</td>
<td>49</td>
</tr>
<tr>
<td>Max. Hourly Number of Posts</td>
<td>15.45</td>
<td>9.84</td>
<td>3</td>
<td>41</td>
<td>49</td>
</tr>
<tr>
<td>Avg. Daily Search Index</td>
<td>40020.87</td>
<td>59157.41</td>
<td>625</td>
<td>283246</td>
<td>49</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

“Avg.” means average over the six months.
Table 3.2: Hourly Number of Posts, Conditional on Posting

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>p75</th>
<th>p95</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Night</td>
<td>1.37</td>
<td>0.92</td>
<td>1.00</td>
<td>3.00</td>
<td>7</td>
<td>308</td>
</tr>
<tr>
<td>Daytime</td>
<td>1.26</td>
<td>0.66</td>
<td>1.00</td>
<td>2.00</td>
<td>13</td>
<td>18424</td>
</tr>
<tr>
<td>Prime Time</td>
<td>2.51</td>
<td>4.08</td>
<td>2.00</td>
<td>10.00</td>
<td>41</td>
<td>5810</td>
</tr>
<tr>
<td>Total</td>
<td>1.56</td>
<td>2.13</td>
<td>1.00</td>
<td>3.00</td>
<td>41</td>
<td>24542</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24542</td>
</tr>
</tbody>
</table>
Table 3.3: Advertising Intensively May Drive Followers Away, Hourly

<table>
<thead>
<tr>
<th></th>
<th>(1) OLS</th>
<th>(2) Fixed Effects</th>
<th>(3) Time Effects</th>
<th>(4) Lag # Tweets</th>
<th>(5) n = 10</th>
<th>(6) n = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag # Followers</td>
<td>1.000***</td>
<td>1.000***</td>
<td>1.000***</td>
<td>1.000***</td>
<td>1.000***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000000359)</td>
<td>(0.0000219)</td>
<td>(0.0000187)</td>
<td>(0.0000187)</td>
<td>(0.0000186)</td>
<td>(0.0000182)</td>
</tr>
<tr>
<td>Hourly # Tweets</td>
<td>30.00***</td>
<td>32.99***</td>
<td>32.46***</td>
<td>22.58***</td>
<td>30.63***</td>
<td>21.73***</td>
</tr>
<tr>
<td>(Hourly # Tweets-6)+</td>
<td>3.319</td>
<td>(9.452)</td>
<td>(9.056)</td>
<td>(8.366)</td>
<td>(9.200)</td>
<td>(7.827)</td>
</tr>
<tr>
<td>(Hourly # Tweets-10)+</td>
<td>-32.10***</td>
<td>-36.33***</td>
<td>-35.73**</td>
<td>-28.94**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.765)</td>
<td>(14.06)</td>
<td>(14.05)</td>
<td>(14.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show Dynamics</td>
<td></td>
<td></td>
<td></td>
<td>(14.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Search Index</td>
<td>16.20***</td>
<td></td>
<td></td>
<td>16.58***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.752)</td>
<td></td>
<td></td>
<td>(3.681)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Season Dummy</td>
<td>12.61***</td>
<td></td>
<td></td>
<td>11.71***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.066)</td>
<td></td>
<td></td>
<td>(3.250)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show Day Dummy</td>
<td>7.512***</td>
<td></td>
<td></td>
<td>4.781***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.538)</td>
<td></td>
<td></td>
<td>(1.170)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show Time Dummy</td>
<td>5.698***</td>
<td></td>
<td></td>
<td>3.248***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.578)</td>
<td></td>
<td></td>
<td>(1.161)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tweet Dynamics</td>
<td>-14.32*</td>
<td></td>
<td></td>
<td>-9.755**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.218)</td>
<td></td>
<td></td>
<td>(3.682)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. # Being Retweeted</td>
<td>5.42***</td>
<td>9.219***</td>
<td>10.50**</td>
<td>12.92**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.542)</td>
<td>(4.902)</td>
<td>(4.651)</td>
<td>(4.570)</td>
<td>(5.118)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. # Likes</td>
<td>86.78***</td>
<td>87.81***</td>
<td>81.33***</td>
<td>73.37***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8.109)</td>
<td>(31.38)</td>
<td>(29.54)</td>
<td>(28.56)</td>
<td>(29.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Containing Pictures</td>
<td>-43.17***</td>
<td>-39.48**</td>
<td>-42.33**</td>
<td>-33.58**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4.836)</td>
<td>(15.87)</td>
<td>(16.48)</td>
<td>(13.99)</td>
<td>(16.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Containing Hashtags</td>
<td>18.46***</td>
<td>7.883</td>
<td>2.844</td>
<td>-8.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4.562)</td>
<td>(11.26)</td>
<td>(11.82)</td>
<td>(16.12)</td>
<td>(11.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Reposts</td>
<td>-9.545</td>
<td>-16.79</td>
<td>-21.83</td>
<td>-23.09**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5.863)</td>
<td>(12.36)</td>
<td>(13.10)</td>
<td>(13.11)</td>
<td>(13.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show Fixed Effects</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hour-of-Day</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day-of-Week</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week(In-Season)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week(Out-of-Season)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>134749</td>
<td>134749</td>
<td>134749</td>
<td>111544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each observation is a show-hour combination. The dependent variable is the number of followers of each show at the end of each hour. Standard errors are clustered by show and reported in parentheses under parameter estimates. *p < 0.1; **p < 0.05; ***p < 0.01.

Then numbers of observations in Column (4) and (6) are smaller because they require four lags and there are some missing observations in the dataset. Those observations without missing values in the four lags are dropped in the regression.
Table 3.4: Advertising Intensively May Drive Followers Away, 4-Hour Period

<table>
<thead>
<tr>
<th></th>
<th>(1) Fixed Effects</th>
<th>(2) Time Effects</th>
<th>(3) Lag # Tweets</th>
<th>(4) n = 24</th>
<th>(5) n = 24, Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag # Followers</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(0.000113)</td>
<td>(0.0000937)</td>
<td>(0.0000923)</td>
<td>(0.0000930)</td>
<td>(0.0000918)</td>
</tr>
<tr>
<td>4-Hour # Tweets</td>
<td>34.18***</td>
<td>31.21***</td>
<td>26.43***</td>
<td>27.35***</td>
<td>25.16***</td>
</tr>
<tr>
<td></td>
<td>-37.05**</td>
<td>-34.06**</td>
<td>-30.29**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(17.23)</td>
<td>(16.32)</td>
<td>(15.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 4-Hour # Tweets</td>
<td></td>
<td></td>
<td>22.31***</td>
<td>13.45***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.780)</td>
<td></td>
<td>(3.695)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-33.45***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(9.403)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Lag 4-Hour # Tweets-12)+</td>
<td></td>
<td></td>
<td>-48.60**</td>
<td>-46.90**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(18.34)</td>
<td>(17.77)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-31.13***</td>
<td>(9.267)</td>
</tr>
<tr>
<td>Time-Varying Show Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Search Index</td>
<td>0.00112***</td>
<td>0.00112***</td>
<td>0.00110***</td>
<td>0.00110***</td>
<td>0.00107***</td>
</tr>
<tr>
<td></td>
<td>(0.000212)</td>
<td>(0.000222)</td>
<td>(0.000220)</td>
<td>(0.000223)</td>
<td>(0.000221)</td>
</tr>
<tr>
<td></td>
<td>(35.56)</td>
<td>(39.74)</td>
<td>(38.84)</td>
<td>(40.54)</td>
<td>(40.08)</td>
</tr>
<tr>
<td>Show Day Dummy</td>
<td>4.188</td>
<td>18.82</td>
<td>18.87</td>
<td>20.24</td>
<td>20.94</td>
</tr>
<tr>
<td></td>
<td>(16.74)</td>
<td>(15.47)</td>
<td>(15.63)</td>
<td></td>
<td>(15.54)</td>
</tr>
<tr>
<td>Show Time Dummy</td>
<td>-8.750</td>
<td>-63.16***</td>
<td>-68.89***</td>
<td>-63.20***</td>
<td>-66.60***</td>
</tr>
<tr>
<td></td>
<td>(20.11)</td>
<td>(21.24)</td>
<td>(21.50)</td>
<td>(22.36)</td>
<td>(22.25)</td>
</tr>
<tr>
<td>Time-Varying Tweet Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hour # Being Retweeted</td>
<td>0.0581***</td>
<td>0.0567***</td>
<td>0.0546***</td>
<td>0.0559***</td>
<td>0.0546***</td>
</tr>
<tr>
<td></td>
<td>(0.0124)</td>
<td>(0.0134)</td>
<td>(0.0135)</td>
<td>(0.0124)</td>
<td>(0.0127)</td>
</tr>
<tr>
<td>Avg. # Comments</td>
<td>0.182</td>
<td>0.171</td>
<td>0.182</td>
<td>0.174</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.121)</td>
<td>(0.127)</td>
<td>(0.118)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>Avg. # Likes</td>
<td>0.0688***</td>
<td>0.0695***</td>
<td>0.0695***</td>
<td>0.0690***</td>
<td>0.0697***</td>
</tr>
<tr>
<td></td>
<td>(0.0080)</td>
<td>(0.0094)</td>
<td>(0.0092)</td>
<td>(0.0090)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>% of Tweets Containing Pictures</td>
<td>-64.80**</td>
<td>-51.78**</td>
<td>-78.82**</td>
<td>-81.00**</td>
<td>-79.41**</td>
</tr>
<tr>
<td></td>
<td>(29.86)</td>
<td>(32.38)</td>
<td>(32.52)</td>
<td>(31.47)</td>
<td>(31.56)</td>
</tr>
<tr>
<td>% of Tweets Containing Hashtag Topics</td>
<td>51.23</td>
<td>39.76</td>
<td>33.60</td>
<td>43.72</td>
<td>38.50</td>
</tr>
<tr>
<td></td>
<td>(31.38)</td>
<td>(31.65)</td>
<td>(31.30)</td>
<td>(32.00)</td>
<td>(31.60)</td>
</tr>
<tr>
<td>% of Tweets that are Reposts</td>
<td>-16.91</td>
<td>-37.01</td>
<td>-39.07</td>
<td>-31.37</td>
<td>-35.18</td>
</tr>
<tr>
<td></td>
<td>(31.27)</td>
<td>(30.65)</td>
<td>(30.48)</td>
<td>(28.73)</td>
<td>(28.63)</td>
</tr>
<tr>
<td>Show Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Day-of-Week</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Time-of-Day</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week(In-Season)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week(Out-of-Season)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>35139</td>
<td>35139</td>
<td>35139</td>
<td>35139</td>
<td>35139</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

We classify each day into six 4-hour periods. Each observation is a show-period combination. The dependent variable is the number of followers of each show at the end of each 4-hour period. Standard errors are clustered by show and reported in parentheses under parameter estimates. *p < 0.1; **p < 0.05; ***p < 0.01.
3.A.3 Proof of Proposition 4

Given that \( \mu_{01} > \mu_{02} \), as we have analyzed in the main text, firm 1’s objective is to minimize \( \Sigma_1 + \Sigma_2 \), and firm 2’s objective is to maximize \( \Sigma_1 + \Sigma_2 \). We denote \( \Sigma_S(n_1, n_2) = \Sigma_S(f_1^*, f_2^*; n_1, n_2) = \Sigma_1(R_1(f_1^*, f_2^*; n_1, n_2)) + \Sigma_2(R_2(f_1^*, f_2^*; n_1, n_2)) \), which is the value of \( \Sigma_1 + \Sigma_2 \) given \( (n_1, n_2) \) and consumers’ unfollowing decision \( (f_1^*, f_2^*) \) under \( (n_1, n_2) \). \( R_j(f_1^*, f_2^*; n_1, n_2) \) is the number of ads that consumers will read from firm \( j \) given their unfollowing decision \( (f_1^*, f_2^*) \) under \( (n_1, n_2) \).

**Step 1: When \( n_1 + n_2 \geq M \), we investigate how \( \Sigma_S(1, 1; n_1, n_2) \) changes with \( n_1, n_2 \).**

When \( n_1 + n_2 \geq M \), \( R_1(1, 1; n_1, n_2) = \frac{n_1}{n_1 + n_2} M, R_2(1, 1; n_1, n_2) = \frac{n_2}{n_1 + n_2} M \). We will consider two cases below. In the first case, \( n_2 > 0 \),

\[
\Sigma_S(1, 1; n_1, n_2) = \Sigma_1 \left( \frac{n_1}{n_1 + n_2} M \right) + \Sigma_2 \left( \frac{n_2}{n_1 + n_2} M \right) = \frac{n_1}{n_1 + n_2} \frac{\sigma_{01}^4}{\omega^2 + \frac{n_1}{n_1 + n_2} \sigma_{01}^2} + \frac{n_2}{n_1 + n_2} \frac{\sigma_{02}^4}{\omega^2 + \frac{n_2}{n_1 + n_2} \sigma_{02}^2} = \frac{\psi}{1 + \psi} \frac{\sigma_{01}^4}{\omega^2 + \frac{\psi}{1 + \psi} \sigma_{01}^2} + \frac{1}{1 + \psi} \frac{\sigma_{02}^4}{\omega^2 + \frac{1}{1 + \psi} \sigma_{02}^2} = f(\psi),
\]

where \( \psi = \frac{n_1}{n_2} (\psi \geq 0) \). The first order derivative of \( f(\psi) \) is

\[
f'(\psi) = \frac{M(1 + \psi)\omega^2[\sigma_{01}^2 \sigma_{02}^2 + (\sigma_{01}^2 + \sigma_{02}^2)\omega^2]}{(M\psi\sigma_{01}^2 + \psi\omega^2 + \omega^2)^2} \cdot \frac{(1 + \psi)(\sigma_{01}^2 - \sigma_{02}^2)\omega^2 - M(\psi - 1)\sigma_{01}^2 \sigma_{02}^2}{(M\psi\sigma_{02}^2 + \psi\omega^2 + \omega^2)^2}
\]
The denominator is positive and $M(1 + \psi)\omega^2[M\sigma^2_{01}\sigma^2_{02} + (\sigma^2_{01} + \sigma^2_{02})\omega^2] > 0$, so 
$\text{sign}(f'(\psi))=\text{sign}((1 + \psi)(\sigma^2_{01} - \sigma^2_{02})\omega^2 - M(\psi - 1)\sigma^2_{01}\sigma^2_{02})$. Denote

$$
g(\psi) = \frac{1}{\sigma^2_{01}\sigma^2_{02}} [(1 + \psi)(\sigma^2_{01} - \sigma^2_{02})\omega^2 - M(\psi - 1)\sigma^2_{01}\sigma^2_{02}]
= \left(\frac{(\sigma^2_{01} - \sigma^2_{02})\omega^2}{\sigma^2_{01}\sigma^2_{02}} - M\right)\psi + \left(\frac{(\sigma^2_{01} - \sigma^2_{02})\omega^2}{\sigma^2_{01}\sigma^2_{02}} + M\right),
$$

and recall our definition that $M_0 \equiv \left|\frac{(\sigma^2_{01} - \sigma^2_{02})\omega^2}{\sigma^2_{01}\sigma^2_{02}}\right|$.

Under $n_2 > 0$, the relationship between $f(\psi)$ and $\psi$ depends on the comparison between $\sigma^2_{01}$ and $\sigma^2_{02}$. There are three cases.

- **If $\sigma^2_{01} > \sigma^2_{02}$**, then $g(\psi) = (M_0 - M)\psi + (M_0 + M)$.
  
  - If $0 < M \leq M_0$, $g(\psi) > 0$ for any $\psi$, so $f(\psi)$ is strictly increasing on $[0, \infty)$.
  
  - If $M > M_0$, $g(\psi) > 0$ if and only if $\psi < \frac{M + M_0}{M - M_0}$. Therefore, $f(\psi)$ increases in $\psi$ when $\psi \leq \frac{M + M_0}{M - M_0}$, and decreases in $\psi$ when $\psi \geq \frac{M + M_0}{M - M_0}$. Notice that given $\sigma^2_{01} > \sigma^2_{02}$, $f(0) = \frac{M\sigma^4_{01}}{\omega^2 + M\sigma^2_{02}} < \frac{M\sigma^4_{02}}{\omega^2 + M\sigma^2_{01}} = \lim_{\psi \to \infty} f(\psi)$, so on $[0, \infty)$, $f(\psi)$ reaches its minimum at $\psi = 0$, and reaches its maximum at $\psi = \frac{M + M_0}{M - M_0}$.

- **If $\sigma^2_{01} < \sigma^2_{02}$**, $g(\psi) = (M - M_0) - (M + M_0)\psi$.
  
  - If $0 < M \leq M_0$, $g(\psi) < 0$ for any $\psi$, so $f(\psi)$ is strictly decreasing on $[0, \infty)$.
  
  - If $M > M_0$, $g(\psi) > 0$ if and only if $\psi < \frac{M - M_0}{M + M_0}$. Therefore, $f(\psi)$ increases in $\psi$ when $\psi \leq \frac{M - M_0}{M + M_0}$, and decreases in $\psi$ when $\psi \geq \frac{M - M_0}{M + M_0}$. Notice that given $\sigma^2_{01} < \sigma^2_{02}$, $f(0) = \frac{M\sigma^4_{02}}{\omega^2 + M\sigma^2_{02}} > \frac{M\sigma^4_{01}}{\omega^2 + M\sigma^2_{01}} = \lim_{\psi \to \infty} f(\psi)$, so on $[0, \infty)$, $f(\psi)$ approaches its infimum when $\psi \to \infty$, and reaches its maximum when $\psi = \frac{M - M_0}{M + M_0}$.

- **If $\sigma^2_{01} = \sigma^2_{02}$**, $g(\psi) = M(1 - \psi)$, so $g > 0$ if and only if $0 < \psi < 1$. Therefore, $f(\psi)$ increases in $\psi$ when $\psi \leq 1$, and decreases in $\psi$ when $\psi \geq 1$. Notice that
given \( \sigma_{01}^2 = \sigma_{02}^2 \), \( f(0) = \frac{M \sigma_{02}^2}{\omega^2 + M \sigma_{02}^2} = \frac{M \sigma_{01}^2}{\omega^2 + M \sigma_{01}^2} < f(\psi) \) for any limited \( \psi > 0 \), so on \([0, \infty)\), \( f(\psi) \) reaches its minimum at \( \psi = 0 \), and reaches its maximum at \( \psi = 1 \).

Table 3.A1 summarizes the three cases above.

Now, let us consider the other case with \( n_2 = 0 \). In this case, consumers will keep following both firms, i.e., \((f_1^*, f_2^*) = (1, 1)\). Correspondingly, \( R_2(f_1^*, f_2^*; n_1, n_2) = 0 \) and \( R_1(f_1^*, f_2^*; n_1, n_2) = \min\{n_1, M\} \).

\[
\Sigma_S(n_1, n_2) = \Sigma_1(\min\{n_1, M\}) = \frac{\min\{n_1, M\} \sigma_{01}^4}{\omega^2 + \min\{n_1, M\} \sigma_{01}^2}.
\]

When \( n_1 < M \), \( \Sigma_S(n_1, n_2) \) increases with \( n_1 \). When \( n_1 \geq M \), \( \Sigma_S(n_1, n_2) = \frac{M \sigma_{01}^4}{\omega^2 + M \sigma_{01}^2} \).

To summarize, for any \( n_1 + n_2 \geq M \), when \( n_2 > 0 \), the relationship between \( \Sigma_S(1, 1; n_1, n_2) \) and \( n_1/n_2 \) is summarized in Table 3.A1; when \( n_2 = 0 \), \( \Sigma_S(n_1, n_2) \) increases in \( n_1 \) when \( n_1 \leq M \), and stays constant when \( n_1 \geq M \).

**Step 2: Equilibrium outcomes under five scenarios**

Now we analyze the two firms’ optimal posting intensity considering consumers’ unfollowing decisions and solve the equilibrium. We impose an equilibrium selection criterion that when a firm is indifferent in a range of advertising intensities, the firm will choose the intensity that is optimal if there is a proportion \( \alpha \) of consumers who will never choose to unfollow firms. \( \alpha \) can be any value in \((0, 1)\).

Recall that firm 1’s objective is to minimize \( \Sigma_S(n_1, n_2) \), whereas firm 2’s objective is to maximize \( \Sigma_S(n_1, n_2) \). For any \( 0 < n_2 \leq M \),

\[
\Sigma_S(1, 1; n_1, n_2) = \begin{cases} 
\frac{n_1 \sigma_{01}^4}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^4}{\omega^2 + n_2 \sigma_{02}^2}, & \text{when } n_1 \leq M - n_2, \\
 f(\frac{n_1}{n_2}), & \text{when } n_1 \geq M - n_2,
\end{cases} \tag{3.A13}
\]

\(^{24}\)When \( n_2 = 0 \), consumers will follow both firms and \( \Sigma_S(1, 1; n_1, n_2) = \frac{\min\{n_1, M\} \sigma_{01}^4}{\omega^2 + \min\{n_1, M\} \sigma_{01}^2} \) is minimized at \( n_1 = 0 \), so firm 1 will choose \( n_1 = 0 \), but \( (n_1, n_2) = (0, 0) \) cannot be an equilibrium since firm 2 has incentive to deviate by increasing \( n_2 \). Thus we can safely ignore the case of \( n_2 = 0 \), and thus \( f(\frac{n_1}{n_2}) \) is always well-defined.
Table 3.A1: When $n_1 + n_2 \geq M, n_2 > 0$, the relationship between $f(\psi) = \Sigma S(1, 1; n_1, n_2)$ and $\psi = \frac{n_1}{n_2}$

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Relationship between $f$ and $\psi$</th>
<th>$\arg\min_{\psi} f(\psi)$</th>
<th>$\arg\max_{\psi} f(\psi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. $\sigma^2_{10} &gt; \sigma^2_{20}$, $0 &lt; M \leq M_0$</td>
<td>$f$ increases in $\psi$</td>
<td>$\psi = 0$</td>
<td>$\psi = \infty$</td>
</tr>
<tr>
<td>B. $\sigma^2_{10} &gt; \sigma^2_{20}$, $M &gt; M_0$</td>
<td>$f$ increases in $\psi$ when $0 &lt; \psi \leq \frac{M+M_0}{M-M_0}$ and $f$ decreases in $\psi$ when $\psi &gt; \frac{M+M_0}{M-M_0}$</td>
<td>$\psi = 0$</td>
<td>$\psi = \frac{M+M_0}{M-M_0}$</td>
</tr>
<tr>
<td>C. $\sigma^2_{10} &lt; \sigma^2_{20}$, $0 &lt; M \leq M_0$</td>
<td>$f$ decreases in $\psi$</td>
<td>$\psi = \infty$</td>
<td>$\psi = 0$</td>
</tr>
<tr>
<td>D. $\sigma^2_{10} &lt; \sigma^2_{20}$, $M &gt; M_0$</td>
<td>$f$ increases in $\psi$ when $0 &lt; \psi \leq \frac{M-M_0}{M+M_0}$ and $f$ decreases in $\psi$ when $\psi &gt; \frac{M-M_0}{M+M_0}$</td>
<td>$\psi = \infty$</td>
<td>$\psi = \frac{M-M_0}{M+M_0}$</td>
</tr>
<tr>
<td>E. $\sigma^2_{10} = \sigma^2_{20}$, $M &gt; 0$</td>
<td>$f$ increases in $\psi$ when $0 &lt; \psi \leq 1$, and $f$ decreases in $\psi$ when $\psi &gt; 1$</td>
<td>$\psi = 0$</td>
<td>$\psi = 1$</td>
</tr>
</tbody>
</table>

We denote $M_0 = \left| \frac{(\sigma^2_{01} - \sigma^2_{02})\sigma^2}{\sigma^2_{01}\sigma^2_{02}} \right|$. 
and for any \( n_2 \geq M \),

\[
\Sigma_S(1, 1; n_1, n_2) = f\left(\frac{n_1}{n_2}\right), \forall n_1 \geq 0.
\]

We discuss the equilibrium in each of the five scenarios classified in Table 3.A1.

- Scenario A & B. \( \sigma^2_{01} > \sigma^2_{02} \).

When \( \sigma^2_{01} > \sigma^2_{02} \), consumers may only unfollow firm 2.

We first think about firm 1’s optimal posting intensity given any \( n_2 > 0 \). In Scenario A, \( \Sigma_S(1, 1; n_1, n_2) \) strictly increases in \( n_1 \) since \( \frac{n_1 \sigma^2_{01}}{\omega^2 + n_1 \sigma^2_{01}} + \frac{n_2 \sigma^2_{02}}{\omega^2 + n_2 \sigma^2_{02}} \) increases in \( n_1 \) and \( f\left(\frac{n_1}{n_2}\right) \) increases in \( \frac{n_1}{n_2} \). In Scenario B, given \( n_2 > 0 \), \( \frac{n_1 \sigma^2_{01}}{\omega^2 + n_1 \sigma^2_{01}} + \frac{n_2 \sigma^2_{02}}{\omega^2 + n_2 \sigma^2_{02}} \) is minimized at \( n_1 = 0 \), and for any \( n_1 > 0 \) we have \( f\left(\frac{n_1}{n_2}\right) > f(0) = \frac{M \sigma^2_{01}}{\omega^2 + M \sigma^2_{02}} \geq \frac{\min(n_2, M) \sigma^2_{01}}{\omega^2 + \min(n_2, M) \sigma^2_{02}} = \Sigma_S(1, 1; 0, n_2) \). Therefore, in both Scenario A and B, \( \Sigma_S(1, 1; n_1, n_2) \) is minimized at \( n_1 = 0 \). Then firm 1’s optimal choice given firm 2’s posting intensity \( n_2 > 0 \) must be \( n_1(n_2) = 0 \), because for any \( n_1 > 0 \), whatever consumers’ optimal unfollowing decision is, \( \Sigma_S(n_1, n_2) = \max\{\Sigma_S(1, 1; n_1, n_2), \Sigma_S(0, 1; n_1, n_2), \Sigma_S(1, 0; n_1, n_2)\} \geq \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2) \).

So far we have shown that for any \( n_2 \), firm 1’s optimal choice is \( n_1(n_2) = 0 \). Given \( n_1 = 0 \), consumers will follow both firms and \( \Sigma_S(n_1, n_2) = \Sigma_S(0, n_2) = \frac{\min(n_2, M) \sigma^2_{01}}{\omega^2 + \min(n_2, M) \sigma^2_{02}} \). To maximize \( \Sigma_S(n_1, n_2) \), firm 2 can choose any \( n_2 \in [M, \bar{n}_2] \). Firm 2 is still indifferent in any \( n_2 \in [M, \bar{n}_2] \) when applying the equilibrium selection rule, as they will lead to the same \( \Sigma_S(1, 1; 0, n_2) = \frac{\min(n_2, M) \sigma^2_{01}}{\omega^2 + \min(n_2, M) \sigma^2_{02}} = \frac{M \sigma^2_{01}}{\omega^2 + M \sigma^2_{02}} \).

- Scenario C. \( \sigma^2_{01} < \sigma^2_{02} \) and \( 0 < M \leq M_0 \).

When \( \sigma^2_{01} < \sigma^2_{02} \), consumers may only unfollow firm 1.

We first think about firm 2’s optimal posting intensity given \( n_1 \). We know from Theorem 1 that in Scenario C, consumers will unfollow firm 1 for any \( n_2 \geq M \). Then \( \Sigma_S(n_1, n_2) = \Sigma_S(0, 1; n_1, n_2) = \Sigma_S(1, 1; 0, n_2) = \frac{M \sigma^2_{01}}{\omega^2 + M \sigma^2_{02}} \) for any \( n_2 \geq M \).

If firm 2 chooses \( n_2 \in (0, M) \) and it is in the region that consumers will unfollow firm 1, \( \Sigma_S(n_1, n_2) = \Sigma_S(0, 1; n_1, n_2) = \frac{n_2 \sigma^2_{02}}{\omega^2 + n_2 \sigma^2_{02}} < \frac{M \sigma^2_{01}}{\omega^2 + M \sigma^2_{02}} \).
If firm 2 chooses $n_2 \in (0, M)$ and it is in the region that consumers will keep following both firms,

$$\Sigma_S(n_1, n_2) = \Sigma_S(1, 1; n_1, n_2) = \begin{cases} \frac{n_1 \sigma_{01}^4}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^2}{\omega^2 + n_2 \sigma_{02}^2} & \text{when } 0 < n_2 \leq M - n_1, \\ f\left(\frac{n_1}{n_2}\right) & \text{when } n_2 \geq M - n_1. \end{cases}$$

When $0 < n_2 \leq M - n_1$,

$$\Sigma_S(1, 1; n_1, n_2) = \frac{n_1 \sigma_{01}^4}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^2}{\omega^2 + n_2 \sigma_{02}^2} \leq \frac{n_1 \sigma_{01}^4}{\omega^2 + n_1 \sigma_{01}^2} + \frac{(M - n_1) \sigma_{02}^2}{\omega^2 + (M - n_1) \sigma_{02}^2} = f\left(\frac{n_1}{M - n_1}\right) < f(0) = \frac{M \sigma_{02}^4}{\omega^2 + M \sigma_{02}^2}. $$

Notice that $f\left(\frac{n_1}{M - n_1}\right) < f(0)$ comes from the fact that $f(\cdot)$ strictly decreases with $\frac{n_1}{n_2}$ in Scenario C.

When $M - n_1 \leq n_2 \leq M$,

$$\Sigma_S(1, 1; n_1, n_2) = f\left(\frac{n_1}{n_2}\right) < f(0) = \frac{M \sigma_{02}^4}{\omega^2 + M \sigma_{02}^2}. $$

Therefore, any $n_2 \in (0, M)$ is an inferior choice for firm 2 than any $n_2 \in [M, \bar{n}_2]$. Given any $n_1$, firm 2 is indifferent in choosing any $n_2 \in [M, \bar{n}_2]$ in equilibrium.

Given the choice of firm 2, consumers will unfollow firm 1 for any $n_1 \in [0, \bar{n}_1]$, and thus firm 1 is indifferent in any $n_1 \in [0, \bar{n}_1]$.

Now we apply the equilibrium selection rule. For consumers who always keep following both firms, $\Sigma_S(n_1, n_2) = \Sigma_S(1, 1; n_1, n_2) = f\left(\frac{n_1}{n_2}\right)$ for $n_1 \in [0, \bar{n}_1], n_2 \in [M, \bar{n}_2]$. $f\left(\frac{n_1}{n_2}\right)$ strictly decreases in $\frac{n_1}{n_2}$, so firm 1 will choose $n_1^* = \bar{n}_1$ to minimize $f\left(\frac{n_1}{n_2}\right)$ and firm 2 will choose $n_2^* = \bar{n}_2$ to maximize $f\left(\frac{n_1}{n_2}\right)$. That is, the selected equilibrium outcome is $(f_1^*, f_2^*; n_1^*, n_2^*) = (0, 1; \bar{n}_1, \bar{n}_2)$. 

106
• Scenario D. $\sigma_{01}^2 < \sigma_{02}^2$ and $M > M_0$.

We first think about firm 1’s optimal posting intensity given $n_2$. For any $n_2 \in (0, M],$

$$\Sigma_S(1, 1; n_1, n_2) = \begin{cases} \frac{n_1 \sigma_{01}^2}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^2}{\omega^2 + n_2 \sigma_{02}^2}, & \text{when } n_1 \leq M - n_2 \\ f(\frac{n_1}{n_2}), & \text{when } n_1 \geq M - n_2, \end{cases}$$

and for any $n_2 \geq M, \Sigma_S(1, 1; n_1, n_2) = f(\frac{n_1}{n_2}), \forall n_1 \geq 0$.

$$\frac{n_1 \sigma_{01}^2}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^2}{\omega^2 + n_2 \sigma_{02}^2} \text{ increases in } n_1. \text{ From Step 1, we know that in Scenario D, } f(\frac{n_1}{n_2}) \text{ increases in } \frac{n_1}{n_2} \text{ when } \frac{n_1}{n_2} \leq \frac{M - M_0}{M + M_0} \text{ and decreases in } \frac{n_1}{n_2} \text{ when } \frac{n_1}{n_2} \geq \frac{M - M_0}{M + M_0}. \text{ Hence given any } n_2 > 0, \Sigma_S(1, 1; n_1, n_2) \text{ first increases with } n_1 \text{ and then decreases with } n_1.$$

Figure 3-A1 illustrates the shape of $\Sigma_S(1, 1; n_1, n_2)$ as a function of $n_1$ in the cases of $n_2 < M$ and $n_2 \geq M^{25}$.

From Theorem 1 we know that in Scenario D, consumers will unfollow firm 1 when $\Sigma_S(1, 1; n_1, n_2) < \Sigma_S(0, 1; n_1, n_2) = \Sigma_S(1, 1; 0, n_2)$, which is equivalent to

$$n_1 \geq \hat{n}_1(n_2) = \begin{cases} \tilde{n}_1(n_2), & \text{if } n_2 \leq M \\ \frac{M - M_0}{M + M_0} n_2, & \text{if } n_2 \geq M. \end{cases}$$

In fact $\hat{n}_1(n_2)$ is the value of $n_1$ that solves $\Sigma_S(1, 1; n_1, n_2) = \Sigma_S(1, 1; 0, n_2)$. Given any $n_2$, if firm 1 chooses $n_1 \geq \hat{n}_1(n_2)$, consumers unfollow firm 1 and $\Sigma_S(n_1, n_2) = \Sigma_S(0, 1; n_1, n_2) = \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2)$; for any $n_1 \in (0, \hat{n}_1(n_2))$, consumers follow both firms and $\Sigma_S(n_1, n_2) = \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2)$.

Therefore, if $\hat{n}_1(n_2) \leq \tilde{n}_1$, firm 1 may choose any $n_1 \in \{0\} \cup [\hat{n}_1(n_2), \tilde{n}_1]$, though consumers will unfollow firm 1 for $n_1 \in [\hat{n}_1(n_2), \tilde{n}_1]$. Firm 1 will not choose $n_1 \in (0, \hat{n}_1(n_2))$ to retain followers because that will lead to higher value of $\Sigma_S$. Applying the equilibrium selection rule, firm 1 will choose $n_1(n_2) = \tilde{n}_1$ since it minimizes $\Sigma_S(1, 1; n_1, n_2)$ in such case.

---

25When $n_2 < M$ and $n_1 \geq M - n_2$, $\Sigma_S(1, 1; n_1, n_2)$ can be either strictly decreasing (if $\frac{M - n_2}{n_2} \geq \frac{M - M_0}{M + M_0}$) or first increasing and then decreasing (if $\frac{M - n_2}{n_2} < \frac{M - M_0}{M + M_0}$). Figure 3-A1(a) only shows the former case, but the two cases will have the implication on $n_1(n_2)$. 

107
Figure 3-A1: \( \Sigma_S(1, 1; n_1, n_2) \) as a function of \( n_1 \) when \( \sigma_{01}^2 < \sigma_{02}^2, M > M_0 \)

If \( \hat{n}_1(n_2) > \bar{n}_1, \Sigma_S(n_1, n_2) = \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2) \) for any \( n_1 \in (0, \bar{n}_1] \) and consumers will follow both firms, so firm 1 should choose \( n_1(n_2) = 0 \) to minimize \( \Sigma_S(n_1, n_2) \).

In Figure 3-A1(a) \( \hat{n}_1(n_2) > \bar{n}_1 \) and in Figure 3-A1(b) \( \hat{n}_1(n_2) < \bar{n}_1 \). In fact \( \hat{n}_1(n_2) \) can be smaller or larger than \( \bar{n}_1 \) in both cases of \( n_2 < M \) and \( n_2 \geq M \).

So far we have shown that firm 1’s optimal choice is either \( n_1(n_2) = 0 \) or \( n_1(n_2) = \bar{n}_1 \).

If \( n_1 = \bar{n}_1 \) in equilibrium, we first argue that rational consumers must choose to unfollow firm 1 in such equilibrium: otherwise firm 1 will have incentive to deviate to \( n_1 = 0 \) since \( \Sigma_S(1, 1; \bar{n}_1, n_2) > \Sigma_S(0, 1; \bar{n}_1, n_2) = \Sigma_S(1, 1; 0, n_2) \). The second argument is that in such equilibrium, firm 2 must choose \( n_2 \geq M \): given that consumers must unfollow firm 1 in such equilibrium, \( \Sigma_S(\bar{n}_1, n_2) = \Sigma_S(0, 1; \bar{n}_1, n_2) = \frac{\min\{n_2, M\} \sigma_{02}^2}{\omega^2 + \min\{n_2, M\} \sigma_{02}^2} \).

If \( n_2 < M \), firm 2 will deviate by increasing \( n_2 \) to increase \( \Sigma_S(\bar{n}_1, n_2) \). Therefore, \( (0, 1; \bar{n}_1, n_2) \) with \( n_2 < M \) cannot be an equilibrium. For \( (0, 1; \bar{n}_1, n_2) \) with \( n_2 \geq M \) to be an equilibrium, we must have

\[
\hat{n}_1(n_2) = \frac{M - M_0}{M_0} n_2 \leq \bar{n}_1
\] (3.A14)
according to our previous analysis. Furthermore, firm 2 cannot increase $n_2$ to some $n'_2$ such that $\hat{n}_1(n'_2) > \bar{n}_1$ (when $n_2 \geq M$, $\hat{n}_1(n_2) = \frac{M-M_0}{M_0} n_2$ is an increasing function of $n_2$), which makes consumers keep following both firms and $\Sigma_S(\bar{n}_1, n'_2) = \Sigma_S(1, 1; \bar{n}_1, n'_2) > \Sigma_S(0, 1; \bar{n}_1, n'_2) = \Sigma_S(0, 1; \bar{n}_1, n_2) = \Sigma_S(\bar{n}_1, n_2)$. Hence for $(0, 1; \bar{n}_1, n_2)$ with $n_2 \geq M$ to be an equilibrium, we also need to have

$$\hat{n}_1(\bar{n}_2) = \frac{M-M_0}{M_0} \bar{n}_2 \leq \bar{n}_1.$$  
(3.15)

Condition (3.15) guarantees that condition (3.14) holds for any $n_2 \in [M, \bar{n}_2]$. Therefore, (3.15) is a sufficient and necessary condition for $(0, 1; \bar{n}_1, n_2)$ with any $n_2 \in [M, \bar{n}_2]$ to be an equilibrium. Applying our equilibrium selection rule, firm 2 should choose $n_2 = \bar{n}_2$ to maximize $\Sigma_S(1, 1; \bar{n}_1, n_2)$ ($\Sigma_S(1, 1; n_1, n_2)$ decreases in $\frac{n_1}{n_2}$ when $\frac{n_1}{n_2} \geq \frac{M-M_0}{M_0}$).

If $n_1 = 0$, consumers will follow both firms and firm 2 will be indifferent in choosing any $n_2 \in [M, \bar{n}_2]$. For any $n_2 \in [M, \bar{n}_2]$, $(0, n_2)$ is indeed an equilibrium outcome when $\hat{n}_1(n_2) = \frac{M-M_0}{M_0} n_2 > \bar{n}_1$, i.e., when $n_2 > \frac{M_0}{M-M_0} \bar{n}_1$. When $\bar{n}_2 > \frac{M_0}{M-M_0} \bar{n}_1$, such $n_2$ exists and $(1, 1; 0, n_2)$ is an equilibrium for $n_2 \in \{\max\{\frac{M_0}{M-M_0} \bar{n}_1, M\}, \bar{n}_2\}$.

To summarize, when $\bar{n}_1 \geq (\frac{M}{M_0} - 1) \bar{n}_2$, the equilibrium outcome will be $(0, 1; \bar{n}_1, \bar{n}_2)$; when $\bar{n}_1 < (\frac{M}{M_0} - 1) \bar{n}_2$, the equilibrium outcome will be $(1, 1; 0, n_2^*)$ where $n_2^* \in (\max\{\frac{M_0}{M-M_0} \bar{n}_1, M\}, \bar{n}_2)$.

- Scenario E. $\sigma_{01}^2 = \sigma_{02}^2, M > 0$.

In this case, consumers will never unfollow either firm. That is, we always have $\Sigma_S(n_1, n_2) = \Sigma_S(1, 1; n_1, n_2)$ in this scenario.

We first think about firm 1’s optimal posting intensity. In equation (3.13), given any $n_2 > 0$, $\frac{n_1 \sigma_{01}^2}{\omega^2 + n_1 \sigma_{01}^2} + \frac{n_2 \sigma_{02}^2}{\omega^2 + n_2 \sigma_{02}^2}$ is minimized at $n_1 = 0$, and according to Table 3.A1, $f'(\frac{n_1}{n_2}) > f(0) = \frac{M \sigma_{01}^2}{\omega^2 + M \sigma_{01}^2} \geq \frac{\min\{n_2, M\} \sigma_{02}^2}{\omega^2 + \min\{n_2, M\} \sigma_{02}^2} = \Sigma_S(1, 1; 0, n_2)$ for any positive $n_1$. Therefore, firm 1’s optimal choice should be $n_1(n_2) = 0$ for any $n_2$ as it minimizes $\Sigma_S(1, 1; n_1, n_2)$.

Given that firm 1 chooses $n_1(n_2) = 0$ for any $n_2$, $\Sigma_S(n_1, n_2) = \Sigma_S(0, n_2) =$
\[
\min\{n_2, M\} \sigma_{02}^{\rho_2}. \quad \text{To maximize } \Sigma_S(0, n_2), \text{ firm 2 is indifferent in } n_2 \in [M, \bar{n}_2].
\]

Therefore the equilibrium outcome is \((1, 1; 0, n_2^*)\) where \(n_2^* \in [M, \bar{n}_2].\)

Summarizing the five cases, when \(\sigma_{01}^2 < \sigma_{02}^2\) and \(M \leq (1 + \frac{n}{n_2})M_0\), the equilibrium outcome is \((f_1^*, f_2^*; n_1^*, n_2^*) = (0, 1; \bar{n}_1, \bar{n}_2)\); when \(\sigma_{01}^2 < \sigma_{02}^2\) and \(M > (1 + \frac{n}{n_2})M_0\), the equilibria are \((1, 1; 0, n_2^*)\) for any \(n_2^* \in \{\frac{M_0}{M-M_0}\bar{n}_1, M\}, \bar{n}_2\); when \(\sigma_{01}^2 \geq \sigma_{02}^2\), the equilibria are \((1, 1; 0, n_2^*)\) for any \(n_2^* \in [M, \bar{n}_2]\).

### 3.A.4 Proof of Proposition 5

We first want to prove that given any \(\omega^2, n_2\), firm 1 will choose \(n_1(n_2) = 0\) or \(n_1(n_2) = \bar{n}_1\), and is indifferent in choosing any \(\omega^2 \in [\omega, \infty)\). In the former case, consumers will keep following both firms, and in the latter case, consumers will unfollow firm 1.

In order to prove this, we first show that for any \(\omega_1^2, \omega_2^2 > 0\) and \(n_2 > 0\), \(\Sigma_S(1, 1; n_1, n_2)\) is either an increasing function of \(n_1\) or an inverse-U shape function of \(n_1\), i.e., first increasing with \(n_1\) and then decreasing with \(n_1\).

Given \(n_2 > 0\),

\[
\Sigma_S(1, 1; n_1, n_2) = \begin{cases} \frac{n_1\sigma_{01}^2}{\omega_1^2 + n_1\sigma_{01}^2} + \frac{n_2\sigma_{02}^2}{\omega_2^2 + n_2\sigma_{02}^2}, & \text{if } n_1 \leq M - n_2, \\ \frac{n_1\sigma_{01}^2}{\omega_1^2 + n_1\sigma_{01}^2} + \frac{n_2\sigma_{02}^2}{\omega_2^2 + n_2\sigma_{02}^2}, & \text{if } n_1 \geq M - n_2. \end{cases} \tag{3.16}
\]

In (3.16), \(\frac{n_1\sigma_{01}^2}{\omega_1^2 + n_1\sigma_{01}^2} + \frac{n_2\sigma_{02}^2}{\omega_2^2 + n_2\sigma_{02}^2}\) increases with \(n_1\). Denote \(\psi = \frac{n_1}{n_2}\) and \(h(\psi) = \frac{\psi\sigma_{01}^2}{\omega_1^2 + \psi\sigma_{01}^2} + \frac{\psi\sigma_{02}^2}{\omega_2^2 + \psi\sigma_{02}^2}\). Its first order derivative is

\[
h'(\psi) = \frac{M}{(M\psi\sigma_{01}^2 + \omega_1^2 + \psi\omega_2^2)(M\psi\sigma_{02}^2 + \omega_2^2 + \psi\omega_2^2)^2}. \quad [2M(1 + \psi)\sigma_{01}^2\sigma_{02}^2(\sigma_{01}^2 - \psi\sigma_{02}^2)\omega_1^2\omega_2^2 + M^2\sigma_{01}^4\sigma_{02}^4(\omega_1^2 - \psi^2\omega_2^2)]
\]

If \(\sigma_{01}^2 \leq \frac{\sigma_{02}^2(\omega_1^2 + \omega_2^2)}{M\sigma_{02}^2 + \omega_2^2}\), \(h'(\psi)\) is negative for any \(\psi > 0\).

\[\text{26} \quad \text{We have argued that } n_2 = 0 \text{ will never happen in equilibrium as firm 1 will choose } n_1 = 0 \text{ and firm 2 has incentive to deviate by increasing } n_2.\]
If \( \omega_1^2 \leq \frac{M}{\omega_4^2} \sigma_{02}^2 \) and \( \sigma_{01}^2 > \frac{\sigma_{02}^2 \omega_1^2}{2 \sigma_{02}^2 + \omega_2^2} \), or \( \omega_1^2 > \frac{M}{\omega_4^2} \sigma_{02}^2 \) and \( \frac{\omega_1^2 \sigma_{02}^2 - \omega_2^2}{\omega_2} < \sigma_{01}^2 \), then

\[
\psi < \frac{M \sigma_{01}^2 \sigma_{02}^2 \omega_1^2 (\sigma_{02}^2 - \sigma_{01}^2) + \omega_1^2 (\sigma_{01}^2 \omega_2^2 - \sigma_{02}^4 \omega_1^2)}{\sigma_{02}^2 (M \sigma_{02}^2 + \omega_1^2)^2 - \sigma_{01}^2 \omega_2^2} + \frac{M \sigma_{01}^2 \sigma_{02}^2 \omega_1^2 (M \sigma_{02}^2 \sigma_{02}^2 + \sigma_{02}^4 \omega_1^2 + \sigma_{01}^2 \omega_2^2)}{\omega_2 [M \sigma_{02}^2 (M \sigma_{01}^2 + \omega_1^2)^2 - \sigma_{01}^2 \omega_2^2]}
\]

and is negative when \( \psi \) is greater than this threshold.

If \( \omega_1^2 > \frac{M}{\omega_4^2} \sigma_{02}^2 \) and \( \sigma_{01}^2 > \frac{\sigma_{02}^2 \omega_1^2}{M \sigma_{02}^2 - \omega_2^2} \), then

\( h'(\psi) \) is positive for any \( \psi > 0 \).

Since \( \Sigma_S(1, 1; n_1, n_2) \) increases in \( n_1 \) when \( n_1 \leq M - n_2 \) and it is continuous at \( n_1 = M - n_2 \), in the case of \( h'(\psi) > 0 \) for any \( \psi > 0 \), \( \Sigma_S(1, 1; n_1, n_2) \) also strictly increases in \( n_1 \). Similarly, in the case of \( h(\psi) \) first increases in \( \psi \) and then decreases in \( \psi \), \( \Sigma_S(1, 1; n_1, n_2) \) also first increases in \( n_1 \) and then decreases in \( n_1 \). In the case of \( h(\psi) \) strictly decreases in \( \psi \) for any \( \psi > 0 \), \( \Sigma_S(1, 1; n_1, n_2) \) first increases in \( n_1 \) when \( n_1 \leq M - n_2 \) and then decreases in \( n_1 \) when \( n_1 \geq M - n_2 \). Thus we have shown that given \( \omega_1^2, \omega_2^2 > 0, n_2 > 0 \), \( \Sigma_S(1, 1; n_1, n_2) \) either strictly increases with \( n_1 \) or first increases and then decreases with \( n_1 \).

If \( \Sigma_S(1, 1; n_1, n_2) \) strictly increases in \( n_1 \), or \( \Sigma_S(1, 1; n_1, n_2) \) first increases in \( n_1 \) and then decreases in \( n_1 \) but \( \Sigma_S(1, 1; \bar{n}_1, n_2) > \Sigma_S(1, 1; 0, n_2) \), firm 1 will choose \( n_1(n_2) = 0 \). This is because for any \( n_1 \in (0, \bar{n}_1] \), \( \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) \), and thus \( \Sigma_S(n_1, n_2) \geq \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2) \).

If \( \Sigma_S(1, 1; n_1, n_2) \) strictly decreases in \( n_1 \), or \( \Sigma_S(1, 1; n_1, n_2) \) first increases in \( n_1 \) and then decreases in \( n_1 \) with \( \Sigma_S(1, 1; \bar{n}_1, n_2) \leq \Sigma_S(1, 1; 0, n_2) \), consumers will unfollow firm 1 when \( n_1 \geq \hat{n}_1(n_2) \) where \( \hat{n}_1(n_2) \) solves \( \Sigma_S(1, 1; n_1, n_2) = \Sigma_S(1, 1; 0, n_2) \) \((\hat{n}_1(n_2) = 0 \text{ if } \Sigma_S(1, 1; n_1, n_2) \text{ strictly decreases in } n_1)\). Firm 1 will be indifferent in any \( n_1 \in \{0\} \cup [\hat{n}_1(n_2), \bar{n}_1] \), as they all lead to \( \Sigma_S(n_1, n_2) = \Sigma_S(0, n_2) \). Firm 1 will not choose \( n_1 \in (0, \hat{n}_1(n_2)) \), because they lead to

\[
\Sigma_S(n_1, n_2) = \max \{ \Sigma_S(1, 1; n_1, n_2), \Sigma_S(1, 0; n_1, n_2), \Sigma_S(0, 1; n_1, n_2) \} \\
\geq \Sigma_S(1, 1; n_1, n_2) > \Sigma_S(1, 1; 0, n_2) = \Sigma_S(0, n_2).
\]
By applying our equilibrium selection rule, firm 1 will choose \( n_1(n_2) = \tilde{n}_1 \), since it minimizes \( \Sigma_S(1, 1; n_1, n_2) \).

So far we have shown that given any \( \omega_2^0, n_2 \), firm 1 will choose either \( n_1(n_2) = 0 \) or \( n_1(n_2) = \tilde{n}_1 \).

When \( n_1(n_2) = 0 \), consumers will keep following both firms and then \( \Sigma_S(n_1, n_2) = \Sigma_S(1, 1; 0, n_2) = \frac{\min\{n_2, M\} \sigma_{y_2}}{\omega_2^0 + \min\{n_2, M\} \sigma_{y_2}} \). Its value does not depend on \( \omega_2^0 \), so firm 1 may choose any \( \omega_1^{*2} \in [\omega_2^0, \infty) \). To maximize \( \Sigma_S(n_1, n_2) \), firm 2 will choose \( \omega_2^{*2} = \omega_2^2 \) and any \( n_2^{*} \in [M, \tilde{n}_2] \) in equilibrium.

When \( n_1(n_2) = \tilde{n}_1 \), consumers must choose to unfollow firm 1: otherwise firm 1 would rather choose \( n_1(n_2) = 0 \). Then \( \Sigma_S(n_1, n_2) = \Sigma_S(0, 1; n_1, n_2) = \frac{\min\{n_2, M\} \sigma_{y_2}}{\omega_2^0 + \min\{n_2, M\} \sigma_{y_2}} \). Its value also does not depend on \( \omega_1^0 \), so firm 1 may choose any \( \omega_1^{*2} \in [\omega_2^0, \infty) \). To maximize \( \Sigma_S(n_1, n_2) \), firm 2 will choose \( \omega_2^{*2} = \omega_2^2 \), and \( n_2^{*} \) must be greater than or equal to \( M \).

Both cases will lead to the same value of \( \Sigma_S(n_1^{*}, n_2^{*}; \omega_1^{*2}, \omega_2^{*2}) = \frac{M \sigma_{y_2}}{\omega_2^0 + M \sigma_{y_2}} \). So \((f_1^{*}, f_2^{*}; n_1^{*}, n_2^{*}; \omega_1^{*2}, \omega_2^{*2}) = (1, 1; 0, n_2^{*}; \omega_1^{*2}, \omega_2^2)\) with any \( \omega_1^{*2} \in [\omega_2^0, \infty) \) and \( n_2^{*} \in [M, \tilde{n}_2] \) is always an equilibrium, and under certain condition, \((f_1^{*}, f_2^{*}; n_1^{*}, n_2^{*}; \omega_1^{*2}, \omega_2^{*2}) = (0, 1; \tilde{n}_1, n_2^{*}; \omega_1^{*2}, \omega_2^2)\) with any \( \omega_1^{*2} \in [\omega_2^0, \infty) \) and \( n_2^{*} \in [M, \tilde{n}_2] \) is also an equilibrium.

### 3.A.5 Online Survey

We recruit 386 Weibo users and they complete our survey online. The questions of the survey are translated as below. The percentage number after each choice is the proportion of subjects who chose this alternative out of all the subjects who answered this question. The number of subjects who answer each question is 386 if not specified.

1. Please estimate your frequency of browsing Weibo.
   - Less than once a day (12.95%)
   - 1~3 times per day (51.55%)
   - 4~6 times per day (18.65%)
   - More than 6 times per day (16.84%)
2. Can you read all the new tweets each time you browse Weibo?

- Never (8.29%)
- Usually not (25.65%)
- Sometimes yes (32.12%)
- Usually yes (27.72%)
- Always yes (6.22%)

3. Please estimate the number of verified accounts you are following.

- 0 (1.04%)
- 1~10 (25.39%)
- 10~20 (37.82%)
- 20~30 (15.8%)
- >30 (19.95%)

4. Do you agree or disagree “Some verified accounts post too intensively, preventing me from reading tweets from other accounts.”

- Strongly disagree (2.62%)
- Disagree (21.2%)
- Neutral (29.32%)
- Agree (43.19%)
- Strongly agree (3.66%)

5. Have you ever unfollowed verified accounts that you follow?

- Yes (74.08%)
- No (25.92%)
6. Do you agree or disagree “The reasons I unfollow the verified accounts include that they post too intensively, preventing me from reading tweets from other accounts.” (Only those who chose “Yes” in the Question 5 answered this question. The number of answers is 283.)

- Strongly disagree (0.35%)
- Disagree (20.85%)
- Neutral (20.49%)
- Agree (53.71%)
- Strongly agree (4.59%)

7. Have you ever followed TV show accounts?

- Yes (75.65%)
- No (24.35%)

8. Do you agree or disagree “The tweets posted by those TV show accounts contain information that is relevant to the content of the show. (Only those who chose “Yes” in Question 7 answered this question. The number of answers is 292.)

- Strongly disagree (0.34%)
- Disagree (2.05%)
- Neutral (5.48%)
- Agree (69.52%)
- Strongly agree (22.6%)

9. Do you agree or disagree “The tweets posted by those TV show accounts can help me know the content of the show. (Only those who chose “Yes” in Question 7 answered this question. The number of answers is 292.)

- Strongly disagree (0%)
- Disagree (1.37%)
- Neutral (8.22%)
- Agree (61.99%)
- Strongly agree (28.42%)

Table 3.A2 summarizes the time of taking the survey and the answers to those 5-level Likert scale questions.

Table 3.A2: Summary Statistics of the 5-level Questions and Time of Taking the Survey

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of Taking Survey</td>
<td>263.37</td>
<td>2639.15</td>
<td>16</td>
<td>51644</td>
<td>386</td>
</tr>
<tr>
<td>Question 2</td>
<td>2.98</td>
<td>1.06</td>
<td>1</td>
<td>5</td>
<td>386</td>
</tr>
<tr>
<td>Question 4</td>
<td>3.24</td>
<td>0.92</td>
<td>1</td>
<td>5</td>
<td>382</td>
</tr>
<tr>
<td>Question 6</td>
<td>3.41</td>
<td>0.88</td>
<td>1</td>
<td>5</td>
<td>283</td>
</tr>
<tr>
<td>Question 8</td>
<td>4.12</td>
<td>0.62</td>
<td>1</td>
<td>5</td>
<td>292</td>
</tr>
<tr>
<td>Question 9</td>
<td>4.17</td>
<td>0.63</td>
<td>2</td>
<td>5</td>
<td>292</td>
</tr>
</tbody>
</table>
Chapter 4

Cooperative Search Advertising

4.1 Introduction

Search advertising is growing rapidly and has become a major advertising channel. In 2015, compared to a 6.0% growth of the entire advertising industry, search advertising grew fast at 16.2%, and has reached a global expenditure of $80 billion dollars\(^1\). Retailing is a major contributor to search advertising. The No.1 spender on Google Adwords is Amazon, and five of the top ten industries contributing most to Google Adwords are related to retailing, which together contribute more than one quarter of Google’s revenue.\(^2\) In all these industries, both brand-owners (manufacturers) and retailers can advertise on the same keywords. For example, Figure 4-1 shows the advertisements on Google in one search query of “laptop”.\(^3\) In this example, some manufacturers advertise products via their own e-commerce sites (Microsoft Surface, Samsung and Google Chromebook), some advertise products via retailers (Asus, Apple, Lenovo, Dell, and Toshiba), and one brand (HP) does both.

What lies behind the scene is the channel coordination between the manufacturers and retailers on search advertising. On one hand, manufacturers and their retailers

\(^1\) Data source: eMarketer.com.

\(^2\) These industries include retailing and general merchandise, home and garden, computer and consumer electronics, vehicles, as well as business and industrial. This is according to a breakdown of Google’s 2011 revenue provided by WordStream.com.

\(^3\) Google provides two kinds of search ads: AdWords ads in texts and product listing ads in pictures. We do not explicitly distinguish them in our work.
compete directly with each other on search engine platforms; on the other hand, it is common for a manufacturer to coordinate with its retailers on search advertising spending. Specifically, a manufacturer can set up a “cooperative advertising” (co-op) fund and reimburse retailers when they advertise the manufacturer’s products on search engines. According to one of the authors’ work experience at Walmart.com, many major brands provide co-op funds to Walmart.com for search ads. According to a survey from Borrell Associates in 2015, 70% of brand managers say that they offer digital marketing co-op programs. From 2012 to 2015, the percentage of local advertisers involved in digital marketing co-op programs increased from 25% to 61%. Among all digital advertising categories, search advertising ranked No. 1 for the highest impact in the minds of both brand managers (68%) and local advertisers (62%). 83% of brand managers and 85% of local advertisers agree that the use of search and display advertising is important in supporting the brand. While 29% brand managers said that they are now actually supporting paid search advertising in their co-op programs, the report by Borrell Associates points out that, “search advertising needs to be a far bigger part of the digital side of co-op.”  

Cooperative advertising is not new, and it prevails in traditional media (Berger 1972, Bergen and John 1997). For example, manufacturers of fast-moving consumer goods can promote their products and get them better displayed on shelves by subsidizing supermarkets; phone manufacturers can get their products featured in the mobile carriers’ TV commercials by cooperating with them on ads spending. The annual total spending on cooperative advertising is estimated to exceed $50 billion in the US alone.  

However, distinct market and institutional features of the search advertising industry make cooperative search advertising very different from cooperative advertising in traditional media.

A prominent feature of search advertising is that the market is highly concentrated, and thus advertisers face stronger competition. In the US, the market is dominated by three players—Google, Bing, and Yahoo, with Google taking more

---

4https://www.netsertive.com/resources/borrell-research-report-changing-face-co-op-programs
5Co-op Advertising Programs Sourcebook, by National Register Publishing Company.
than 60% of the market share. For any business trying to reach customers via search ads, it is almost imperative to use Google’s service. As a result, a manufacturer will inevitably compete with its own retailers directly for consumers’ clicks on Google. In contrast, in traditional media, a manufacturer and its retailers face much weaker competition on advertisements with each other, as their ads can appear in various media outlets. Furthermore, in traditional advertising, advertisements from manufacturers and retailers often serve different purposes and have positive spillovers (Bergen and John 1997). For example, a manufacturer can make national ads to increase brand awareness, while its retailers can make local ads about the product availability, price, promotion, etc. A consumer who has watched a TV ad from the manufacturer may choose to buy the product from a retailer, and a consumer may see ads from retailer A but make the purchase from retailer B. In contrast, in paid search advertising, both a manufacturer and its retailers’ ads are mainly made to drive conversions. Since their differentiation becomes much smaller, the competition becomes stronger.

Another prominent feature is that search engine platforms sell keywords via generalized second-price position auctions, and this becomes crucial for cooperative advertising due to the high market concentration of the search advertising industry. It is important for advertisers to understand the auction mechanism, when formulating their competitive strategies on search advertising. In the position auctions, a manufacturer and its retailers compete directly in bidding for a better position. One advertiser’s higher bid will either increase other advertisers’ costs or decrease their demands by moving them to less attractive positions. As a result, higher advertising cost does not necessarily lead to larger demand, but can be burning the manufacturer’s own money. This is in stark contrast with the case in traditional advertising, where ads from a manufacturer and its retailers usually complement with each other.

These new features of cooperative search advertising have been noticed by brand managers, and potentially prevents them from operating cooperative search advertising effectively. The survey by Borell indicated that brand managers and local advertisers are not participating more heavily in co-op search advertising due to the complexity of digital co-op programs and lack of understanding of the advertising
landscape, leaving more advertising dollars—about $14 billion—unused. A report regarding digital cooperative advertising by Interactive Advertising Bureau (IAB) also pointed out that the key barriers to online co-op advertising are the complexity of digital channels, lack of knowledge required to advertise in digital channels, lack of guidelines and requirements. They quoted from a senior manager of search engine and mobile marketing at HP that “We’re driving each other’s bidding up. HP’s perspective is we don’t think co-op is that positive if we’re all going after the same term.”

In this chapter, we build a game-theoretic model to investigate how a manufacturer and its retailers should coordinate in search advertising, and prescribe the manufacturer’s optimal cooperative search advertising strategy. Specifically, we aim to answer the following research questions. (1) Should a manufacturer cooperate with all its retailers? Is it indeed burning its own money by invoking competition on the search advertising platform? (2) If a manufacturer should not cooperate with all retailers, which one(s) should it cooperate with? (3) Given the profit margin via direct sales is often higher than that via retailers, should a manufacturer advertise directly to consumers instead of via retailers? (4) How does retailers’ price competition influence the manufacturer’s cooperative advertising strategy? (5) How to coordinate the channel by using both wholesale and cooperative advertising contracts?

Throughout the chapter, we consider the setting where one manufacturer sells via two retailers. This is the simplest setting under which we can study a manufacturer’s choice of how many and which retailer(s) to cooperate with on search advertising. We build our basic model based on the assumption of exogenous wholesale contracts and retail prices. This is a reasonable assumption when search advertising is only part of the demand source, and thus it is reasonable to assume that prices and wholesale contracts have been determined before the manufacturer and retailers make decisions on search advertising. We consider a simple coordination mechanism where a manufacturer covers a fixed percentage—the so-called participation rate—of a retailer’s spending on search advertising. This coordination mechanism has been widely ac-

---

cepted in industry as well as in previous literature (Dutta et al. 1995, Bergen and John 1997, Nagler 2006). The game proceeds as follows. First, a manufacturer determines the participation rate for each retailer. Then retailers and other advertisers submit their bids to a search engine platform. Finally, the auction outcome realizes.

Following the literature of position auctions (Varian 2007, Edelman et al. 2007), we do not explicitly model consumers’ searching and clicking behaviors, but assume an exogenous click-through rate for each position, which is independent of advertisers’ identities (an extension that allows the click-through rates to depend on advertiser identities is investigated in Section 4.1). We focus on the intra-brand coordination and competition among a manufacturer and its retailers, but also account for inter-brand competition by incorporating outside advertisers in the position auction.

The basic model (Section 4.2) considers the case in which the manufacturer cannot directly participate in search advertising. This happens, for example, when the manufacturer does not have its own e-commerce site that sells to consumers directly. This enables us to focus on the manufacturer’s problem of how many retailers to sponsor, which retailer(s) to sponsor, and how much sponsorship to provide. In determining how many retailers to cooperate with, a manufacturer makes a tradeoff between larger demand brought by more retailers and higher bidding cost associated with intensified competition in bidding. We find that it may not be optimal for a manufacturer to cooperate with all the retailers, even if they are ex ante the same. This finding points to an important difference between cooperative advertising in traditional media and in search advertising. In both cases, it has been observed that the utilization rate of manufacturers’ funding is low. In traditional cooperative advertising, the reason could be that not all retailers satisfy the qualifications to join the co-op program, and the manufacturer selects retailers based on their characteristics when giving out cooperative advertising money. In cooperative search advertising, our result shows that even if two retailers are exactly the same, a manufacturer may still decide to sponsor only one of them. Thus there is an additional reason for low utilization of the co-op funding—the manufacturer takes hold of the cooperative advertising money to reduce potential competition among retailers.
With two asymmetric retailers, the manufacturer should support the retailer with higher channel profit per click to get a higher position than the other retailer. That is, when deciding the two retailers’ relative positions, the manufacturer acts as if the channel is fully coordinated, even though it is actually not. This demonstrates the effectiveness of this simple coordination mechanism and thus provides a rationale for its prevalence in industry. We discuss the optimality of the participation rate mechanism by considering the integrated channel as a benchmark, and find that channel integration can lead to lower channel profit than the participation rate mechanism, because the outside advertiser may strategically raise the bid knowing that the integrated channel has higher willingness-to-pay per click.

In Section 4.3, we study a more general setup in which the manufacturer can endogenously decide whether to participate in search advertising directly by bidding for its own website. The key finding in the basic model still holds, i.e., the retailer with the higher channel profit per click still gets a higher position than the other retailer in equilibrium. We also find that when the manufacturer’s profit per click via direct sales is higher than a retailer’s channel profit per click, the manufacturer should get a higher position than the retailer. The manufacturer will definitely participate in the auction directly, when his profit per click via direct sales is higher than both retailers’ channel profit per click; otherwise, he may or may not participate directly. Our result explains why we commonly observe retailers’ search ads in practice, given that manufacturers earn higher profits by advertising and selling directly to consumers.

There are two reasons. First, a retailer may have a higher conversion rate, which means a higher return for each click. Second, selling through a retailer may result in a higher channel profit margin, since retailers may operate more efficiently and have lower operational cost. When deciding whether to advertise directly to consumers or to advertise via retailers, the manufacturer should rely on the comparison of channel profits instead of its own profits. A retailer with high profit margin has a strong incentive to bid, and therefore does not require much cooperative sponsorship from the manufacturer to win the search ad auction; on the other hand, if the retailer is not sponsored, it will become a strong competitor to the manufacturer, which makes...
the manufacturer's winning more costly.

We further consider three extensions to our basic model in Section 4.4. In Section 4.1, we extend our basic model by allowing the click-through rate to depend on the advertiser’s identity. Our main results generalize nicely. With two retailers, the manufacturer will support the retailer with higher channel profit per impression to get a higher position than the other retailer.

In Section 4.2, we incorporate endogenous retail price with price competition. Consistent with the findings from the basic model, a manufacturer may optimally provide sponsorship to only one retailer, given two ex ante symmetric retailers. Particularly, by sponsoring both retailers, the manufacturer encourages retail price competition, which lowers retail prices and increases demand. However, a lower retail price and thus a lower retail margin reduce retailers’ incentives to bid high and win the position auctions. Consequently, the manufacturer has to provide more sponsorship to both retailers to get them displayed in the position auctions. Generally speaking, the manufacturer tradeoffs demand expansion and sponsorship cost, and chooses to sponsor only one retailer when the consumer heterogeneity between the two retailers is relatively small.

In Section 4.3, we endogenize both retail prices and wholesale contracts, without explicitly considering retail price competition. We illustrates how a manufacturer can uses the two devices—wholesale and cooperative advertising contracts—to coordinate the channel. We find that when the wholesale contracts are restricted to be linear, it can still be optimal for a manufacturer to cooperate with one retailer on advertising, but it is never optimal to support both retailers. When the wholesale contracts are two-part tariffs, we show that cooperative advertising is no longer needed. This is consistent with the general viewpoint that a sufficiently flexible wholesale contract can fully coordinate the channel and there is no need for a manufacturer to cooperate with retailers separately on advertising. Nevertheless, in practice we still observe that cooperative search advertising can play an important role. The reason could be that search advertising is one of the demand sources and changing wholesale contracts will affect the profits from other demand sources, or could be that the market conditions
of search advertising vary across time whereas the wholesale contracts cannot be adjusted frequently.

To the best of our knowledge, this is the first work that addresses cooperative search advertising. It is closely related to the literature of cooperative advertising in conventional channels (Berger 1972, Desai 1997, Bergen and John 1997, Kim and Staelin 1999, etc.). Berger (1972) is the earliest work that analytically investigates traditional cooperative advertising. It solves for the best cooperative advertising plan for one manufacturer and one retailer, treating coop funds as price subsidies. Desai (1997) studies a different mechanism for cooperative advertising: he shows that a franchisor can charge an advertising fee from each franchisee and decide where and how the advertising dollars are spent, thus overcoming free-riding problems among franchisees. Our research is closest to Bergen and John (1997), in terms of two aspects: (1) we both consider the participation rate mechanism, and (2) we both assume that retailers are independent—manufacturers cannot directly control their pricing and advertising decisions, and they cannot be forced to participate in a co-op plan. However, Bergen and John (1997) shows that a manufacturer will provide identical co-op plans to ex ante symmetric retailers, whereas we find that in cooperative search advertising, it can be optimal for the manufacturer to sponsor only one of two retailers even when they are symmetric. The distinct results are driven by the distinct market and institutional features of the search advertising industry discussed earlier.

This chapter also relates to a large literature of competitive strategies in search advertising. Existing theoretical works have studied the impact of click fraud on advertisers’ bidding strategies and the search engine’s revenue (Wilbur and Zhu 2009), the interaction between firms’ advertising auction and price competition (Xu et al. 2011), the interplay between organic and sponsored links (Katona and Sarvary 2010), the bidding strategies of vertically differentiated firms (Jerath et al. 2011), the competitive poaching strategy (Sayedi et al. 2014 and Desai et al. 2014), the impact of advertisers’ budget constraints on their own profits and the platform’s revenue (Lu et al. 2015), the effect of real-time bidding on advertisers’ strategies and profits (Sayedi 2017), etc. We contribute to the literature by incorporating channel coordination in
Lastly, we contribute to the literature about position auctions. The auction mechanism design and equilibrium properties have been investigated extensively (Edelman et al. 2007, Varian 2007, Feng 2008, Chen and He 2011, Athey and Ellison 2011, Zhu and Wilbur 2011, Dellarocas 2012, etc.), but these studies all assume that bidders are independent. In our setting, a manufacturer’s profit comes from not only its own website, but also its retailers’ websites, so the bidders are not independent from each other any more. Our analysis of the position auctions with non-independent bidders makes a contribution to this stream of literature. There are also recent works investigating collusive bidding behaviors (Decarolis et al. 2017, Decarolis and Rovigatti 2017), where competing bidders delegate their bidding decisions to a common marketing agency. Different from our work, these papers have not considered vertical relationships in distribution channels.

The chapter unfolds as the following. In Section 4.2, we lay out and analyze our basic model. We incorporate the manufacturer’s direct participation in search advertising in Section 4.3, and consider three extensions from Section 4.1 to 4.3. Lastly, Section 4.5 concludes the chapter.

4.2 Basic Model

4.1 Position Auctions

We first introduce the assumptions on position auctions, and also briefly review the equilibrium analysis of position auctions closely following Varian (2007).

We consider a generalized second-price (GSP) position auction with two positions. The highest bidder wins the first position and pays the second highest bid; the second highest bidder wins the second position, and pays the third highest bid; all other bidders with lower bids will not get a position nor clicks, and pay zero.\(^7\) We consider

\(^7\)In practice, search engines usually use a modified GSP mechanism that adjusts the ranking according to advertisers’ “quality score”, which, roughly speaking, is the prediction of an advertiser’s click through rate. In the basic model, we consider the simple case that all advertisers have the same quality score. In Section 4.1, we will incorporate heterogeneous advertiser quality scores and show
a pay-per-click mechanism, which has been widely adopted in industry. The click through rate (CTR) of the $i$-th position is denoted as $d_i$ ($i = 1, 2$), which is defined as the fraction of clicks on this position out of all impressions displayed to consumers. Suppose there are $n \geq 3$ bidders in the market. To simplify notations, we can view the current position auction equivalently as the one with $n$ positions where the positions 3 up to $n$ have zero CTR, i.e., $d_i = 0$ for $i = 3 \cdots n$. A higher position is assumed to have a higher CTR, i.e., $d_1 \geq d_2 \geq 0$.

Furthermore, the CTR for each position is assumed to be independent of the identity of the advertiser who takes that position. We will relax this assumption in Section 4.1. Following Varian (2007) and Edelman et al. (2007), we assume the auction is a complete-information simultaneous game, where each bidder knows others' valuations or payoffs per click. This assumption can be justified by considering that bidding takes place frequently, and as a result, after many rounds of bidding, the bidders will be able to infer each other's valuations.

It turns out that there are infinite Nash equilibria for the position auction. Varian (2007) and Edelman et al. (2007) have come up with some equilibrium refinement rules. We recap their results with two positions and $n$ independent bidders by the following lemma.

Lemma 1 Consider $n$ independent bidders ($n \geq 3$) competing for two positions, with payoffs per click $v_1 > v_2 > \cdots > v_n$. In equilibrium, bidder $i$ will get the $i$-th position ($i = 1, \cdots, n$). The equilibrium bids by bidder $i \geq 2$ are respectively,

$$b_2^* = \frac{d_1 - d_2}{d_1} v_2 + \frac{d_2}{d_1} v_3,$$

$$b_i^* = v_i, \ i = 3, \cdots, n,$$

and bidder 1's equilibrium bid $b_1^* \geq b_2^*$.

It is worthwhile to understand the result above, because we will analyze equilibria for more complex bidding games with non-independent bidders below. To understand the

---

8 Notice that a second-price auction with only one position is a special case of our model with $d_2 = 0$. 

Notice that the findings of the basic model generalize nicely.
result in Lemma 1, let's first consider bidder 2's one potential deviation—by bidding higher, it may be able to get the first position. To guard against this deviation, we must have \( d_2(v_2 - b_3) \geq d_1(v_2 - b_1) \). Varian (2007) proposed the concept of symmetric Nash equilibrium (SNE) by requiring that \( d_2(v_2 - b_3) \geq d_1(v_2 - b_2) \). This is a stronger condition because \( b_1 \geq b_2 \), hence the SNE is a subset of Nash equilibria (NE). A nice property of SNE is that bidders with higher payoff per click will always get a higher position. Varian (2007) further proposed the equilibrium selection criterion LB (short for lower bound), which selects the lowest possible bids from SNE. The LB rule implies that \( d_1(v_2 - b_2) \geq d_2(v_2 - b_3) \). The interpretation of this requirement is that, if it happens that bidder 1 bid so low that bidder 2 slightly exceeded bidder 1's bid and moved up to the first position, bidder 2 would earn at least as much profit as it makes now at the second position. Combining the SNE and LB criteria, we have \( d_2(v_2 - b_3^*) = d_1(v_2 - b_2^*) \), from which we can get the expression of \( b_2^* \) in equation (4.1). One can verify that \( b_i^* = v_i \) for \( i \geq 3 \) is also an SNE and satisfies the LB equilibrium selection rule. Therefore, under Varian (2007)'s SNE and LB equilibrium selection criteria, the bidders that do not win a position will bid its true valuation, and the bidders who get a position will underbid.

4.2 Coordination Game

After introducing the position auctions above, let us continue to specify the assumptions on the game of channel coordinations. Consider a channel with one manufacturer, who produces and sells one product via two retailers. In the basic model, we assume that the manufacturer does not sell directly to consumers, and thus does not participate in search ads auctions directly. This assumption not only simplifies the equilibrium analysis of the basic model, but also helps isolate the manufacturer's key managerial question—how many and which retailer(s) to cooperate with in search advertising. In Section 4.3, we will relax this assumption by allowing the manufac-

---

9Edelman et al. (2007) proposed the concept of "locally envy-free equilibria" which requires that each bidder cannot improve its payoff by exchanging bids with the bidder ranked one position above it, and it yields the same result as the LB and SNE in Varian (2007).
turer to sell directly to consumers, and consider its endogenous participation in the position auction.

The manufacturer first signs a wholesale contract with each retailer. The wholesale contract between the manufacturer and each retailer $i$ essentially determines the manufacturer's and retailer $i$'s profit margins for each product sold via the retailer, which are denoted as $m_i$ and $r_i$ respectively. $m_i$ and $r_i$ are assumed to be exogenously given when the manufacturer and retailers make their search advertising decisions. In reality, the manufacturer and retailers may sell products via both online and offline channels, and even for the online channel, search advertising is only one of the demand sources, so it is reasonable to assume that the wholesale and retail prices have been determined before they make search advertising decisions. We will consider endogenous retail prices and wholesale contracts in Section 4.3.

The conversion rate at retailer $i$'s site is denoted as $\theta_i$, which means that $\theta_i$ fraction out of all clicks on the retailer's site will convert into purchases eventually. We assume that the conversion rate is independent of the position and hence the bid of the sponsored link. This assumption is consistent with some recent empirical findings (e.g., Narayanan and Kalyanam 2015). Given these assumptions, retailer $i$'s profit per click is $\theta_i r_i$, the manufacturer's profit per click is $\theta_i m_i$, and the channel profit per click is $\theta_i (m_i + r_i)$.

We incorporate inter-brand competition in the game by introducing one strategic outside advertiser representing other brands or other retailers. The outside advertiser is denoted as $A$, and its profit per click is assumed to be $v_A$. To summarize, in our basic model, there are in total $n = 3$ bidders—two retailers and one outside advertiser.

We consider the following game. First, the manufacturer decides the participation rate $\alpha_i \in [0, 1)$ for each retailer $i = 1, 2$, which means that the manufacturer will contribute $\alpha_i$ percentage of retailer $i$'s spending on search advertising, and retailer $i$ only needs to pay the remaining $1 - \alpha_i$ percentage. Second, each retailer $i$ decides its bid $b_i$, and outside advertiser $A$ decides its bid $b_A$. Lastly, given everyone's bid, the auction outcome realizes, and the advertisers' demands and profits realize.
4.3 Equilibrium Analysis

We can solve the game by backward induction.

We first consider each retailer’s bidding strategy. The following lemma characterizes how the participation rate $\alpha_i$ from the manufacturer changes retailer $i$’s bidding strategy.

**Lemma 2** Given the manufacturer’s participation rate $\alpha_i$, retailer $i$’s equivalent profit per click in the position auction will be $\theta_i r_i / (1 - \alpha_i)$. In other words, its bidding strategy will be the same as if its profit per click was $\theta_i r_i / (1 - \alpha_i)$ and there was no support from the manufacturer.

**Proof.** Suppose a retailer will get position $i$ and pay $p_i$ in equilibrium. The Nash equilibrium condition that guards against the retailer to deviate to position $j \neq i$ is that,

$$d_i[\theta_i r_i - (1 - \alpha_i)p_i] \geq d_j[\theta_j r_j - (1 - \alpha_j)p_j], \quad (4.3)$$

which is equivalent to

$$d_i[\theta_i r_i/(1 - \alpha_i) - p_i] \geq d_j[\theta_j r_j/(1 - \alpha_j) - p_j]. \quad (4.4)$$

Similarly, we can write down and transform the SNE and LB conditions as above. Therefore, the retailer’s equilibrium bid will be the same as if its profit per click was $\theta_i r_i / (1 - \alpha_i)$ and there was no support from the manufacturer. ■

Lemma 2 shows that, by choosing participation rate $\alpha_i$, the manufacturer essentially determines retailer $i$’s equivalent profit per click on $[\theta_i r_i, +\infty)$. The manufacturer can incentivize retailer $i$ to bid as high as possible by choosing $\alpha_i$ close to one; however, the manufacturer is not able to force the retailer to bid lower than $\theta_i r_i$, because $\alpha_i \geq 0$. We will discuss the advantage and disadvantage of using negative $\alpha_i$ to coordinate channel in Section 4.4. Another implication of Lemma 2 is that, in the position auction, outside bidders do not need to observe the participation rate $\alpha_i$ or the profit per click $\theta_i r_i$; instead, they only need to know each retailer’s equivalent
profit per click $\theta_i r_i/(1 - \alpha_i)$, which has been assumed to be observable due to repeated bidding. Therefore, the results of our model do not rely on the observability of the channel coordination contract $\alpha_i$, which has been shown to be a critical assumption that greatly influences the equilibrium channel structure (Coughlan and Wernerfelt 1989).

Before continuing to analyze the manufacturer’s strategy in equilibrium, we set a restriction on the parameter space, which will facilitate the equilibrium analysis. We assume that $v_A > \theta_i r_i$ ($i = 1, 2$). Under this condition, the two retailers are not able to win position 1 without the manufacturer’s support, and different levels of participation rates can potentially move retailers to different positions. This is the most interesting case to study because it allows us to investigate whether it is optimal for the manufacturer to sponsor one or both retailers, and how much support it should provide to each retailer. We will analyze the case where $v_A$ is between $\theta_1 r_1$ and $\theta_2 r_2$ later as a robustness check. The case with $v_A$ below both $\theta_1 r_1$ and $\theta_2 r_2$ is trivial, because by Lemma 1, we know that in this case, the two retailers can get the first and second positions even without the manufacturer’s sponsorship.

Given the manufacturer’s participation rates $\alpha_1$ and $\alpha_2$, the position rank of the three bidders is determined by the order of $\theta_1 r_1/(1 - \alpha_1)$, $\theta_2 r_2/(1 - \alpha_2)$, and $v_A$. This results in six possible position configurations. For each position configuration, we can write down the manufacturer’s profit function and maximize it with respect to $\alpha_1$ and $\alpha_2$. Then we compare the manufacturer’s profits under all six position configurations to determine the manufacturer’s optimal choice of participation rates $\alpha_1^*$ and $\alpha_2^*$.

We use an array $(X, Y)$ to denote a position configuration in which bidder $X$ takes the first position and bidder $Y$ takes the second position. Without loss of generality, we adopt the following tie-breaking rule: when $R_1$ bids the same as $R_2$ or $A$, $R_1$ wins; when $R_2$ bids the same as $A$, $R_2$ wins.

We first consider position configuration $(R_1, R_2)$, i.e., the configuration in which retailer 1 gets position 1, and retailer 2 gets position 2. According to Lemma 1, we need to have $\theta_1 r_1/(1 - \alpha_1) \geq \theta_2 r_2/(1 - \alpha_2) \geq v_A$ for this position configuration to be the equilibrium. The outside advertiser’s equilibrium bid is $v_A$, and retailer 2’s
equilibrium bid is \( \frac{d_1-d_2}{d_1} \theta_2 r_2 + \frac{d_2}{d_1} v_A \). The manufacturer’s profit will be,

\[
\pi_M(\alpha_1, \alpha_2) = d_1 \left[ \theta_1 m_1 - \alpha_1 \left( \frac{d_1-d_2}{d_1} \theta_2 r_2 + \frac{d_2}{d_1} v_A \right) \right] + d_2 (\theta_2 m_2 - \alpha_2 v_A),
\]

which decreases with both \( \alpha_1 \) and \( \alpha_2 \). Therefore, the manufacturer will choose the smallest participation rates that ensure \( \theta_1 r_1/(1-\alpha_1) \geq \theta_2 r_2/(1-\alpha_2) \geq v_A \), which are

\[
\alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A}, \quad (4.6)
\]
\[
\alpha_2^* = 1 - \frac{\theta_2 r_2}{v_A}. \quad (4.7)
\]

Correspondingly, the manufacturer’s profit under the optimal participation rates will be

\[
\pi_M(\alpha_1^*, \alpha_2^*) = d_1 [\theta_1 (m_1 + r_1) - v_A] + d_2 [\theta_2 (m_2 + r_2) - v_A]. \quad (4.8)
\]

Following the same procedure, we can work out the optimal participation rates and the manufacturer’s profit under the other five position configurations. By comparing the manufacturer’s profits among all six position configurations, we can determine the equilibrium position configuration. The following theorem summarizes the equilibrium outcome, with the proof in the appendix.

**Theorem 2** Consider a position auction, the participants of which consist of a manufacturer’s two retailers and an outside advertiser. The manufacturer supports retailers in search advertising by sharing a fraction of each retailer’s advertising expense. Assume that \( v_A > \theta_i r_i \) for \( i = 1, 2 \).

In equilibrium, the retailer \( i \) with a higher channel profit per click \( \theta_i(m_i + r_i) \) will take a higher position than the other retailer. If it is optimal for the manufacturer to support only one retailer, it will choose to support this retailer.

Suppose \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \) without loss of generality, then the possible position configurations are \((R_1, R_2)\), \((R_1, A)\), and \((A, R_1)\).
The equilibrium position configuration is \((R_1, R_2)\) if and only if
\[
\frac{d_1 - d_2}{d_2} \theta_1 m_1 + \theta_2 m_2 \geq \frac{d_1 + d_2}{d_2} v_A - \frac{d_1 - d_2}{d_2} \theta_1 r_1 - \theta_2 r_2 - \max \{\theta_1 r_1, \theta_2 r_2\}.
\]

The corresponding equilibrium participation rates are \(\alpha_i^* = 1 - \theta_i r_i/v_A\) for \(i = 1, 2\).

The equilibrium position configuration is \((R_1, A)\) if and only if
\[
\frac{d_1 - d_2}{d_2} \theta_1 m_1 \geq \frac{d_1 - d_2}{d_2} v_A + \min \{\theta_1 r_1, \theta_2 r_2\} - \frac{d_1 - d_2}{d_2} \theta_1 r_1 - \frac{\theta_1 r_1 \cdot \theta_2 r_2}{v_A},
\]
\[
\theta_2 m_2 < v_A - \theta_2 r_2 + \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}.
\]

The corresponding equilibrium participation rates are \(\alpha_1^* = 1 - \theta_1 r_1/v_A\) and \(\alpha_2^* = 0\) for \(i \neq j = 1, 2\).

The equilibrium position configuration is \((A, R_1)\) if and only if
\[
\frac{d_1 - d_2}{d_2} \theta_1 m_1 < \frac{d_1 - d_2}{d_2} v_A + \min \{\theta_1 r_1, \theta_2 r_2\} - \frac{d_1 - d_2}{d_2} \theta_1 r_1 - \frac{\theta_1 r_1 \cdot \theta_2 r_2}{v_A},
\]
\[
\frac{d_1 - d_2}{d_2} \theta_1 m_1 + \theta_2 m_2 < \frac{d_1 + d_2}{d_2} v_A - \frac{d_1 - d_2}{d_2} \theta_1 r_1 - \theta_2 r_2 - \max \{\theta_1 r_1, \theta_2 r_2\}.
\]

The corresponding equilibrium participation rates are \(\alpha_1^* = \max \{1 - \theta_1 r_1/(\theta_2 r_2), 0\}\) and \(\alpha_2^* = 0\) for \(i \neq j = 1, 2\).

The theorem above completely characterizes the manufacturer’s optimal sponsorship strategy, by essentially providing answers to three managerial questions: (1) which retailer to sponsor, (2) how much sponsorship to provide, and (3) how many retailers to sponsor. We will discuss the implications of Theorem 2 on these three questions one by one below.

Regarding the first question, the theorem above shows that the two retailers’ relative positions in equilibrium are entirely determined by their channel profits per click. When deciding which retailer to support, the manufacturer should rely its
decision on the channel profit per click instead of its own profit per click. In this sense, the manufacturer acts as if the channel is fully coordinated when deciding the two retailers’ relative positions. The intuition is that, in cooperative search advertising, the manufacturer needs to consider not only its own profit per click, but also how much it needs to pay for the retailer to get a good position. Hence there are two reasons why not only the manufacturer’s own profit per click but also the retailers’ profits per click matter. First, when a retailer’s own profit per click is relatively high, it already has relatively high willingness-to-pay for the position, and thus the manufacturer can sponsor this retailer to get a higher position with a relatively low cost. Second, if the manufacturer chooses to sponsor a retailer with relatively low profit per click, the other retailer with relatively high profit per click will become a strong competitor in the position auction, which will raise the bidding cost and thus the manufacturer’s sponsorship cost. Taking both rationales into consideration, the manufacturer should sponsor the retailer with a higher channel profit per click to get a higher position.

We should notice that we obtain the results above based on the assumption that $v_A \geq \theta_i r_i$ ($i = 1, 2$). We also consider the case where $v_A$ is between $\theta_1 r_1$ and $\theta_2 r_2$, with details of the analysis relegated in the appendix. We find that given $\theta_i r_i \geq v_A \geq \theta_j r_j$, retailer $i$ will get a higher position when $\theta_i (m_i + r_i) \geq \theta_j (m_j + r_j)$; but retailer $j$ may not get a higher position when $\theta_j (m_j + r_j) > \theta_i (m_i + r_i)$. In other words, in this case, a retailer with higher channel profit per click may not necessarily get a higher position; yet, a retailer with both higher total channel profit and higher own profit per click will always get a higher position than the other retailer. Therefore, the results in Theorem 2 rely on the assumption that $v_A \geq \theta_i r_i$ ($i = 1, 2$), which we do think is a more reasonable case to study. This is because given $\theta_i r_i \geq v_A \geq \theta_j r_j$, the manufacturer will never provide positive participation rate to retailer $i$, thus will have no control of this retailer’s bidding. It only needs to decide whether and how much to support retailer $j$. Thus retailer $i$ is like an outside advertiser that is not of the manufacturer’s concern in terms of cooperative advertising. In contrast, when $v_A \geq \theta_i r_i$ ($i = 1, 2$), the manufacturer will consider supporting both retailers under
Regarding the second question about optimal sponsorship rates, we interpret Theorem 2 under each position configuration respectively. In configuration \((R_1, R_2)\), the manufacturer provides positive participation rates to both retailers. Specifically, it carefully chooses the participation rates such that the equivalent profits per click for both retailers are equal to \(v_A\), which is also the price per click for both positions. As a result, both retailers earn zero profit, and the manufacturer collects the entire channel profit. In configuration \((R_1, A)\), the manufacturer only provides support to retailer 1. It will not provide a positive participation rate to retailer 2 because that will raise the equilibrium bid of advertiser \(A\) and thus increase the price per click for retailer 1. In configuration \((A, R_1)\), the manufacturer still provides zero participation rate to retailer 2, and it needs to provide a positive participation rate to retailer 1 only when this retailer's own profit is lower than that of retailer 2.

Lastly, to answer the third question about how many retailers to sponsor, Theorem 2 describes the exact conditions for the manufacturer to sponsor two, one, or zero retailer(s) given two asymmetric retailers. The key tradeoff here is higher demand versus higher bidding cost resulting from intensified competition. Specifically, by supporting more retailers, the manufacturer will get more demand; but at the same time, the bidding costs will go up as retailers now compete with each other by bidding higher. To understand the key tradeoff more intuitively, we consider a special case where the two retailers are ex ante symmetric, i.e., \(\theta_1 = \theta_2 = \theta\), \(m_1 = m_2 = m\), and \(r_1 = r_2 = r\), and plot Figure 4-2 that completely characterizes a manufacturer's optimal cooperative search advertising strategy for two symmetric retailers.

Figure 4-2 indicates that the manufacturer provides positive participation rates to both retailers when the channel profit per click is relatively high; it provides support to neither retailers if its own profit per click and each retailer's profit per click are both relatively low. More interestingly, we find that the manufacturer will provide positive participation rate to only one retailer when its profit per click is relatively high but the retailers' profit per click is relatively low. In this case, retailers need high participation rates to move up to a higher position, but it is too expensive for
the manufacturer to support both retailers. For cooperative advertising, it has been noticed that the manufacturer should provide identical co-op programs to symmetric retailers (Bergen and John 1997), whereas our finding indicates that the manufacturer may optimally offer the co-op program only to a subset of retailers even if they are ex ante symmetric. The reason is that in traditional advertising, ads from different retailers usually appear in different places and have positive spillovers, i.e., a consumer may see ads from retailer A but make the purchase from retailer B. In search advertising, different retailers compete with each other directly in the position auction, and their competition for demand also becomes stronger due to smaller differentiation. Therefore, the manufacturer takes hold of the cooperative advertising money to reduce potential competition among retailers.

4.4 Integrated Channel

It is worthwhile to compare our basic model with the case of channel integration, where the manufacturer has full control of both retailers’ bids. First, we notice that the integration does not necessarily generate a higher channel profit (McGuire and Staelin 1983), because there is a strategic outsider advertiser who can choose its bid based on whether the channel is integrated. If the outside advertiser chooses to bid higher when the channel is integrated, the bidding cost for the integrated channel can get higher, and the channel profit can get lower. This is exactly what we find under certain circumstances below.

When the channel is integrated, the two retailers are no longer independent bidders in the auction. The integrated channel as a whole gets profits from two retailers, and one retailer’s bid will influence the channel profit by affecting the other retailer’s cost per click even when the two retailers’ positions are given, which differs from the case of independent bidders where each bidder’s bid does not affect its own profit given its position. As a result, for any given position configuration, we cannot directly apply Lemma 1 to determine each player’s equilibrium bid at this position configuration. Instead, we use the following rule to determine each player’s equilibrium bids at a given position configuration.
• For the outside advertiser, since it is still independent from other players, we still use Varian (2007)'s Symmetric Nash Equilibrium (SNE) and Lower Bound (LB) equilibrium selection rules to determine its equilibrium bid.\textsuperscript{10}

• For the two retailers, the SNE rule is no longer a sensible refinement rule to select equilibrium bids at a given position configuration. This is because SNE posits a stronger condition than NE that raises the bid of the retailer at the lower position. As the lower-ranked retailer’s bid gets higher, the higher-ranked retailer’s cost per click increases and its profit decreases. Therefore, equilibrium bids selected by SNE do not maximize the channel profit, and the channel can deviate to other bids that lead to higher channel profit. Thus the bids selected by SNE can fail to be a NE. Therefore, we use a new equilibrium selection rule—profit maximization (PM) criterion—to determine the two retailers’ equilibrium bids at a given position configuration. Particularly, given a position configuration, among all bids $b_1$ and $b_2$ that satisfy the Nash equilibrium condition, we select the equilibrium that maximizes the integrated channel’s profit. Given our objective is to show that under some circumstances, the integrated channel can result in a lower channel profit than a non-integrated channel, the PM rule is conservative in that it selects the equilibrium with the maximum integrated channel profit.\textsuperscript{11}

We determine the equilibrium position configuration in the following way. For

\textsuperscript{10}This means that the outside advertiser can anticipate the integrated channel’s bids $b_1$ and $b_2$, and then decides its bid $b_A$ according to Lemma 1.

\textsuperscript{11}The PM criterion also facilitates the equilibrium analysis. Particularly, the SNE and LB rules for the outside advertiser will determine its bid as a function of the integrated channel’s bids, like what happens in Lemma 1. Then when the integrated channel follows the PM criterion to pick the most profitable equilibrium, it essentially chooses its bids after taking into account the outsider advertiser’s response to its bids. Mathematically, this is as if the integrated channel chooses its bids first while anticipating the outside advertiser’s response, and then the outside advertiser observes the integrated channel’s choice and then decides its bid. Therefore, the PM criterion essentially transforms the current simultaneous bidding game equivalently into a dynamic game with the integrated channel bidding first. The equilibrium outcome is the same whether the outside advertiser anticipates or observes the integrated channel’s bids. In fact, in Varian (2007)’s setup with independent bidders, consider a dynamic game where one bidder decides its bid first and other bidders observe the bid and decide their bids subsequently. This dynamic game will yield the same equilibrium outcome as a static game where everyone bids simultaneously, as long as the SNE and LB equilibrium selection rules are maintained.
a position auction with two retailers and one outside advertiser, there are six possible position configurations in total. For each position configuration, we determine the integrated channel and the outside advertiser’s equilibrium bids, using the selection rules we described above. Given that the integrated channel’s objective is to maximize profit and it must have no incentive to deviate to any other position configuration, we can get the Nash equilibrium condition, i.e., the parameter space that this position configuration can hold in equilibrium. It turns out that the Nash equilibrium conditions for the six position configurations are not mutually exclusive. For the parameter setting where only one position configuration can exist, this position configuration will be the one in equilibrium. For the parameter setting where multiple position configurations co-exist, the integrated channel will select the equilibrium position configuration as the one with the highest channel profit. The idea is that the integrated channel can choose which position configuration to be in equilibrium by choosing its bids $b_1$ and $b_2$.

The detailed analysis is in the appendix. We summarize the main findings below.

**Proposition 6** Consider a channel with one manufacturer and two retailers. The two retailers and one outside advertiser $A$ compete for two positions. Assume $v_A > \theta_i r_1$, for $i = 1, 2$.

- In both integrated and non-integrated channels, the retailer with a higher channel profit per click $\theta_i(m_i + r_i)$ gets a higher position than the other retailer in equilibrium.

- The channel profit is the same for the integrated and non-integrated channels when the two retailers win both positions.

- Under certain condition, the channel profit could be lower in the integrated channel than in the non-integrated channel.\(^\text{12}\)

\(^{12}\)Suppose $\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)$, then the channel profit in the integrated case is lower compared with the non-integrated case when $v_A - \frac{d_1 (v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A} < \theta_2(m_2 + r_2) < v_A$ and $\theta_1(m_1 + r_1) \geq \frac{d_1}{d_1 - d_2} \theta_2(m_2 + r_2)$. This is a sufficient but not a necessary condition.
The first two findings suggest the efficiency of the participation rate mechanism. We have discussed the intuition for the first finding. The intuition for the second finding is as follows. For the integrated channel, when the two retailers get the two positions, the channel maximizes channel profit by letting the retailer who takes the second position submit the lowest possible bid $v_A$. For the non-integrated channel, the manufacturer can choose the participation rates so that the equivalent profit per click for both retailers are $v_A$ and the retailer who takes the second position bids exactly $v_A$, and as a result, the manufacturer collects the total channel profit by itself, which is the same as that in the integrated case.

The third finding is surprising. It implies that it is not an easy task to come up with a "first best" benchmark that maximizes the channel profit. That’s the reason why we instead focus on a simple and widely used coordination mechanism—participation rates. The intuition for this finding is as follows. Suppose $\theta_1(m_1+r_1) \geq \theta_2(m_2+r_2)$. When the channel is integrated, retailer 2’s profit per click is $\theta_2(m_2+r_2)$, whereas when the channel is non-integrated, retailer 2’s profit per click is $\theta_2 r_2$ if there is no sponsorship from the manufacturer. At the position configuration $(R_1, A)$, retailer 2 in the integrated channel will have a higher incentive to move up to the second position compared with the non-integrated channel, since it has a higher profit per click. Knowing this, the outside advertiser A will bid higher in equilibrium to keep its position. As a result, the cost per click for retailer 1 at the first position will be higher when the channel is integrated, which leads to a lower total channel profit.

Given the finding above, we can further consider whether there is a more efficient way to coordinate the channel than the channel integration. The answer is yes. In our participation-rate mechanism, we restrict that the participation rate $\alpha_i$ must be non-negative. If we allow $\alpha_i$ to be negative, which means that the retailer needs to pay the manufacturer a certain proportion of its spending on search advertising, then under position configuration $(R_1, A)$ and $(A, R_1)$, the manufacturer can force $R_2$ to bid 0 by choosing $\alpha_2 \to -\infty$. This minimizes the cost per click for $R_1$ and achieves a higher channel profit. However, it is worthwhile noticing that negative participation rates increase the channel profit at the price of a lower retailer profit. Therefore,
retailers may have no incentive to participate in a cooperative advertising program with negative participation rates, unless more complex contracts are involved, such as two-part tariffs. In contrast, by restricting the participation rates to be non-negative in our basic model, we guarantee that both the manufacturer and retailers have the incentive to participate in the cooperative advertising program.

4.3 Manufacturer’s Direct Participation in Search Advertising

In the basic model, we have assumed that the manufacturer does not participate in the position auction directly so that we can focus on the manufacturer’s decision of which retailer(s) to support and how much support to provide. In this section, we relax this assumption by allowing the manufacturer to participate in the search advertising directly. We will show that the results from the basic model are robust. Moreover, we have some new insights about when the manufacturer wants to advertise directly and when to advertise via the retailer(s).

We will have four bidders in the auction: one manufacturer, its two retailers, and an outside advertiser. We denote $m_0$ as the manufacturer’s profit margin when selling directly to consumers, which could be either greater or less than $m_i + r_i$, the channel profit margin when selling via retailer $i$. We further denote $\theta_0$ as the conversion rate at the manufacturer’s site. Therefore, the manufacturer’s profit per click on its own site is $\theta_0 m_0$. The manufacturer’s bid is denoted as $b_0$.

The game still proceeds in two stages. In the first stage, the manufacturer not only decides the participation rate $\alpha_i \in [0, 1)$ for each retailer $i = 1, 2$, but also decides whether to participate in the auction directly in the second stage. We assume that the manufacturer can credibly commit not to participate in the auction. The reason why the manufacturer may want to commit to not participate in the auction is that the manufacturer’s participation may increase the outside advertiser’s bid, leading to higher bidding costs and lower profit for the manufacturer. In the second

\footnote{The manufacturer can commit by, for example, not setting up a search engine marketing team.}
stage, each retailer $i$ decides its bid $b_i$, outside advertiser $A$ decides its bid $b_A$, and the manufacturer decides its bid $b_0$ if it has decided to participate. Lastly, given everyone's bid, the auction outcome realizes, and the advertisers' demands and profits realize.

In the position auction, if the manufacturer participates in bidding directly, the manufacturer makes profits from its own site as well as the retailers' sites; therefore, the manufacturer and retailers are non-independent bidders in the auction. Similar to the case of channel integration in Section 4.4, we cannot directly apply Lemma 1 to determine each player’s equilibrium bid at any given position configuration. For the retailers and the outside advertiser, we still use Varian (2007)’s Symmetric Nash Equilibrium (SNE) and Lower Bound (LB) equilibrium selection rules to determine their equilibrium bids, because they make profits only from their own sites and their choice of bids does not affect their own profits directly at any given position. However, for the manufacturer, its choice of bid does affect its profit by affecting the price per click on the retailers’ sites when it is at a lower position than one or both retailers. In such case, the SNE is no longer a sensible criterion to select the manufacturer’s bid, since it fails to be a NE. Therefore, for the position configurations in which the manufacturer is below one or both retailers, we will use the manufacturer’s profit maximization (PM) criterion in determining the manufacturer’s equilibrium bid (we have introduced the PM criterion in Section 4.3). For the position configurations in which the manufacturer is above the retailer, the manufacturer’s bid does not affect its profit directly, so we still use the SNE and LB criterion to select the manufacturer’s bid.

We solve the game by backward induction. In the second stage, given the manufacturer’s participation decision and participation rates, each bidder’s (equivalent) profit per click will be determined. Given four bidders competing for two positions, there are a total of $4 \times 3 = 12$ possible position configurations. For each position configuration, we determine the bids of the manufacturer, the retailers, and the outside advertiser, and the manufacturer’s optimal participation rates $\alpha_1, \alpha_2$. If the manufacturer’s optimal bid $b_M^*$ is 0, then the manufacturer should commit to quit the bidding in the first stage. Back to the first stage, the manufacturer compares
its profits over the twelve position configurations to determine the equilibrium. The equilibrium condition for one position configuration is that the manufacturer's profit under this position configuration is greater than that under all other position configurations. We find that when this equilibrium condition holds, the Nash equilibrium condition for this position configuration always holds. The detailed equilibrium analysis is relegated to the appendix. Similarly, we restrict our attention to the case with $\theta_i r_i < v_A$ and $\theta_i r_i < \theta_0 m_0$ for $i = 1, 2$. That is, without the manufacturer's support, neither retailer can get displayed. Again, this is the most interesting case because it can be optimal for the manufacturer to sponsor 0, 1, or 2 retailers and different levels of participation rates will move the retailers to different positions. The following proposition characterizes the equilibrium, with proof in the appendix.

**Theorem 3** Consider a position auction, the participants of which consist of a manufacturer, its two retailers and an outside advertiser. The manufacturer supports the two retailers in search advertising by providing certain participation rates. Assume that neither retailer can get displayed without support ($\theta_i r_i < v_A$, and $\theta_i r_i < \theta_0 m_0$ for both $i = 1, 2$).

- The retailer who has a higher channel profit per click will get a higher position than the other retailer in equilibrium.

- The manufacturer will take a higher position than a retailer when its profit per click via direct sales exceeds the channel profit per click of the retailer.

- The manufacturer will definitely participate in bidding when its profit per click is higher than both retailers' channel profits per click, and may participate in bidding when its profit per click is lower than one or both retailers' channel profit per click but not too low.

- When the manufacturer participates in bidding, it may still sponsor one retailer but will not sponsor both retailers. The manufacturer will support both retailers only when its profit per click is lower than both retailers' channel profits per click and the manufacturer quits the bidding itself.
Theorem 3 shows that one key finding in the basic model remains true when the manufacturer is allowed to participate in the search ad auction directly, i.e., the retailer with a higher channel profit per click will get a higher position than the other retailer in equilibrium. Furthermore, the manufacturer will get a higher position than both retailers if its profit per click via direct sales is higher than both retailers’ channel profit per click. On the other hand, the manufacturer may want to sponsor one or both of the retailers to get a higher position than itself when they get a higher channel profit per click, even though the manufacturer may earn a higher profit margin from direct sales than via the retailers (i.e., $m_0 > m_1, m_2$).

Similar to the basic model, we find that in deciding whether to participate in search advertising directly or indirectly via retailers, the manufacturer compares its profit per click with the retailers’ channel profit per click. The intuition is that when a retailer’s profit per click $\theta_i r_i$ is high, the retailer will bid high without the manufacturer’s support, so sponsoring this retailer to get a higher position is less costly. On the other hand, if the manufacturer wants to get a high position for its own site, it needs to bid higher than the retailer, which can be very costly. As a result, it can be optimal for the manufacturer to support the retailer(s) rather than to get a higher position by itself.

The manufacturer’s equilibrium bidding and sponsoring strategies can be very complicated under a general setting, as shown in the appendix. Therefore, it is informative to consider a special case where the two retailers are symmetric and their channel profit is the same with the manufacturer’s profit (i.e., $\theta_1 = \theta_2 = \theta_0 = \theta$, $m_1 = m_2 = m$, $r_1 = r_2 = r$, and $m + r = m_0$). In this case, $(M, R), (R, M), (R, R)$ will lead to exactly the same profit for the manufacturer, as long as the channel takes both positions; $(A, M), (A, R)$ will also lead to the same profit for the manufacturer, as long as the channel takes the second position; whereas $(M, A)$ will always be more profitable for the manufacturer than $(R, A)$. Figure 4-3 illustrates the manufacturer’s optimal bidding and sponsoring strategies.

We find that, the manufacturer will sponsor one or both retailers if and only if $\theta m + 2\theta r \geq 2v_A$; otherwise, the manufacturer will bid directly and takes the first
position when the channel profit per click is higher than the outside advertiser’s profit per click, i.e., \( \theta(m + r) = \theta m_0 \geq v_A \), and he will neither bid nor sponsor when \( \theta(m + r) = \theta m_0 < v_A \). The result indicates the key role of channel profit per click in determining the relative position rank, and also implies that there exists the situation where the manufacturer sponsors a retailer and bid by itself at the same time (under position configurations \((R, M)\) or \((M, R)\)).

### 4.4 Extensions

#### 4.1 Identity-Dependent Click-Through Rate

In the basic model, we assume that the CTR at each position is independent of the identity of the advertiser. In reality, this assumption may not hold. For example, some consumers may be loyal to a retailer and more likely to click on its sponsored ads even if the retailer is not at the top position. In this subsection, we consider an extension of the basic model that allows the CTR of a sponsored ad to depend on both its position and the identity of its advertiser.

Specifically, we assume that the CTR of advertiser \(i\) at position \(j\), \(d_{ij}\) can be decomposed as \(d_{ij} = e_i x_j\), where \(e_i\) is the “identity effect”, which measures the attractiveness of the advertiser, and \(x_j\) is the “position effect”, which measures the attractiveness of the position.\(^{14}\) Similar as before, we assume that given an advertiser, the higher position it takes, the more clicks it will get, i.e., \(x_j\) decreases in \(j\). Moreover, it is common for search advertising platforms to adjust the position rank of advertisers according to their identity effects. For example, when deciding the positions ranks, Google augments each advertiser’s bid with its quality score, which, roughly speaking, is a measure of the advertiser’s predicted CTR, i.e., the identity effect, besides other less important considerations such as landing page quality. We follow Varian (2007) to assume that the positions of advertisers are ranked according to \(e_i b_i\) in descending order, where \(b_i\) denotes the bid of the advertiser at position \(i\).

\(^{14}\)We have not fully modeled consumers’ searching and clicking behaviors, which is beyond the scope of the chapter. See Athey and Ellison (2011), Chen and He (2011), Jerath et al. (2011), etc.
The advertiser at position \(i\) then pays \((e_{i+1}/e_i)b_{i+1}\) per click.

The equilibrium analysis of the model here parallels with the basic model above (with details in the appendix). We summarize the findings by the following theorem.

**Theorem 4** Consider a position auction with identity-dependent CTR, participated by an outside advertiser, and two retailers who are supported by a manufacturer. Assume that \(e_{AVA} > e_i\theta_i r_i\) for \(i = 1, 2\). In equilibrium, the retailer with higher total channel profit per impression will always take a higher position than the other retailer. If it is optimal for the manufacturer to support only one retailer, it will choose to support this retailer.

Theorem 4 generalizes the results in Theorem 2 nicely. With identity-dependent CTR, the retailer with higher total channel profit per impression, i.e., \(e_i\theta_i(m_i + r_i)\), will always take a higher position than the other retailer in equilibrium.

### 4.2 Endogenous Retail Prices with Price Competition

As argued before, the wholesale prices and retail prices can be seen as exogenously given when search advertising is only part of the demand source for the manufacturer and retailers. However, when search advertising is the main source of demand for the channel, it is more reasonable to consider endogenous wholesale contracts and retail prices. That is, when manufacturers and retailers determine their wholesale and retail prices, they take the search advertising strategy into account.

In this section, we extend the basic model to consider the case where search advertising is the main source of demand for the retailers. The retailers set retail prices endogenously, taking into account the potential price competition between them. We will analyze the effect of price competition on manufacturer's sponsorship strategy and investigate whether the insights from the main model generalize here. That is, whether the manufacturer may still sponsor only one retailer, even when the two retailers are symmetric. We take the manufacturer's wholesale contracts as exogenously given in this section, and we will analyze endogenous wholesale contracts in next section.
The overall effect of retailers' price competition on the manufacturer's sponsorship strategy is ambiguous. On the one hand, price competition between retailers can lead to lower retail prices and thus higher total demand for the manufacturer. On the other hand, as price competition forces retailers to charge a lower retail price, they have less incentive to win the search ad auction, and consequently, the manufacturer may need to provide higher participation rates to the retailers in order to help them get displayed. In general, the manufacturer faces the tradeoff between demand expansion and sponsorship costs, both of which depend on the intensity of retail price competition.

The same as in the main model, we consider the setting where one manufacturer sells via two retailers, and the manufacturer does not participate in the auction directly. Instead of having only one outside advertiser, we assume that there are two outside advertisers with the same profit per click, $v_A$. The game proceeds in three stages. First, the manufacturer decides the participation rates, $\alpha_1$ and $\alpha_2$; second, the two retailers decide their prices $p_1$ and $p_2$ simultaneously; lastly, the two retailers and the outside advertisers bid in the position auction with two positions. To focus on the price competition between retailers, we assume exogenous and symmetric wholesale price for the two retailers, $w$. For tractability of the analysis, we do not explicitly consider the outside advertisers' pricing decisions. We restrict ourselves to text ads, for which, consumers do not see the price information before clicking on an ad.

In practice, upon seeing two ads, some consumers may form rational expectations of their prices, and only click on the ad that they prefer more; while others may be less sophisticated or more interested in evaluating both options, and thus click on both ads before making a choice. To avoid the complexity of explicitly modeling consumers' clicking behaviors, we will focus on the extreme case where each consumer who clicks on an ad will click on both ads before making a choice. That is, the click-

---

15If there is only one outside advertiser, when retailer 1 and the outside advertiser get displayed, their cost per click is the profit per click of retailer 2. However, in this case, retailer 2 is not displayed and thus does not compete with retailer 1 directly. Therefore, retailer 2's profit per click is ambiguous. In order to get over the complexity of defining equilibrium selection rules for retailer 2's profit per click when retailer 2 does not get a position, we assume that there are two outside advertisers in the market. With two outside advertisers, when retailer 1 and one outside advertiser get displayed, their cost per click is the profit per click of the other outside advertiser.
through rates for the two positions are equal, $d_1 = d_2$, which are normalized as one. This is also the case where the price competition is most severe.\textsuperscript{16} We assume that the two retailers are symmetric \textit{ex ante}, i.e., out of all consumers who click on their ads, $\theta$ of them will consider purchasing from one of the two retailers.

Consumers who consider purchasing from one of the two retailers are assumed to distribute uniformly on a Hotelling line in $[0, 1]$. Retailer 1 is located at 0 and retailer 2 at 1. Given the two retailers' prices $p_1$ and $p_2$, a consumer at location $x$ derives utility $v - tx - p_1$ if purchasing from retailer 1 and $v - t(1 - x) - p_2$ if purchasing from retailer 2. We assume that $v > w$, i.e., the wholesale price is lower than consumer's highest valuation of the product. $t$ is the unit traveling cost that measures the degree of consumer heterogeneity. We assume that $v$ is sufficiently high such that the market is covered (otherwise there is no direct price competition between the two retailers). We will explicitly provide the market coverage condition below. The consumer at location $x^* = \frac{1}{2} + \frac{p_2 - p_1}{2t}$ will be indifferent between the two retailers.

We will investigate whether the manufacturer will sponsor both, one, or neither retailers, given symmetric retailers and symmetric wholesale prices. Similar to the main model, we assume that neither retailer can get displayed without support from the manufacturer. As we will show later, this amounts to some restrictions on $t$.

We first consider the case where the manufacturer sponsors both retailers to get displayed. Given the retail prices, the conversion rate for retailer 1 is $\theta x^*$, and the conversion rate for retailer 2 is $\theta (1 - x^*)$. Since the two positions have the same CTR, the price per click for both positions is the profit per click of the first bidder who cannot get displayed, i.e., the outside advertisers' profit per click, $v_A$. Retailer 1's profit will be $\pi_1 = \theta x_1(p_1 - w) - (1 - \alpha_1)v_A$, and retailer 2's profit will be $\pi_2 = \theta (1 - x_1)(p_2 - w) - (1 - \alpha_2)v_A$. By solving the best response function of the two retailers, we get the equilibrium retail prices $p_1^* = p_2^* = w + t$, and the equilibrium market share $x^* = 1 - x^* = 1/2$. Given the participation rates, the two retailers'\footnote{We essentially consider the case where consumers' cost of clicking is zero. If consumers face a positive cost of clicking, and if price is the only information consumers learn after clicking, the classic Diamond Paradox (Diamond 1971) implies that each retailer will set a monopolistic price, no matter how small the cost of clicking is, and there is no price competition.}
profits in equilibrium are \( \pi_1^* = \theta t/2 - (1 - \alpha_1)v_A, \pi_2^* = \theta t/2 - (1 - \alpha_2)v_A \), and the manufacturer’s profit is \( \pi_M = \theta w - (\alpha_1 + \alpha_2)v_A \). The manufacturer’s profit decreases in \( \alpha_1, \alpha_2 \), so the manufacturer will choose the lowest participation rates that can help the two retailers get displayed, i.e., \( \alpha_i^* = 1 - \frac{\theta t}{2v_A}, i = 1, 2 \). Notice that under the condition \( t < 2v_A/\theta \), \( \alpha_i^* > 0 \). The manufacturer’s profit under the optimal participation rates is \( \pi_M^* = \theta(w + t) - 2v_A \).

In the appendix, we analyze the other two cases in which the manufacturer sponsors one or neither retailer. Comparing the manufacturer’s profits under these the three cases (sponsoring two, one, zero retailers), we get the following theorem, with proof in the appendix.

**Theorem 5** Suppose that neither retailer can get displayed without support from the manufacturer and the market is covered when the manufacturer sponsors both retailers to get displayed. These two conditions amount to \( t \in (\tilde{t}, \bar{t}] \), where \( \tilde{t} = \frac{2}{3}(v - w) \), and

\[
\tilde{t} = \begin{cases} 
\frac{\theta}{4v_A}(v - w)^2, & \text{if } \frac{3}{8}(v - w) < \frac{v_A}{\theta} < \frac{v - w}{2} \\
v - w - \frac{v_A}{\theta}, & \text{if } \frac{v_A}{\theta} \geq \frac{v - w}{2}.
\end{cases}
\]

The manufacturer’s optimal sponsorship strategy depends on the profit per click of the outside advertisers \( v_A \) and the degree of consumer heterogeneity \( t \) in the following way:

- If \( v_A \geq \theta v \), the manufacturer sponsors neither retailer.
- If \( \frac{3}{2}(v + w) \leq v_A < \theta v \), the manufacturer sponsors one retailer when \( t \leq t_1 \), and sponsors neither retailer when \( t > t_1 \).
- If \( \frac{3\theta}{8}(v - w) \leq v_A < \frac{\theta}{2}(v + w) \), the manufacturer sponsors one retailer for \( t < t_2 \), sponsors neither retailer when \( t_2 \leq t < t_3 \), and sponsors both retailers when \( t \geq t_3 \).

The thresholds \( t_1, t_2, \) and \( t_3 \) are functions of \( v, w, v_A \) and \( \theta \).
The above theorem completely characterizes the manufacturer's optimal sponsorship strategy. The expressions of the thresholds $t_1$, $t_2$, and $t_3$ are complex and provided in the appendix. Notice that we need to have $v_A > \frac{3\theta}{8}(v - w)$ in order for having $t < \bar{t}$.

The results of Theorem 5 are intuitive. When $v_A$ is so high that it exceeds the highest possible profit per click the channel can get, $\theta v$, the manufacturer would sponsor neither retailer get displayed and earn zero profit. As long as $v_A < \theta v$, the manufacturer will consider sponsoring one or two retailers. When the degree of consumer heterogeneity $t$ is relatively small, one retailer can already cover a large portion of consumers, so the manufacturer can already get a relatively high demand by sponsoring just one retailer. At the same time, if the manufacturer sponsors both retailers to get displayed, the spatial competition between retailers is intense when $t$ is small, which makes the retail price and retailers' profits low. Consequently, the manufacturer needs to provide high participation rates to the two retailers when sponsoring both retailers. Therefore, it is optimal for the manufacturer to sponsor only one retailer when $t$ is small. On the other hand, when the degree of consumer heterogeneity $t$ is relatively large, the manufacturer will lose a significant portion of the market by sponsoring only one retailer. At the same time, with relatively large $t$, the price competition is weak, which provides retailers enough incentive to bid high. Thus it is more profitable for the manufacturer to sponsor both retailers, as it increases the total demand with relatively low sponsorship costs. Combining both sides, it is more profitable for the manufacturer to sponsor one retailer when $t$ is small, and to sponsor both retailers when $t$ is large. For $t$ in the middle range, the benefits from both sides are not large enough to overcome the cost of outbidding the outside advertisers, and thus the manufacturer may optimally sponsor neither retailer, earning zero profit.

One thing to notice is that when $\frac{\theta}{2}(v + w) \leq v_A < \theta v$, $\bar{t} < t_1$ always holds, and when $\frac{3\theta}{8}(v - w) < v_A < \frac{\theta}{2}(v + w)$, $\bar{t} < t_2$ always holds. This implies that as long as $v_A < \theta v$, there always exists the case that it is optimal for the manufacturer to sponsor only one retailer given two symmetric retailers.
4.3 Endogenous Wholesale Contracts and Retail Prices

In this section, we further consider an extension when search advertising is the main source of demand for both manufacturer and retailers, so that the manufacturer has to take search advertising into account when setting wholesale prices. We incorporate endogenous wholesale contracts and retail prices into the basic model, without explicitly considering retail price competition.

Let's consider a three-stage game with the timeline of events shown in Figure 4-4. First, the manufacturer chooses the wholesale contracts and participation rates $\alpha_1, \alpha_2$ for the two retailers. Second, given the manufacturer's choices, the two retailers decide retail prices $p_1, p_2$. Lastly, after observing the retail prices, the two retailers submit bids $b_1, b_2$ in a position auction. By specifying the timeline above, we have made several assumptions, and here are our justifications. First, we notice that both wholesale and co-op contracts require heavy administrative work, and thus cannot be altered frequently. In contrast, online retailers can adjust their prices weekly or even daily, so they can treat the wholesale contracts and participation rates as given when determining the retail prices. Moreover, it is not uncommon for retailers to adjust their bids almost continuously with the help of automatic bidding support systems. Therefore it is reasonable to assume that when choosing their bids, the retailers can treat the retail prices as given.

For the wholesale contracts, we will consider both linear contracts and two-part tariffs below. Generally speaking, two-part tariffs generate a higher channel profit, but linear contracts are simpler to implement and more robust to accommodate against changing environments or incomplete contracts (Villas-Boas 1998). We denote the wholesale prices as $w_i$ for retailer $i$.

For sake of simplicity of the analysis, we assume that the click-through rate only depends on the rank of the position, independent of the retail price and the identity of the product. However, the price will affect how likely a consumer will make a purchase after they click the ad. That is, retail prices will influence the conversion rate. Consumers' valuation of the product is assumed to be uniformly distributed on
Therefore, the conversion rate of retailer $i$ can be written as $\theta_i = \bar{\theta}_i (1 - p_i)$, where $\bar{\theta}_i$ is a constant. Retailer $i$'s profit margin is $r_i = p_i - w_i$, and the manufacturer's profit margin is $m_i = w_i - c$, where $c$ is the marginal production cost. Retailers are \textit{ex ante} symmetric, i.e., $\theta_1 = \theta_2 = \bar{\theta}$.

We solve the equilibrium by backward induction. In fact, in the last stage when the wholesale contracts, participation rates and retail prices have been determined, retailers face the exactly same problem as the basic model above, and we have solved for the retailers' optimal bids given their profit per click. Now we consider the two retailers’ decisions on retail prices. Notice that $p_i$ enter into retailer $i$'s profit function only via $\theta_i r_i$. Retailer $i$'s equivalent profit per click is

$$v_i = \frac{\theta_i r_i}{1 - \alpha_i} = \frac{\bar{\theta}(1 - p_i)(p_i - w_i)}{1 - \alpha_i},$$

which reaches the maximum value $v_i^* = \frac{\bar{\theta}(1 - w_i)^2}{4(1 - \alpha_i)}$ when $p_i = (1 + w_i)/2$, and takes the minimum value of zero when $p_i$ is equal to $w_i$ or 1. Therefore, when choosing the retail price $p_i \in [w_i, 1]$, retailer $i$ is essentially choosing $v_i \in [0, v_i^*]$. Given the two retailers’ choice of retail prices, their positions will be determined by the order of $v_1$, $v_2$, and $v_A$. The following lemma shows that in fact, the bidders’ positions in equilibrium are completely determined by the rank of $v_1^*$, $v_2^*$, and $v_A$ (with proof in the appendix).

**Lemma 3** In equilibrium, retail prices are set at $p_i^* = (1 + w_i)/2$, and the positions of retailer 1, retailer 2, and the outside advertiser are given by the descending order of $v_1^*$, $v_2^*$, and $v_A$.

**Linear Wholesale Contracts**

Given the retailers’ retail prices and bids, now we consider the manufacturer’s problem. Let’s first consider linear wholesale contracts. By symmetry, without loss of generality, we can assume that retailer 1 takes a higher position than retailer 2. Similar to the basic model, there are three possible position configurations. Under each position configuration, we can write down the manufacturer’s profit function, and
then maximize its profit with respect to $\alpha_1, \alpha_2, w_1, w_2$ subject to the conditions for the position configuration to hold in equilibrium. By comparing the manufacturer’s profit among the three cases, we can get its wholesale prices and participation rates in equilibrium. We relegate the details of calculations in the appendix. The following theorem completely characterizes the equilibrium.

**Theorem 6** Consider a position auction, the participants of which are two ex-ante symmetric retailers supported by a manufacturer as well as an outside advertiser. The manufacturer provides linear wholesale contracts and participation rates to the two retailers. In equilibrium, the manufacturer’s wholesale prices and participation rates are,

\[
\begin{pmatrix}
    w_1^* \\
    w_2^* \\
    \alpha_1^* \\
    \alpha_2^*
\end{pmatrix} = \begin{cases}
    \left( \frac{1+c}{2}, \frac{1+c}{2}, 0, 0 \right)^T \\
    (1 - 2\sqrt{\frac{\vartheta_A}{\theta}}, 1 - 2\sqrt{\frac{\vartheta_A}{\theta}}, 0, 0)^T \\
    \left( \frac{d_1 + d_2 + d_2 \sqrt{\frac{d_1 + d_2}{d_1 + 3d_2}}}{d_1 + 3d_2}, 1, 1 - \frac{\vartheta(1-c)^2d_1^2}{4\vartheta_A(d_1 + d_2)^2}, 0 \right)^T \\
    \left( \frac{2d_1 + d_2}{d_1 + d_2}, \frac{\theta(1-c)^2}{4} \right)^T \\
    \left( \frac{1+c}{2}, 1, 0, 0 \right)^T
\end{cases}
\]

where $(X, Y)$ at the end of each row indicates the equilibrium position configuration under such condition. The retail prices in equilibrium are,

\[
p_i^* = \frac{1 + w_i^*}{2}, \ i = 1, 2.
\]
per click $v_A$, the manufacturer will optimally apply one or both devices to coordinate the channel so as to maximize its profit. We go through the four cases in equation (4.10) together to understand the manufacturer’s optimal wholesaling and advertising strategies. First, when $v_A$ is very low, the two retailers will get the top two positions without the help from the manufacturer. In this case, the manufacturer sets the monopolistic wholesale prices $(1 + c)/2$, and provides zero participation rates. As we increase $v_A$ to the interval in the second case, the manufacturer still wants to keep the two retailers at the top two positions. It will achieve this goal by providing lower wholesale prices and thus higher profit margins for the retailers but still keeping the participation rates at zero. Intuitively, lowering the wholesale prices not only increases the retailers’ profit margins and thus helps the retailers outbid the outside advertiser, but also increases the demand as the retail prices go down; whereas increasing participation rates only has the first effect. This is why the wholesale prices are the manufacturer’s first choice when fighting against downstream competition in bidding. As we further increase $v_A$ to the third case, wholesale prices alone do not suffice to grant the retailers the winners of the auction. The manufacturer will set a low wholesale price and at the same time provide a positive participation rate to retailer 1 so as to keep it at the first position. It will entirely drop retailer 2 by not selling to it, and as a result retailer 2 will take the third position with zero demand. Lastly, when $v_A$ is very high, the manufacturer has to give up winning the auction. It will set the monopolistic wholesale price $(1 + c)/2$ again for retailer 1 and provides zero participation rate to it. The manufacturer will not sell to retailer 2, who will take the third position.

Figure 4-5 clearly illustrates the manufacturer’s optimal wholesaling and cooperative advertising strategies as described above. By Lemma 3, the equilibrium retail price $p_i^* = (1 + w_i^*)/2$. The average retail price for all consumers will be $\bar{p} = (d(1)p_1^* + d(2)p_2^*)/(d(1) + d(2))$, where $d(i)$ denotes the CTR for retailer $i$ given its position. According to equation (4.10), it is straightforward to show that $\bar{p} = p_1^* = (1 + w_1^*)/2$. Therefore, the relationship between $\bar{p}$ and $v_A$ will be very similar to the relationship between $w_1^*$ and $v_A$ in Figure 4-5. As the outside advertiser’s profit per click increases,
the average retail price first decreases then increases.

Two-Part Tariffs

Now suppose the manufacturer uses two-part tariff wholesale contracts for channel coordination. Each retailer $i$ pays the manufacturer $w_i$ for each product, as well as a fixed franchise fee. Under two-part tariffs, the retailers' problem is the same as before, with their positions determined by the order of $v_1^*, v_2^*$ and $v_A$; however, the manufacturer's objective now is to maximize the total channel profit by choosing $w_1$, $w_2$, $\alpha_1$ and $\alpha_2$. It uses the franchise fee to divide the channel profit with retailers. Similarly, we analyze the equilibrium given each of the three position configurations in the appendix. By comparing the manufacturer's profits among the three position configurations, we get its equilibrium wholesale prices and participation rates. The following theorem characterizes the equilibrium.

**Theorem 7** Suppose a manufacturer's two ex-ante symmetric retailers and one outside advertiser participates in the position auction. The manufacturer provides two-part tariff wholesale contracts and linear participation rates to the two retailers. In
equilibrium, the manufacturer's wholesale prices and participation rates are,

\[
\begin{align*}
\left( \begin{array}{c} w_1^* \\ w_2^* \\ \alpha_1^* \\ \alpha_2^* \\ \end{array} \right) &= \begin{cases} 
\left( c, \left( 1 - \frac{d_1}{d_t} \right) + \frac{d_2}{d_t} c, 0, 0 \right)^T, & \text{when } v_A \leq \frac{d_2^2}{4d_t^2} \frac{\ddot{\theta} (1-c)^2}{4}, \ (R_1, R_2), \\
\left( c, 1 - 2 \sqrt{\frac{v_A}{d_t}}, 0, 0 \right)^T, & \text{when } \frac{d_2^2}{4d_t^2} \frac{\ddot{\theta} (1-c)^2}{4} \leq v_A \leq \frac{\ddot{\theta} (1-c)^2}{9}, \ (R_1, R_2), \quad \text{if } \frac{d_1}{d_2} \geq \frac{3}{2}. 
\end{cases} \quad (4.12)
\end{align*}
\]

\[
\begin{align*}
\left( \begin{array}{c} w_1^* \\ w_2^* \\ \alpha_1^* \\ \alpha_2^* \\ \end{array} \right) &= \begin{cases} 
\left( c, \left( 1 - \frac{d_1}{d_t} \right) + \frac{d_2}{d_t} c, 0, 0 \right)^T, & \text{when } v_A \leq \frac{d_2^2}{4d_t(3d_2-d_1)} \frac{\ddot{\theta} (1-c)^2}{4}, \ (R_1, R_2), \\
\left( c, 1, 0, 0 \right)^T, & \text{when } \frac{d_2^2}{4d_t(3d_2-d_1)} \frac{\ddot{\theta} (1-c)^2}{4} < v_A \leq \frac{\ddot{\theta} (1-c)^2}{4}, \ (R_1, A), \quad \text{if } \frac{d_1}{d_2} < \frac{3}{2}. 
\end{cases} \quad (4.13)
\end{align*}
\]

The retail prices in equilibrium are,

\[
p_i^* = \frac{1 + w_i^*}{2}, \quad i = 1, 2. \quad (4.14)
\]

Retailer 1, 2, and outsider advertiser's positions in equilibrium are given by the descending order of \( \ddot{\theta} (1 - w_i^*)^2/[4(1 - \alpha_i^*)] \), \( \ddot{\theta} (1 - w_2^*)^2/[4(1 - \alpha_2^*)] \), and \( v_A \).

Compared with linear contracts, we find that cooperative advertising in forms of participation rates is never used to coordinate the channel with two-part tariffs. This is consistent with the general observation that a sufficiently flexible wholesale contract will fully coordinate the channel and there will be no need for a manufacturer to cooperate with retailers separately on advertising. From a different angle, we can also view the lump-sum payment between the manufacturer and the retailers as a form of cooperative search advertising.
From Theorem 7, we find that the manufacturer always sets the wholesale price as the marginal production cost for retailer 1, who takes a higher position than the other retailer. This not only eliminates the double marginalization conflicts in retail prices, but also maximizes the retailer’s profit per click thus chance of winning the auction. In contrast, it takes the manufacturer more deliberations when setting the wholesale price for retailer 2, who takes a lower position than retailer 1. When the outside advertiser’s profit per click is relatively high, the manufacturer will not sell to retailer 2, who ends up in the third position; on the other hand, when the outside advertiser’s profit per click is relatively low, the manufacturer will support retailer 2 to take the second position by providing a wholesale price that is higher than the marginal production cost, in an effort to balance between supporting retailer 2 to outbid the outside advertiser and lowering the price per click for retailer 1.

4.5 Conclusion

To the best of our knowledge, this is the first research that addresses the problem of channel coordination in the context of search advertising. Our work studies cooperative search advertising by considering a simple form of coordination contract—a manufacturer shares a fixed percentage of a retailer’s spending on search advertising. We consider intra-brand coordination and competition among one manufacturer and two retailers, as well as inter-brand competition with outside advertisers. We find that it may not be optimal for the manufacturer to sponsor both retailers even if they are ex ante the same. This reflects the manufacturer’s tradeoff between higher demand and higher bidding cost resulting from intensified competition. We also find that the relative positions of retailers in equilibrium are determined by their channel profit per click. This illustrates the efficiency of this simple coordination mechanism despite that it does not fully coordinate the channel.

In general, our main results carry through the several extensions we consider. First, when the manufacturer can submit bids directly, the relative positions of retailers are still determined by the rank of their channel profit per click, and the
manufacturer should get a higher position itself if its profit per click via direct sales is higher than both retailers' channel profit per click. Second, with endogenous retail price competition, the manufacturer may still optimally sponsor one retailer given two symmetric retailers in the market. Third, with endogenous linear wholesale contracts and retail prices, the manufacturer may still optimally sponsor one retailer; however, it is never optimal to sponsor both retailers and it is no longer necessary to use cooperative advertising with two-part tariffs. Lastly, when click-through rates depend on advertisers' identities, we demonstrate that our main results generalize nicely—now one's total channel profit per impression will determine its position rank.

This work has several limitations. First, we do not make a distinction between branded and generic keywords, which could be a topic for future research. Second, we have not fully modeled consumers' search and click behaviors, and thus cannot evaluate consumers' welfare in the context of cooperative search advertising. Last but not least, we discuss the problem of one focal manufacturer, incorporating the competition in position auction from other brands by considering an outside advertiser, but we do not explicitly model the strategic interaction between manufacturers, especially when they sell to common retailers.
4.A Appendix

4.A.1 Figures

Figure 4-1: An Example of Google Ads on “Laptop”.

![Google Ads Screenshot](image-url)
Figure 4-2: Manufacturer's Optimal Cooperative Search Advertising Strategy, Given Two Symmetric Retailers.

Retailer's Profit per Click $\theta r$

Support Both

Support One

Support None

Manufacturer's Profit per Click $\theta m$

$2v_A$

$\frac{d_2}{d_1}(1+\frac{d_2}{d_1})v_A$

$0$

$(1-\frac{d_2}{d_1})v_A$

$0$

$0$

$v_A$

$v_A$
Figure 4-3: Manufacturer’s Optimal Advertising and Cooperative Search Advertising Strategies, Given Two Symmetric Retailers and Equal Channel Profit.

Figure 4-4: Timeline of Events with Endogenous Wholesale Contracts and Retail Prices.

Manufacturer chooses wholesale contracts and participation rates $\alpha_1, \alpha_2$. Retailers choose bids $b_1$ and $b_2$ respectively.

Retailers choose prices $p_1$ and $p_2$ respectively. Positions in the auction realize; consumer clicks and purchases realize.
Figure 4-5: Manufacturer’s Optimal Wholesaling and Cooperative Advertising Strategies under Linear Wholesale Contracts.
4.A.2 Proof of Theorem 2

We first consider the three cases in which retailer 1 gets a higher position than retailer 2.

- Consider the position configuration \((R_1, R_2)\).

We have lay out the equilibrium analysis for this case in the main text. The manufacturer's optimal participation rates are

\[
\alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A},
\]
\[
\alpha_2^* = 1 - \frac{\theta_2 r_2}{v_A}.
\]

The manufacturer's optimal profit in this case is

\[
\pi_M(\alpha_1^*, \alpha_2^*) = d_1 [\theta_1 (m_1 + r_1) - v_A] + d_2 [\theta_2 (m_2 + r_2) - v_A].
\] (4.A15)

Accordingly, profits of the two retailers and the channel under \(\alpha_1^*\) and \(\alpha_2^*\) are

\[
\pi_{R_1}(\alpha_1^*, \alpha_2^*) = 0,
\]
\[
\pi_{R_2}(\alpha_1^*, \alpha_2^*) = 0,
\]
\[
\pi_C(\alpha_1^*, \alpha_2^*) = d_1 [\theta_1 (m_1 + r_1) - v_A] + d_2 [\theta_2 (m_2 + r_2) - v_A].
\]

- Consider the position configuration \((R_1, A)\).

According to Lemma 1, we need to have \(\theta_1 r_1/(1 - \alpha_1) \geq v_A > \theta_2 r_2/(1 - \alpha_2)\) for this position configuration to be the equilibrium. Retailer 2 will bid its equivalent value \(\theta_2 r_2/(1 - \alpha_2)\), and the outside advertiser's equilibrium bid is \(\frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \theta_2 r_2\).

The manufacturer's profit will be

\[
\pi_M(\alpha_1, \alpha_2) = d_1 \left[ \theta_1 m_1 - \alpha_1 \left( \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \theta_2 r_2 \right) \right],
\]

which decreases in both \(\alpha_1\) and \(\alpha_2\). Therefore, the manufacturer will choose the
smallest $\alpha_1$ and $\alpha_2$ that ensure $\theta_1 r_1/(1 - \alpha_1) > v_A > \theta_2 r_2/(1 - \alpha_2)$, which are

$$\alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A}, \quad \alpha_2^* = 0.$$  

The manufacturer’s profit under the optimal participation rates is

$$\pi_M(\alpha_1^*, \alpha_2^*) = d_1 \left[ \theta_1 (m_1 + r_1) - v_A \right] + d_2 \left( \frac{v_A - \theta_1 r_1 (v_A - \theta_2 r_2)}{v_A} \right). \quad (4.16)$$

Accordingly, profits of the two retailers and the channel under $\alpha_1^*$ and $\alpha_2^*$ are

$$\pi_{R_1}(\alpha_1^*, \alpha_2^*) = d_1 \left[ \theta_1 r_1 - (1 - \alpha_1^*) \left( \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \frac{\theta_2 r_2}{1 - \alpha_2^*} \right) \right] = d_2 \theta_1 r_1 \left( 1 - \frac{\theta_2 r_2}{v_A} \right),$$

$$\pi_{R_2}(\alpha_1^*, \alpha_2^*) = 0,$$

$$\pi_C(\alpha_1^*, \alpha_2^*) = d_1 \left[ \theta_1 (m_1 + r_1) - v_A \right] + d_2 \left[ v_A - \theta_2 r_2 \right].$$

- Consider the position configuration $(A, R_1)$.

According to Lemma 1, we need to have $v_A > \theta_1 r_1/(1 - \alpha_1) \geq \theta_2 r_2/(1 - \alpha_2)$ for this position configuration to be the equilibrium. Retailer 2’s bid at position 3 will be its equivalent profit per click, $\theta_2 r_2/(1 - \alpha_2)$, and this is the cost per click for retailer 1. The manufacturer’s profit will be

$$\pi_M(\alpha_1, \alpha_2) = d_2 \left( \theta_1 m_1 - \alpha_1 \frac{\theta_2 r_2}{1 - \alpha_2} \right),$$

which decreases in $\alpha_1$ and $\alpha_2$. Therefore, the manufacturer will choose the smallest $\alpha_1$ and $\alpha_2$ that satisfy $v_A > \theta_1 r_1/(1 - \alpha_1) > \theta_2 r_2/(1 - \alpha_2)$. The optimal participation rates are

$$\alpha_1^* = \max \left\{ 1 - \frac{\theta_1 r_1}{\theta_2 r_2}, 0 \right\},$$

$$\alpha_2^* = 0.$$
The manufacturer's profit under the optimal participation rates is

\[ \pi_M(\alpha_1^*, \alpha_2^*) = d_2 [\theta_1(m_1 + r_1) - \max\{\theta_1r_1, \theta_2r_2\}] . \]  \hspace{1cm} (4.A17)

Accordingly, profits of the two retailers and the channel under \( \alpha_1^* \) and \( \alpha_2^* \) are,

\[ \pi_{R_1}(\alpha_1^*, \alpha_2^*) = d_2 \left[ \theta_1r_1 - (1 - \alpha_1^*) \frac{\theta_2r_2}{1 - \alpha_2^*} \right] = \max\{\theta_1r_1 - \theta_2r_2, 0\} , \]

\[ \pi_{R_2}(\alpha_1^*, \alpha_2^*) = 0 , \]

\[ \pi_C(\alpha_1^*, \alpha_2^*) = \pi_M(\alpha_1^*, \alpha_2^*) + \pi_{R_1}(\alpha_1^*, \alpha_2^*) + \pi_{R_2}(\alpha_1^*, \alpha_2^*) = d_2 [\theta_1(m_1 + r_1) - \theta_2r_2] . \]

So far, we have analyzed the three position configurations in which retailer 1 gets a higher position than retailer 2. We find that given \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \), in each case, the manufacturer's profit will get lower if retailer 1 and retailer 2 are exchanged. That is, retailer 1 will always take a higher position than retailer 2 in equilibrium.

In order to know which position configuration will be chosen by the manufacturer in equilibrium, we only need to compare the manufacturer's profits under the three cases—equations (4.A15), (4.A16), and (4.A17).

**4.A.3 Analysis of the Case with \( v_A \) between \( \theta_1r_1 \) and \( \theta_2r_2 \)**

In the following analysis, we do not presume that \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \). Without loss of generality, suppose \( \theta_1r_1 \geq v_A \geq \theta_2r_2 \). Without any support from the manufacturer, retailer 1 will get position 1, the outside advertiser A will get position 2, and retailer 2 will get position 3. The manufacturer's profit only comes from retailer 1, and it equals to,

\[ \pi_M(0, 0) = d_1\theta_1m_1 . \]

Obviously, the manufacturer has no incentive to give retailer 1 any support, since retailer 1 can already get the first position and thus the manufacturer can get the maximum demand from retailer 1 without paying anything for the clicks. Now let's see whether the manufacturer would like to support retailer 2 to move up.
If the manufacturer moves retailer 2 up to position 2, it needs to provide participation rate $\alpha_2$ to retailer 2 such that

$$\theta_1 r_1 \geq \frac{\theta_2 r_2}{1 - \alpha_2} \geq v_A. \tag{4.18}$$

The outside advertiser’s bid at position 3 is $v_A$, and retailer 2’s bid at position 2 is

$$(d_1 - d_2) / d_1 \cdot \theta_2 r_2 / (1 - \alpha_2) + d_2 / d_1 \cdot v_A,$$

which is the price per click for retailer 1. The manufacturer’s profit is then,

$$\pi_M(0, \alpha_2) = d_1 \theta_1 m_1 + d_2 (\theta_2 m_2 - \alpha_2 v_A),$$

which decreases in $\alpha_2$, so the manufacturer will choose the smallest $\alpha_2$ that satisfies (4.18), $\alpha_2^* = 1 - \theta_2 r_2 / v_A$. Correspondingly, the manufacturer’s profit is

$$\pi_M(0, \alpha_2^*) = d_1 \theta_1 m_1 + d_2 [\theta_2 (m_2 + r_2) - v_A].$$

If the manufacturer further supports retailer 2 to move up to position 1, it needs to provide participation rate $\alpha_2$ to retailer 2 such that

$$\frac{\theta_2 r_2}{1 - \alpha_2} > \frac{\theta_2 r_2}{\alpha_2} \geq v_A. \tag{4.19}$$

The outside advertiser’s bid at position 3 is $v_A$, and retailer 1’s bid at position 2 is

$$(d_1 - d_2) / d_1 \cdot \theta_1 r_1 + d_2 / d_1 \cdot v_A.$$

Similarly, the manufacturer will choose the smallest $\alpha_2$ that satisfies (4.19), so $\alpha_2^* = 1 - \theta_2 r_2 / (\theta_1 r_1)$. The manufacturer’s profit is then,

$$\pi_M(0, \alpha_2^*) = d_1 [\theta_2 m_2 - \alpha_2 \left(\frac{d_1 - d_2}{d_1} \theta_1 r_1 + \frac{d_2}{d_1} v_A\right)] + d_2 \theta_1 m_1$$

$$= d_1 [\theta_2 (m_2 + r_2) - \theta_1 r_1] + d_2 \left[\theta_1 (m_1 + r_1) - \theta_2 r_2 - v_A \left(1 - \frac{\theta_2 r_2}{\theta_1 r_1}\right)\right].$$

Comparing the manufacturer’s profits in the three scenarios, we can see that the
manufacturer will support retailer 2 to get position 2 if and only if,

\[ \theta_1(m_1 + r_1) + \frac{d_2}{d_1 - d_2} \frac{\theta_1 r_1 - v_A}{\theta_1 r_1} \theta_2 r_2 \geq \theta_2(m_2 + r_2) \geq v_A, \]

and the manufacturer will support retailer 2 to get position 1 if and only if,

\[ \theta_2(m_2 + r_2) > \theta_1(m_1 + r_1) + \frac{d_2}{d_1 - d_2} \frac{\theta_1 r_1 - v_A}{\theta_1 r_1} \theta_2 r_2. \]

To summarize, when \( v_A \) is between \( \theta_1 r_1 \) and \( \theta_2 r_2 \), the retailer with higher total channel profit does not necessarily get a higher position. Specifically, given \( \theta_1 r_1 \geq v_A \geq \theta_2 r_2 \), retailer 1 will get a higher position when \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \); but retailer 2 may not get a higher position when \( \theta_2(m_2 + r_2) > \theta_1(m_1 + r_1) \). In other words, we need a stricter condition than \( \theta_2(m_2 + r_2) > \theta_1(m_1 + r_1) \), as shown above, to grant retailer 2 a higher position than retailer 1.

The reason is that, in the case that \( v_A \) is between \( \theta_1 r_1 \) and \( \theta_2 r_2 \), the manufacturer can choose the participation rates such that the bid at position 2 equals to \( v_A \) and both retailers pay \( v_A \) per click, when the two retailers take the top two positions. Now, consider the case that \( \theta_1 r_1 \geq v_A \). When retailer 1's takes position 2 and retailer 2 takes position 1, retailer 1's bid is a linear combination of \( v_A \) and \( \theta_1 r_1 \), which is greater than \( v_A \). As a result, the price per click at position 1 is higher than \( v_A \). On the other hand, when retailer 1 takes position 1 and retailer 2 takes position 2, the manufacturer can choose retailer 2's participation rate such that the bid at position 2 equals to \( v_A \) and both retailers pay \( v_A \) per click. Therefore, having retailer 2 at position 1 is more costly and thus demand a stronger condition to ensure it as the equilibrium.

4.A.4 Proof of Proposition 6

**Step 1**

We will first show that for the integrated channel, the retailer with a higher channel profit per click \( \theta_i(m_i + r_i) \) gets a higher position than the other retailer.
To prove this, we write down the integrated channel's profit under each position configuration given the retailers' and advertiser's bids.

\[
\pi^{(R_1,R_2)}(b_1, b_2, b_A) = d_1 [\theta_1(m_1 + r_1) - b_2] + d_2 [\theta_2(m_2 + r_2) - b_A],
\]

\[
\pi^{(R_1,A)}(b_1, b_2, b_A) = d_1 [\theta_1(m_1 + r_1) - b_A],
\]

\[
\pi^{(A,R_1)}(b_1, b_2, b_A) = d_2 [\theta_1(m_1 + r_1) - b_2],
\]

where the superscripts denote the position configurations. We have only written down the profit functions for the three cases in which retailer 1 gets a higher position than retailer 2. The profit functions for the other three cases can be obtained by symmetry. It is straightforward to verify that when \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \),

\[
\pi^{(R_1,R_2)}(b_1, b_2, b_A) \geq \pi^{(R_2,R_1)}(b_1, b_2, b_A),
\]

\[
\pi^{(R_1,A)}(b_1, b_2, b_A) \geq \pi^{(R_2,A)}(b_1, b_2, b_A),
\]

\[
\pi^{(A,R_1)}(b_1, b_2, b_A) \geq \pi^{(A,R_2)}(b_1, b_2, b_A)
\]

for any \( b_1, b_2 \) and \( b_A \). This implies that given \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \), retailer 1 must get a higher position than retailer 2 in equilibrium and the only possible equilibrium position configurations are \((R_1,R_2)\), \((R_1,A)\), and \((A,R_1)\).

**Step 2**

Now we assume \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \) without loss of generalizability. For each of the three position configuration in which retailer 1 gets a higher position than retailer 2, we will work out the parameter space (i.e., requirement on the relationship between \( v_A, \theta_1(m_1 + r_1) \), and \( \theta_2(m_2 + r_2) \)) where the position configuration can be an equilibrium, the channel’s optimal choice of bids in the parameter space and the channel’s optimal profit.

- Consider the position configuration \((R_1,R_2)\).

In order for this position configuration to be an equilibrium, the bids should satisfy

\[
b_1 \geq b_2 \geq b_A. \quad (4.A20)
\]

According to SNE and LB that guard against advertiser A’s deviation to position 2, the outside advertiser A’s equilibrium bid should be \( b_A = v_A \). Then (4.A20) implies
\( b_1 \geq v_A \), and thus it is not profitable for advertiser \( A \) to deviate to position 1.

To prevent the channel deviating from \((R_1, R_2)\) to \((R_1, A)\), the bids should satisfy
\[
d_1[\theta_1(m_1 + r_1) - b_2] + d_2[\theta_2(m_2 + r_2) - b_A] \geq d_1[\theta_1(m_1 + r_1) - b_A].
\]
Given \( b_A = v_A \), this is equivalent to
\[
\theta_2(m_2 + r_2) \geq \frac{d_1}{d_2}(b_2 - v_A) + v_A. \tag{4.A21}
\]

To prevent the channel deviating from \((R_1, R_2)\) to \((A, R_1)\), the bids should satisfy
\[
d_1[\theta_1(m_1 + r_1) - b_2] + d_2[\theta_2(m_2 + r_2) - b_A] \geq d_2[\theta_1(m_1 + r_1) - b_A],
\]
which is equivalent to
\[
\frac{d_1 - d_2}{d_1}\theta_1(m_1 + r_1) + \frac{d_2}{d_1}\theta_2(m_2 + r_2) \geq b_2. \tag{4.A22}
\]

The channel profit is
\[
\pi_C(b_1, b_2) = d_1[\theta_1(m_1 + r_1) - b_2] + d_2[\theta_2(m_2 + r_2) - v_A]. \tag{4.A23}
\]

According to (4.A23), the channel profit decreases with \( b_2 \) at this position configuration, and the two NE conditions (4.A21) and (4.A22) cover a larger parameter space when \( b_2 \) is smaller, so the channel should choose the smallest possible \( b_2 \) that satisfy (4.A20), i.e., \( b_2^* = v_A \). Retailer 1’s equilibrium bid \( b_1^* \) can be any value higher than or equal to \( v_A \). The channel profit under configuration \((R_1, R_2)\) is
\[
\pi_C(b_1^*, b_2^*) = d_1[\theta_1(m_1 + r_1) - v_A] + d_2[\theta_2(m_2 + r_2) - v_A],
\]
and the equilibrium can exist when \( \theta_2(m_2 + r_2) \geq v_A \). ((4.A22) is automatically satisfied given this and \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \)).

- Consider the position configuration \((R_1, A)\).

In order for this position configuration to be an equilibrium, the bids should satisfy \( b_1 \geq b_A > b_2 \).

Given any \( b_2 \), to prevent the outside advertiser \( A \) deviating from position 2 to position 1, advertiser \( A \)’s bid in equilibrium is \( b_A = \frac{d_1 - d_2}{d_1}v_A + \frac{d_2}{d_1}b_2 \) according to SNE and LB.

To prevent advertiser \( A \) deviating from position 2 to position 3, we need \( d_2(v_A -
\( b_2 > 0 \), i.e., \( b_2 < v_A \).

To prevent the channel deviating from \((R_1, A)\) to \((R_1, R_2)\), we need \( d_1[\theta_1(m_1 + r_1) - b_A] > d_1[\theta_1(m_1 + r_1) - b_A] + d_2[\theta_2(m_2 + r_2) - b_A] \), which is equivalent to

\[
b_A > \theta_2(m_2 + r_2). \tag{4.A24}
\]

To prevent the channel deviating from \((R_1, A)\) to \((A, R_1)\), we need \( d_1[\theta_1(m_1 + r_1) - b_A] > d_2[\theta_1(m_1 + r_1) - b_2] \). Given the expression of \( b_A \), this is equivalent to

\[
\theta_1(m_1 + r_1) \geq v_A.
\]

The channel profit is

\[
\pi_C(b_1, b_2) = d_1[\theta_1(m_1 + r_1) - b_A] = d_1 \left[ \theta_1(m_1 + r_1) - \frac{d_1 - d_2}{d_1} v_A - \frac{d_2}{d_1} b_2 \right]. \tag{4.A25}
\]

According to (4.A25), the channel profit decreases in \( b_2 \) given the position configuration. Thus retailer 2 will choose the lowest possible bid that satisfy the equilibrium conditions. Notice the requirement of (4.A24) and \( b_A = \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} b_2 \), when \( \theta_2(m_2 + r_2) < \frac{d_1 - d_2}{d_1} v_A \), retailer 2 can choose \( b_2^* = 0 \) in equilibrium; however, when \( \theta_2(m_2 + r_2) > \frac{d_1 - d_2}{d_1} v_A \), in order to sustain the equilibrium, retailer 2 has to choose \( b_2^* \) such that \( \theta_2(m_2 + r_2) = \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} b_2^* \).

The channel profit under position configuration \((R_1, A)\) is

\[
\pi_C(b_1^*, b_2^*) = d_1 \left[ \theta_1(m_1 + r_1) - \max \left\{ \frac{d_1 - d_2}{d_1} v_A, \theta_2(m_2 + r_2) \right\} \right],
\]

and the NE holds when \( \theta_1(m_1 + r_1) \geq v_A > \theta_2(m_2 + r_2) \) (the latter inequality comes from \( b_2 < v_A \)).

- Consider the position configuration \((A, R_1)\).

In order for this position configuration to be an equilibrium, the bids should satisfy

\[ b_A > b_1 \geq b_2. \]

\[ ^{17} \text{In other words, in order for } (R_1, A) \text{ to be an equilibrium, advertiser A has to bid no less than } \theta_2(m_2 + r_2). \]

168
To prevent advertiser A deviating from position 1 to position 2, we need \(d_1(v_A - b_1) > d_2(v_A - b_2)\), which is equivalent to \(b_1 < \frac{d_1 - d_2 v_A}{d_1} + \frac{d_2}{d_1} b_2\). To prevent advertiser A deviating from position 1 to position 3, we need \(d_1(v_A - b_1) > 0\), which is equivalent to \(b_1 < v_A\).

To prevent the channel deviating from \((A, R_1)\) to \((R_1, R_2)\), we need \(d_2[\theta_1(m_1 + r_1) - b_2] > d_1[\theta_1(m_1 + r_1) - b_A] + d_2[\theta_2(m_2 + r_2) - b_A]\). To prevent the channel deviating from \((A, R_1)\) to \((R_1, A)\), we need \(d_2[\theta_1(m_1 + r_1) - b_2] > d_1[\theta_1(m_1 + r_1) - b_A]\). Given any \(b_2\), these two conditions can be satisfied when \(b_A\) is sufficiently large.

The channel profit is

\[\pi_C(b_1, b_2) = d_2[\theta_1(m_1 + r_1) - b_2]\]  \hspace{1cm} (4.A26)

According to (4.A26), the channel profit decreases in \(b_2\) at this position configuration. Thus retailer 2 will choose the lowest possible bid in equilibrium, i.e., \(b_2^* = 0\). Retailer 1’s equilibrium bid \(b_1^*\) can be any value between 0 and \(\frac{d_1 - d_2 v_A}{d_1}\). The channel profit under position configuration \((R_1, A)\) is

\[\pi_C(b_1^*, b_2^*) = d_2\theta_1(m_1 + r_1),\]

and this equilibrium can exist on the entire parameter space given \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\).

**Step 3**

So far, for each of the three position configuration, we have worked out the largest possible parameter space on which the position configuration can be an equilibrium, the channel’s optimal choice of bids in the parameter space and the channel’s optimal profit. Table 4.A1 summarizes the Nash equilibrium condition for each position configuration and the integrated channel’s corresponding profit.

Then we can divide the parameter space of \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\) into four mutually exclusive parts. In each part, if only one position configuration can exist according to the above analysis, it will be the equilibrium position configuration, and if there are multiple possible position configurations, the channel will choose the one
Table 4.A1: Profits and NE Condition for each Position Configuration – Integrated Channel

<table>
<thead>
<tr>
<th>Position Configuration</th>
<th>( \pi_C )</th>
<th>NE Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>((R_1, R_2))</td>
<td>( d_1 [\theta_1(m_1 + r_1) - v_A] + d_2 [\theta_2(m_2 + r_2) - v_A] )</td>
<td>( \theta_2(m_2 + r_2) \geq v_A )</td>
</tr>
<tr>
<td>((R_1, A))</td>
<td>( d_1 [\theta_1(m_1 + r_1) - \max{\frac{d_1-d_2}{d_1} v_A, \theta_2(m_2 + r_2)}] )</td>
<td>( \theta_1(m_1 + r_1) \geq v_A &gt; \theta_2(m_2 + r_2) )</td>
</tr>
<tr>
<td>((A, R_1))</td>
<td>( d_2 \theta_1(m_1 + r_1) )</td>
<td>any ( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) )</td>
</tr>
</tbody>
</table>

with the highest channel profit in equilibrium.

- **Part 1:** \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \geq v_A \).

  The possible position configurations in equilibrium are \((R_1, R_2)\) and \((A, R_1)\).

  The channel will choose \((A, R_1)\) when \( \theta_1(m_1 + r_1) < \frac{d_1+d_2}{d_1-d_2} v_A - \frac{d_2}{d_1-d_2} \theta_2(m_2 + r_2) \), and will choose \((R_1, R_2)\) otherwise.

- **Part 2:** \( \theta_1(m_1 + r_1) \geq v_A, \theta_2(m_2 + r_2) < \frac{d_1-d_2}{d_1} v_A \).

  The possible position configurations are \((R_1, A)^a\) and \((A, R_1)^a\).\(^{18}\) The channel will always choose \((R_1, A)^a\) in this area.

- **Part 3:** \( \theta_1(m_1 + r_1) \geq v_A, \frac{d_1-d_2}{d_1} v_A \leq \theta_2(m_2 + r_2) < v_A \).

  The possible configurations are \((R_1, A)^b\) and \((A, R_1)\). The channel will choose \((A, R_1)\) when \( \theta_1(m_1 + r_1) < \frac{d_1-d_2}{d_1-d_2} \theta_2(m_2 + r_2) \), and will choose \((R_1, A)^b\) otherwise.

- **Part 4:** \( \theta_2(m_2 + r_2) \leq \theta_1(m_1 + r_1) \leq v_A \).

  The only possible position configuration is \((A, R_1)\).

To summarize, the condition for each position configuration to be chosen is:

- **\((R_1, R_2)\):** \( \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) \geq v_A \) and \( \theta_1(m_1 + r_1) \geq \frac{d_1+d_2}{d_1-d_2} v_A - \frac{d_2}{d_1-d_2} \theta_2(m_2 + r_2) \);

- **\((R_1, A)\):** \( \theta_1(m_1 + r_1) \geq v_A > \theta_2(m_2 + r_2) \) and \( \theta_1(m_1 + r_1) \geq \frac{d_1}{d_1-d_2} \theta_2(m_2 + r_2) \);

\(^{18}\)For the position configuration \((R_1, A)\), we denote it as \((R_1, A)^a\) when \( \theta_2(m_2 + r_2) < \frac{d_1-d_2}{d_1} v_A \), and denote it as \((R_1, A)^b\) when \( \frac{d_1-d_2}{d_1} v_A \leq \theta_2(m_2 + r_2) < v_A \).
Table 4.A2: Profits under each Position Configuration for Integrated Channel and Non-integrated Channel

<table>
<thead>
<tr>
<th></th>
<th>Integrated</th>
<th>Non-integrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>((R_1, R_2))</td>
<td>(d_1 [\theta_1(m_1 + r_1) - v_A] + )</td>
<td>(d_1 [\theta_1(m_1 + r_1) - v_A] + )</td>
</tr>
<tr>
<td></td>
<td>(d_2 [\theta_2(m_2 + r_2) - v_A] )</td>
<td>(d_2 [\theta_2(m_2 + r_2) - v_A] )</td>
</tr>
<tr>
<td>((R_1, A))</td>
<td>(d_1 [\theta_1(m_1 + r_1)] )</td>
<td>(d_1 [\theta_1(m_1 + r_1) - v_A] + )</td>
</tr>
<tr>
<td></td>
<td>(-\max{d_1 \frac{\partial}{\partial r_2} v_A, \theta_2(m_2 + r_2)})</td>
<td>(d_2 (v_A - \theta_1 r_1)(v_A - \theta_2 r_2))</td>
</tr>
<tr>
<td>((A, R_1))</td>
<td>(d_2 \theta_1(m_1 + r_1))</td>
<td>(d_2 \theta_1(m_1 + r_1) - \max{\theta_1 r_1, \theta_2 r_2})</td>
</tr>
</tbody>
</table>

- \((A, R_1)\): \(\theta_2(m_2 + r_2) \leq \theta_1(m_1 + r_1) \leq v_A\), or \(\theta_1(m_1 + r_1) < \min\{\frac{d_1}{d_1 - d_2} \theta_2(m_2 + r_2)\}\).

Step 4

Now we compare the integrated channel’s profit versus the non-integrated channel’s profit.

Table 4.A2 summarizes the integrated channel’s profit under each position configuration, and the manufacturer’s profit and channel profit when the channel is non-integrated.

One observation is that under position configuration \((R_1, R_2)\), the profit of the non-integrated channel and that of the integrated channel are the same. When the channel profits per click for both retailers are relatively high, both the integrated channel and the non-integrated channel will have position configuration \((R_1, R_2)\) in equilibrium, and will earn the same channel profit in equilibrium. This is an evidence of the efficiency of participation rate mechanism.

An important and surprising finding is that the total channel profit could be lower in the integrated case than in the non-integrated case. When \(\frac{d_1}{d_1 - d_2} v_A \leq \theta_2(m_2 + r_2) < v_A\) and \(\theta_1(m_1 + r_1) \geq \frac{d_1}{d_1 - d_2} \theta_2(m_2 + r_2)\), the equilibrium position configuration for the integrated channel is \((R_1, A)\) and the equilibrium channel profit is \(\pi_{\text{inte-C}}^* = d_1 [\theta_1(m_1 + r_1) - \theta_2(m_2 + r_2)]\). Within this parameter space, when \(\theta_2(m_2 + r_2) \in\)
(v_A - \frac{d_2(v_A-\theta_1r_1)(v_A-\theta_2r_2)}{v_A}, v_A), \quad 19

\pi^*_{inte-C} = d_1 \left[ \theta_1(m_1 + r_1) - \theta_2(m_2 + r_2) \right]
< d_1 \left[ \theta_1(m_1 + r_1) - v_A \right] + d_2 \frac{(v_A - \theta_1r_1)(v_A - \theta_2r_2)}{v_A}
= \pi^*_{noninte-M}
< \pi^*_{noninte-M}.

The last inequality comes from the fact that when the channel is non-integrated, the manufacturer chooses the position configuration that maximizes its own profit in equilibrium. Further noticing that at any position configuration, \( \pi^*_{noninte-M} \leq \pi^*_{noninte-C} \), we know the equilibrium position configuration also satisfies \( \pi^*_{noninte-M} \leq \pi^*_{noninte-C} \). Therefore, when \( \theta_2(m_2 + r_2) \in (v_A - \frac{d_2(v_A-\theta_1r_1)(v_A-\theta_2r_2)}{v_A}, v_A) \) and \( \theta_1(m_1 + r_1) \geq \frac{d_1}{d_1-d_2}\theta_2(m_2 + r_2) \), \( \pi^*_{inte-C} < \pi^*_{noninte-C} \), meaning that under such condition, the integrated channel's profit in equilibrium is lower than the non-integrated channel's profit in equilibrium.

4.A.5 Proof of Theorem 3

We solve the game by backward induction. In the second stage, given the manufacturer's participation decision and participation rates, each bidder's (equivalent) profit per click will be determined. Given four bidders competing for two positions, there are a total of \( 4 \times 3 = 12 \) possible position configurations. In four of the twelve position configurations, the manufacturer is above both retailers, and in the other eight, the manufacturer is below one or both retailers. For each position configuration with the manufacturer above both retailers, we determine the bids of the manufacturer, the retailers, and the outside advertiser, and then maximize the manufacturer's profit with respect to participation rates \( \alpha_1, \alpha_2 \), similar to the analysis in the main model; for each position configuration with the manufacturer below one or both retailers, we determine the retailer and outside advertiser's bids first, and then maximize the man-

\(^{19}\)Notice that \( \frac{d_1-d_2}{d_1}v_A < v_A - \frac{d_2(v_A-\theta_1r_1)(v_A-\theta_2r_2)}{v_A} < v_A \), so the set \( (v_A - \frac{d_2(v_A-\theta_1r_1)(v_A-\theta_2r_2)}{v_A}, v_A) \) is a non-empty subset of \((\frac{d_1-d_2}{d_1}v_A, v_A)\).
ufacturer’s profit with respect to $\alpha_1, \alpha_2$ and its bid $b_M$, without imposing the Nash equilibrium condition that guards against the manufacturer’s deviation to other positions. This is because the analysis gets very complicated by taking into account the Nash equilibrium condition directly. Instead, we will show below that this actually does not influence the equilibrium outcome, i.e., the Nash equilibrium conditions for all position configurations are not binding. If the manufacturer’s optimal bid $b^*_M$ is 0, then the manufacturer should commit to quit the bidding in the first stage, because otherwise, the manufacturer may find it profitable to deviate in the second stage, and we cannot safely omit the Nash equilibrium condition. Back to the first stage, the manufacturer compares its profits over the twelve position configurations to determine the equilibrium. The equilibrium condition for one position configuration is that the manufacturer’s profit under this position configuration is greater than that under all other position configurations. We find that when this equilibrium condition holds, the Nash equilibrium condition for this position configuration always holds. This means that, the Nash equilibrium condition is not binding.

**Step 1**

We first consider the four position configurations in which the manufacturer gets a higher position than both retailers: $(M, R_1)$, $(M, R_2)$, $(M, A)$, $(A, M)$.

- Consider the position configuration $(M, R_1)$.

For this position configuration to be the equilibrium, we need to have

$$b_M \geq b_1 \geq b_A, b_2.$$  \hspace{1cm} (4.A27)

The SNE condition and the LB selection rule that guard against the outside advertiser A and $R_2$’s deviation from position 3 to position 2 imply that,

$$b_A = v_A, b_2 = \frac{\theta_2 r_2}{1 - \alpha_2}.$$  \hspace{1cm} (4.A28)

Given $b_1 > b_A, b_2$ and the equation above, we can show that the SNE condition that guards against the retailer’s deviation to position 1 is has been satisfied.

The SNE condition and the LB selection rule that guard against $R_1$’s deviation
from position 2 to position 1 imply that,

\[ b_1 = \frac{d_1 - d_2}{d_1} \frac{\theta_1 r_1}{1 - \alpha} + \frac{d_2}{d_1} \max\{b_A, b_2\}. \]  \hspace{1cm} (4.A29)

The SNE condition that guards against the R1’s deviation to position 3 is automatically satisfied given \( b_1 > b_A, b_2 \) and the equation above.

The manufacturer chooses \( \alpha_1, \alpha_2 \) to maximize its profit \( \pi_M(\alpha_1, \alpha_2) \), subject to (4.A27)-(4.A29), where,

\[
\begin{align*}
\pi_M(\alpha_1, \alpha_2) &= d_1(\theta_0 m_0 - b_1) + d_2(\theta_1 m_1 - \alpha_1 \max\{b_A, b_2\}) \\
&= d_1 \left[ \theta_0 m_0 - \left( \frac{d_1 - d_2}{d_1} \frac{\theta_1 r_1}{1 - \alpha_1} + \frac{d_2}{d_1} \max\{v_A, \frac{\theta_2 r_2}{1 - \alpha_2}\} \right) \right] \\
&\quad + d_2 \left( \theta_1 m_1 - \alpha_1 \max\{v_A, \frac{\theta_2 r_2}{1 - \alpha_2}\} \right),
\end{align*}
\]

which decreases in \( \alpha_1 \) and \( \alpha_2 \). Therefore, the manufacturer will provide no support to \( R_2 (\alpha_2^* = 0) \) and choose the smallest \( \alpha_1 \) that satisfies \( \frac{\theta_1 r_1}{1 - \alpha_1} > v_A \), i.e., \( \alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A} \).

To prevent the manufacturer from deviating to position 2, we need to have

\[ d_1(\theta_0 m_0 - b_1) + d_2(\theta_1 m_1 - \alpha_1 b_A) \geq d_1(\theta_1 m_1 - \alpha_1 b_A) + d_2(\theta_0 m_0 - b_A), \]

which is equivalent to \( \theta_0 m_0 \geq \theta_1 (m_1 + r_1) \) given the choice of \( \alpha_1 \).

To prevent the manufacturer from deviating to position 3, we need to have

\[ d_1(\theta_0 m_0 - b_1) + d_2(\theta_1 m_1 - \alpha_1 b_A) \geq 0, \]

which is equivalent to \( d_1(\theta_0 m_0 - v_A) + d_2(\theta_1 (m_1 + r_1) - v_A) \geq 0 \) given the choice of \( \alpha_1 \).

To summarize, the optimal participation rates and manufacturer’s profit in the
the case of \((M, R_1)\) are

\[
\begin{align*}
\alpha_1^* &= 1 - \frac{\theta_1 r_1}{v_A}, \\
\alpha_2^* &= 0, \\
\pi_M(\alpha_1^*, \alpha_2^*) &= d_1(\theta_0 m_0 - v_A) + d_2[\theta_1(m_1 + r_1) - v_A],
\end{align*}
\]

(4.A30)

and the equilibrium exists when the parameters satisfy \(\theta_0 m_0 \geq \theta_1(m_1 + r_1)\) and

\[
d_1(\theta_0 m_0 - v_A) + d_2(\theta_1(\theta_0 m_0 - v_A) + d_2[\theta_1(m_1 + r_1) - v_A]) \geq 0.
\]

The manufacturer’s optimal profit in the case of \((M, R_2)\) can be obtained by exchanging \(R_1\) and \(R_2\) in (4.A30), i.e., \(\pi_M(\alpha_1^*, \alpha_2^*) = d_1(\theta_0 m_0 - v_A) + d_2[\theta_2(m_2 + r_2) - v_A]\), which is smaller than its optimal profit under \((M, R_1)\) given \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\). Therefore, \((M, R_2)\) cannot be the position configuration in equilibrium.

- Consider the position configuration \((M, A)\).

For the position configuration to be the equilibrium, we need to have

\[
b_M \geq b_A > \max\{b_1, b_2\}.
\]

(4.A31)

Similarly as above, the SNE and LB conditions that guard against \(R_1\) and \(R_2\)’s deviations from position 3 to a higher position will lead to

\[
b_1 = \frac{\theta_1 r_1}{1 - \alpha_1}, \quad b_2 = \frac{\theta_2 r_2}{1 - \alpha_2}.
\]

The SNE condition and the LB selection rule that guard against the outside advertiser A’s deviation from position 2 to position 1 imply that,

\[
b_A = \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \max \left\{ \frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2} \right\}.
\]

(4.A32)

Given this equation and \(b_A > \max \left\{ \frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2} \right\}\), the SNE condition that guards against the outside advertiser A’s deviation to position 3 has been satisfied.

To prevent the manufacturer from deviating to position 2, we need to have

\[
d_1(\theta_0 m_0 - b_A) \geq d_2(\theta_0 m_0 - \max\{b_1, b_2\}).
\]
Given the expression of $b_A$ as in (4.A32), this is equivalent to $\theta_0 m_0 \geq v_A$.\(^{20}\)

The manufacturer’s profit is

$$\pi_M(\alpha_1, \alpha_2) = d_1(\theta_0 m_0 - b_A)$$

$$= d_1(\theta_0 m_0 - v_A) + d_2\left(v_A - \max\left\{\frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2}\right\}\right),$$

which decreases in $\alpha_1, \alpha_2$. Therefore, the manufacturer will choose $\alpha_1^* = 0, \alpha_2^* = 0$, which satisfies (4.A31) and maximizes its profit.

To prevent the manufacturer from deviating to position 3, we need to have

$$d_1(\theta_0 m_0 - b_A) \geq 0,$$

which is equivalent to $\theta_0 m_0 \geq \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \max\{\theta_1 r_1, \theta_2 r_2\}$ and is satisfied given $\theta_0 m_0 \geq v_A$ and $v_A > \theta_1 r_1, \theta_2 r_2$.

To summarize, the optimal participation rates and manufacturer’s profit in the case of $(M, A)$ are

$$\alpha_1^* = 0,$$

$$\alpha_2^* = 0,$$

$$\pi_M(\alpha_1^*, \alpha_2^*) = d_1(\theta_0 m_0 - v_A) + d_2\left(v_A - \max\{\theta_1 r_1, \theta_2 r_2\}\right),$$

and the equilibrium exists when the parameters satisfy $\theta_0 m_0 \geq v_A$.

- Consider the position configuration $(A, M)$.

For the position configuration to be the equilibrium, we need to have

$$b_A > b_M \geq \max\{b_1, b_2\}.$$  \hspace{1cm} (4.A33)

Similarly as above, the SNE and LB conditions that guard against $R_1$ and $R_2$’s deviations from position 3 to a higher position will lead to $b_1 = \frac{\theta_1 r_1}{1 - \alpha_1}, b_2 = \frac{\theta_2 r_2}{1 - \alpha_2}$.

The SNE condition and the LB selection rule that guard against the outside

\(^{20}\)This requirement is independent of the choice $\alpha_1, \alpha_2$.\]
advertiser M's deviation from position 2 to position 1 imply that,

\[ b_M = \frac{d_1 - d_2}{d_1} \theta_0 m_0 + \frac{d_2}{d_1} \max \left\{ \frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2} \right\}. \]  

(4.A34)

Given this equation and \( b_M > \max \left\{ \frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2} \right\} \), the SNE condition that guards against the manufacturer's deviation to position 3 has been satisfied.

To prevent advertiser A from deviating to the second position, we need to have

\[ d_1 (v_A - b_M) > d_2 (v_A - \max \{b_1, b_2\}), \]

which is equivalent to \( v_A > \theta_0 m_0 \). Similarly as above, the condition of preventing A from deviating to the third position is automatically satisfied given this and \( v_A > \theta_1 r_1, \theta_2 r_2 \).

The manufacturer's profit is

\[ \pi_M(\alpha_1, \alpha_2) = d_2 \left( \theta_0 m_0 - \max \left\{ \frac{\theta_1 r_1}{1 - \alpha_1}, \frac{\theta_2 r_2}{1 - \alpha_2} \right\} \right) \]

which decreases in \( \alpha_1, \alpha_2 \). Therefore, the manufacturer will choose \( \alpha_1^* = 0, \alpha_2^* = 0 \), which satisfies (4.A33) and maximizes its profit.

To summarize, the optimal participation rates and manufacturer's profit in the case of \((A, M)\) are

\[ \alpha_1^* = 0, \]
\[ \alpha_2^* = 0, \]
\[ \pi_M(\alpha_1^*, \alpha_2^*) = d_2 (\theta_0 m_0 - \max \{\theta_1 r_1, \theta_2 r_2\}), \]

and the equilibrium exists when the parameters satisfy \( v_A > \theta_0 m_0 \).

**Step 2**

Now we consider the eight position configurations in which the manufacturer gets a lower position than one or both retailers: \((R_1, M), (R_2, M), (R_1, R_2), (R_2, R_1), (R_1, A), (R_2, A), (A, R_1), (A, R_2)\).
Consider the position configuration \((R_1, M)\).

For this position configuration to be the equilibrium, we need to have

\[ b_1 > b_M > b_A, b_2. \]  
\[ (4.A35) \]

The SNE condition and the LB selection rule that guard against the outside advertiser \(A\) and \(R_2\)'s deviation from position 3 to position 2 imply that,

\[ b_A = v_A, b_2 = \frac{\theta_2 r_2}{1 - \alpha_2}. \]  
\[ (4.A36) \]

Given \(b_M > b_A, b_2\) and the equation above, we can show that the SNE condition that guards against advertiser \(A\) and \(R_2\)'s deviation to position 1 is has been satisfied.

Given \(b_A\) and \(b_M\), the NE condition that guards against \(R_1\)'s deviation from position 1 to position 2 imply that,

\[ d_1[\theta_1 r_1 - (1 - \alpha_1)b_M] > d_2[\theta_1 r_1 - (1 - \alpha_1)b_A]. \]  
\[ (4.A37) \]

Given \(b_M > b_A\) and the inequality above, the SNE condition that guards against \(R_1\)'s deviation to position 3 must be satisfied.

The manufacturer chooses \(\alpha_1, \alpha_2\) and \(b_M\) to maximize its profit \(\pi_M(\alpha_1, \alpha_2, b_M)\), subject to conditions \((4.A35)-(4.A37)\), where,

\[
\pi_M(\alpha_1, \alpha_2, b_M) = d_1(\theta_1 m_1 - \alpha_1 b_M) + d_2(\theta_0 m_0 - \max\{b_A, b_2\})
= d_1(\theta_1 m_1 - \alpha_1 b_M) + d_2\left(\theta_0 m_0 - \max\{v_A, \frac{\theta_2 r_2}{1 - \alpha_2}\}\right),
\]
\[ (4.A38) \]

which decreases with \(\alpha_1, \alpha_2\) and \(b_M\). Thus the manufacturer will choose \(\alpha_2^* = 0\). The smallest bid \(b_M\) that satisfies \((4.A35)\) is \(b_M = v_A\). Notice that the bound of \(\alpha_1\) given by \((4.A37)\) decreases with \(b_M\), so the smallest possible \(\alpha_1\) that satisfies \((4.A37)\) is \(\alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A}\).

To prevent the manufacturer from deviating to the first position, we need to have

\[ d_1(\theta_1 m_1 - \alpha_1 b_M) + d_2(\theta_0 m_0 - v_A) > d_1(\theta_0 m_0 - b_1) + d_2(\theta_1 m_1 - \alpha_1 v_A), \]
which can be
satisfied as long as $b_1$ is large enough.  

To prevent the manufacturer from deviating to the third position, we need to have 
\[ d_1(\theta_1m_1 - \alpha_1b_M) + d_2(\theta_0m_0 - v_A) \geq d_1(\theta_1m_1 - \alpha_1b_A), \]
which is equivalent to $\theta_0m_0 \geq v_A$.

To summarize, the solution to this optimization problem is that,
\[
\begin{align*}
\alpha_1^* &= 1 - \frac{\theta_1r_1}{v_A}, \\
\alpha_2^* &= 0, \\
b_M^* &= v_A, \\
\pi_M(\alpha_1^*, \alpha_2^*, b_M^*) &= d_1[\theta_1(m_1 + r_1) - v_A] + d_2(\theta_0m_0 - v_A).
\end{align*}
\]

The equilibrium holds when $\theta_0m_0 \geq v_A$.

Notice that when configuration is the equilibrium, the manufacturer gives a positive participation rate while bidding by itself. The manufacturer submits the lowest possible bid $v_A$ that can keep the second position for itself and minimize the cost per click for $R_1$ at the same time.

By exchanging $R_1$ and $R_2$ in the solution above, we can get the manufacturer’s optimal profit under $(R_2, M)$, 
\[ \pi_M(\alpha_1^*, \alpha_2^*, b_M^*) = d_1[\theta_2(m_2 + r_2) - v_A] + d_2(\theta_0m_0 - v_A), \]
which is smaller than the manufacturer’s optimal profit under $(R_1, M)$ given $\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)$. Therefore, $(R_2, M)$ cannot be the position configuration in equilibrium.

- Consider the position configuration $(R_1, A)$.

For this position configuration to be the equilibrium, we need to have
\[ b_1 > b_A > \max\{b_M, b_2\}. \tag{4.A39} \]

The SNE condition and the LB selection rule that guard against $R_2$’s deviation from position 3 to position 2 imply that,
\[ b_2 = \frac{\theta_2r_2}{1 - \alpha_2} \tag{4.A40} \]

\textsuperscript{21}The channel profit is not affected by $b_1$ given that it is higher than $b_M, b_A, b_2$.  

\[ 179 \]
Given (4.A39) and (4.A40), we can show that the SNE condition that guards against 
$R_2$'s deviation to position 1 is has been satisfied.

Given $b_M$, the SNE condition and the LB selection rule that guard against the 
outside advertiser A's deviation from position 2 to position 1 imply that,

$$b_A = \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \max\{b_M, b_2\}. \quad (4.A41)$$

Given this equation and $b_A > \max\{b_M, b_2\}$, the SNE condition that guards against 
the outside advertiser A's deviation to position 3 has been satisfied.

The NE condition that guard against $R_1$'s deviation from position 1 to position 2 imply that,

$$d_1[\theta_1 r_1 - (1 - \alpha_1)b_A] > d_2[\theta_1 r_1 - (1 - \alpha_1)\max\{b_M, b_2\}]. \quad (4.A42)$$

Given (4.A41) and (4.A42), the NE condition that guards against $R_1$'s deviation to 
position 3 has been satisfied.

The manufacturer chooses $\alpha_1, \alpha_2$ and $b_M$ to maximize its profit $\pi_M(\alpha_1, \alpha_2, b_M)$ 
subject to (4.A39)-(4.A42), where

$$\pi_M(\alpha_1, \alpha_2, b_M) = d_1(\theta_1 m_1 - \alpha_1 b_A),$$

$$= d_1 \left[ \theta_1 m_1 - \alpha_1 \left( \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \max\{b_M, \frac{\theta_2 r_2}{1 - \alpha_2}\} \right) \right]. \quad (4.A43)$$

which decreases in $\alpha_1, \alpha_2$ and $b_M$. Thus the manufacturer will choose $\alpha_2^* = 0, b_M^* = 0$.

Substituting (4.A41) to (4.A42), we get $\alpha_1^* = 1 - \frac{\theta_1 r_1}{v_A}$.

However, in order to prevent the manufacturer from deviating to the second po-
sition, we need to have $d_1(\theta_1 m_1 - \alpha_1 b_A) > d_1(\theta_1 m_1 - \alpha_1 b_A) + d_2(\theta_0 m_0 - b_A)$, which 
is equivalent to $b_A > \theta_0 m_0$. Given $b_M = 0, \alpha_2 = 0$, $b_A = \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \theta_2 r_2$. If 
$\frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \theta_2 r_2 < \theta_0 m_0$, advertiser A needs to bid at least $\theta_0 m_0$ to secure its position 
at the second position, and this will lower the manufacturer's profit as it increases 
the price per click for $R_1$. To avoid this, the manufacturer should commit not to par-
ticipate in the bidding in the first stage if it finds that to stay in the third position
(i.e., not to get displayed) is optimal. In such case, the equilibrium is the same as 
\((R_1, A)\) in the basic model.

Therefore, the solution to this optimization problem is

\[
\begin{align*}
\alpha_1^* &= 1 - \frac{\theta_1 r_1}{v_A}, \\
\alpha_2^* &= 0, \\
b_M^* &= 0, \\
\pi_M(\alpha_1^*, \alpha_2^*, b_M^*) &= d_1[\theta_1(m_1 + r_1) - v_A] + d_2 \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}.
\end{align*}
\]

The equilibrium conditions are satisfied given \(v_A > \theta_1 r_1, \theta_2 r_2\).

If we exchange the position of \(R_1\) and \(R_2\), we can get the manufacturer's optimal profit under \((R_2, A)\), i.e.,
\[
\pi_M(\alpha_1^*, \alpha_2^*, b_M^*) = d_1[\theta_2(m_2 + r_2) - v_A] + d_2 \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}.
\]
Again, it's lower than the manufacturer's optimal profit under \((R_1, A)\) given \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\), so \((R_2, A)\) cannot be the equilibrium position configuration.

- Consider the position configuration \((R_1, R_2)\).

Following similar reasoning as above, the manufacturer should commit not to participate in the bidding under this position configuration, so the optimal participation rates, equilibrium bids, and the manufacturer's optimal profit are the same as the position configuration \((R_1, R_2)\) in the basic model, i.e.,

\[
\begin{align*}
\alpha_1^* &= 1 - \frac{\theta_1 r_1}{v_A}, \\
\alpha_2^* &= 1 - \frac{\theta_2 r_2}{v_A}, \\
b_M^* &= 0, \\
\pi_M(\alpha_1^*, \alpha_2^*, b_M^*) &= d_1[\theta_1(m_1 + r_1) - v_A] + d_2[\theta_2(m_2 + r_2) - v_A].
\end{align*}
\]

Similarly, the manufacturer's profit cannot be improved if we exchange the positions of \(R_1, R_2\), so the position configuration \((R_2, R_1)\) can be the equilibrium configuration.

- Consider the position configuration \((A, R_1)\).
Following similar reasoning as above, the manufacturer quits the bidding under this position configuration, so the optimal participation rates, equilibrium bids, and the manufacturer’s optimal profit are the same as the position configuration \((A, R_1)\) in the basic model, i.e.,

\[
\alpha_1^* = \max \left\{ 1 - \frac{\theta_1 r_1}{\theta_2 r_2}, 0 \right\}, \\
\alpha_2^* = 0, \\
b_M^* = 0, \\
\pi_M(\alpha_1^*, \alpha_2^*, b_M^*) = d_2[\theta_1(m_1 + r_1) - \max\{\theta_1 r_1, \theta_2 r_2\}].
\]

Similarly, the manufacturer’s profit gets lower if we exchange the positions of \(R_1, R_2\), so the position configuration \((A, R_2)\) cannot be the position configuration in equilibrium.

So far we have analyzed all the possible position configurations in equilibrium for the 1M2R model. We have the following conclusions.

First, given the assumption \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\), retailer 2 still cannot get a higher position than retailer 1 in equilibrium.

Given that retailer 1 always gets a higher position than retailer 2, there are seven possible position configurations in equilibrium. Table 4.A3 summarizes the manufacturer’s profit and the optimal participation rates under each position configuration.

In the first stage, i.e., when the manufacturer chooses the participation rates, it will compare its optimal profit under these possible position configurations and choose the position that leads to the highest profit \(\pi_M(\alpha_1^*, \alpha_2^*, b_M^*)\). It will choose the optimal participation rates correspondingly. If the manufacturer finds that to stay in the third position is optimal for itself, it will commit to quit the bidding.

To compare the optimal profits under the seven possible position configurations is too complicated and the result will not be informative. So we focus on two managerially important questions: (1) when the manufacturer should participate in bidding, and (2) whether the manufacturer should sponsor the retailer(s) when it participates in bidding.
Table 4.A3: Optimal Participation Rates and Manufacturer's Profit.

<table>
<thead>
<tr>
<th>Position</th>
<th>( \pi_M(\alpha_1^<em>, \alpha_2, b_M^</em>) )</th>
<th>( (\alpha_1^<em>, \alpha_2^</em>) )</th>
<th>Requirement for NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>((M, R_1))</td>
<td>(d_1(\theta_0m_0 - v_A) + d_2[\theta_1(m_1 + r_1) - v_A] )</td>
<td>( (1 - \frac{\theta_1 r_1}{v_A}, 0) )</td>
<td>( \theta_0 m_0 \geq \theta_1(m_1 + r_1) ) and ( \frac{d_1}{d_1 + d_2} \theta_0 m_0 + \frac{d_2}{d_1 + d_2} \theta_1(m_1 + r_1) \geq v_A )</td>
</tr>
<tr>
<td>((R_1, M))</td>
<td>(d_1[\theta_1(m_1 + r_1) - v_A] + d_2(\theta_0m_0 - v_A) )</td>
<td>( (1 - \frac{\theta_1 r_1}{v_A}, 0) )</td>
<td>( \theta_0 m_0 \geq v_A )</td>
</tr>
<tr>
<td>((R_1, R_2))</td>
<td>(d_1[\theta_1(m_1 + r_1) - v_A] + d_2[\theta_2(m_2 + r_2) - v_A] )</td>
<td>( (1 - \frac{\theta_1 r_1}{v_A}, 1 - \frac{\theta_2 r_2}{v_A}) )</td>
<td></td>
</tr>
<tr>
<td>((M, A))</td>
<td>(d_1(\theta_0m_0 - v_A) + d_2(v_A - \max{\theta_1 r_1, \theta_2 r_2}) )</td>
<td>( (0, 0) )</td>
<td>( \theta_0 m_0 \geq v_A )</td>
</tr>
<tr>
<td>((R_1, A))</td>
<td>(d_1[\theta_1(m_1 + r_1) - v_A] + d_2 \left( \frac{v_A - \theta_1 r_1}{v_A} \right) \left( v_A - \theta_2 r_2 \right) )</td>
<td>( (1 - \frac{\theta_1 r_1}{v_A}, 0) )</td>
<td></td>
</tr>
<tr>
<td>((A, M))</td>
<td>(d_2(\theta_0m_0 - \max{\theta_1 r_1, \theta_2 r_2}) )</td>
<td>( (0, 0) )</td>
<td>( v_A \geq \theta_0 m_0 )</td>
</tr>
<tr>
<td>((A, R_1))</td>
<td>(d_2[\theta_1(m_1 + r_1) - \max{\theta_1 r_1, \theta_2 r_2}] )</td>
<td>( (\max{1 - \frac{\theta_1 r_1}{\theta_2 r_2}, 0}, 0) )</td>
<td></td>
</tr>
</tbody>
</table>
Before we answer these two questions, we first notice that it is relatively easier to classify the seven position configuration into three categories (the channel takes both positions, takes the first position, and takes the second position) and compare the manufacturer’s profit within each category.

- **Category 1:** the channel takes both positions, i.e., \((M, R_1), (R_1, M), (R_1, R_2)\). Within this category, the optimal position configuration is determined by the rank of channel profit per click. That is, if \(\theta_0 m_0 \geq \theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2)\), the optimal position configuration out of the three is \((M, R_1)\); if \(\theta_1(m_1 + r_1) > \theta_0 m_0 \geq \theta_2(m_2 + r_2)\), the optimal one is \((R_1, M)\); and if \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) > \theta_0 m_0\), the optimal one is \((R_1, R_2)\).

- **Category 2:** the channel takes only the first position, i.e., \((M, A), (R_1, A)\). Within this category, \((M, A)\) is the more profitable one if and only if

\[
\theta_0 m_0 \geq \theta_1(m_1 + r_1) - \frac{d_2}{d_1} \left( v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A} \right),
\]

and \((R_1, A)\) is the more profitable one if and only if

\[
\theta_0 m_0 + \frac{d_2}{d_1} \left( v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A} \right) < \theta_1(m_1 + r_1).
\]

Notice that \(v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A} > 0\), so (4.4A44) is a weaker condition than \(\theta_0 m_0 \geq \theta_1(m_1 + r_1)\). Therefore, when \(\theta_0 m_0 \geq \theta_1(m_1 + r_1)\), \((M, A)\) is definitely more profitable than \((R_1, A)\), but \((R_1, A) > (M, A)\) needs a stronger condition than \(\theta_1(m_1 + r_1) > \theta_0 m_0\).

- **Category 3:** the channel takes only the second position, i.e., \((A, M), (A, R_1)\). \((A, M)\) is the more profitable one if \(\theta_0 m_0 \geq \theta_1(m_1 + r_1)\), and \((A, R_1)\) is the more profitable one if and only if \(\theta_1(m_1 + r_1) > \theta_0 m_0\).

The manufacturer can first select the winner from each category, and then compare the candidates selected from the three categories.
We divide the parameter space into three cases according to the relationship between the manufacturer’s channel profit per click and the two retailers’ channel profit per click.

- **Case 1:** The manufacturer’s channel profit per click is higher than both retailers, i.e., $\theta_0 m_0 \geq \theta_1 (m_1 + r_1) \geq \theta_2 (m_2 + r_2)$.

  The candidates from the three categories are $(M, R_1)$, $(M, A)$, and $(A, M)$. No matter which one out of the three is the most profitable, the manufacturer participates in bidding in equilibrium. The manufacturer sponsors $R_1$ when $(M, R_1)$ is the position configuration in equilibrium, i.e., when $\theta_1 (m_1 + r_1) \geq 2v_A - \max \{\theta_1 r_1, \theta_2 r_2\}$.

- **Case 2:** The manufacturer’s channel profit per click is higher than $R_2$ but lower than $R_1$, i.e., $\theta_1 (m_1 + r_1) > \theta_0 m_0 \geq \theta_2 (m_2 + r_2)$.

  The candidates from the first and the third categories are $(R_1, M)$ and $(A, R_1)$ respectively. In the second category, the winner is $(R_1, A)$ if $\theta_1 (m_1 + r_1) > \theta_0 m_0 + \frac{d_2}{d_1} (v_A - \max \{\theta_1 r_1, \theta_2 r_2\}) - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}$, and is $(M, A)$ if $\theta_1 (m_1 + r_1) > \theta_0 m_0 \geq \theta_1 (m_1 + r_1) - \frac{d_2}{d_1} (v_A - \max \{\theta_1 r_1, \theta_2 r_2\}) - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}$.

  The manufacturer participates in bidding when $(R_1, M)$ or $(M, A)$ is the position configuration in equilibrium, and sponsors $R_1$ in the former case.

  $(R_1, M)$ is more profitable than the other three position configurations if and only if

  $\theta_0 m_0 - v_A \geq \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}$,

  $\theta_1 (m_1 + r_1) - v_A \geq \frac{d_1 - d_2}{d_1} (\theta_0 m_0 - v_A) + \frac{d_2}{d_1} (v_A - \max \{\theta_1 r_1, \theta_2 r_2\})$,

  and $\theta_1 (m_1 + r_1) - v_A \geq \frac{d_2}{d_1} \left( \theta_1 (m_1 + r_1) - \theta_0 m_0 + v_A - \max \{\theta_1 r_1, \theta_2 r_2\} \right)$.

  Notice that the requirement for NE is automatically satisfied. Therefore, $(R_1, M)$

  \(^{22}(M, R_1) > (M, A)$ is equivalent to $\theta_1 (m_1 + r_1) \geq 2v_A - \max \{\theta_1 r_1, \theta_2 r_2\}$, which implies $\theta_0 m_0 \geq \theta_1 (m_1 + r_1) > v_A$ and thus indicates $(M, R_1) > (M, A) > (A, M)$. The requirement for $(M, R_1)$ to be a NE is also automatically satisfied.\[185\]
is the position configuration in equilibrium when the conditions above are satisfied.

Similarly, \((M, A)\) is the position configuration in equilibrium if and only if

\[
\begin{align*}
\theta_1(m_1 + r_1) - v_A &< \frac{d_1 - d_2}{d_1} (\theta_0 m_0 - v_A) + \frac{d_2}{d_1} (v_A - \max\{\theta_1 r_1, \theta_2 r_2\}), \\
\theta_1(m_1 + r_1) - \theta_0 m_0 &\leq \frac{d_2}{d_1} (v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}), \\
\text{and } \theta_0 m_0 - v_A &\geq \frac{d_2}{d_1} [\theta_1(m_1 + r_1) - v_A].
\end{align*}
\]

- Case 3: The manufacturer’s channel profit per click is lower than both retailers, i.e., \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) > \theta_0 m_0\).

The candidates from the first and the third categories are \((R_1, R_2)\) and \((A, R_1)\).

In the second category, the winner is \((R_1, A)\) if \(\theta_1(m_1 + r_1) > \theta_0 m_0 + \frac{d_2}{d_1} (v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A})\), and is \((M, A)\) if \(\theta_1(m_1 + r_1) \geq \theta_2(m_2 + r_2) > \theta_0 m_0 \geq \theta_1(m_1 + r_1) - \frac{d_2}{d_1} (v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A})\).

The manufacturer participates in bidding only when \((M, A)\) is the position configuration in equilibrium, i.e., when

\[
\begin{align*}
\theta_1(m_1 + r_1) - v_A &< \frac{d_2}{d_1} \left(2v_A - \theta_2(m_2 + r_2) - \max\{\theta_1 r_1, \theta_2 r_2\}\right), \\
\theta_1(m_1 + r_1) - \theta_0 m_0 &\leq \frac{d_2}{d_1} (v_A - \max\{\theta_1 r_1, \theta_2 r_2\} - \frac{(v_A - \theta_1 r_1)(v_A - \theta_2 r_2)}{v_A}), \\
\text{and } \theta_0 m_0 - v_A &\geq \frac{d_2}{d_1} [\theta_1(m_1 + r_1) - v_A].
\end{align*}
\]

In this position configuration, the manufacturer will not sponsor the retailers.

From the results above, we can see that the manufacturer will definitely participate in bidding when its profit per click is higher than both retailers, and may participate in bidding when its profit per click is lower than one or both retailers but not too low. When the manufacturer participates in bidding, it may still sponsor one retailer but will not sponsor both retailers. The manufacturer will support both retailers only when its profit per click is lower than both retailers and the manufacturer quits the
bidding itself.

4.A.6 Proof of Theorem 4

As a counterpart to Lemma 1, Varian (2007) has shown that in equilibrium, \( e_1 v_1 \geq e_2 v_2 \geq e_3 v_3 \), where \( v_i \) denotes the profit per click for the advertiser at position \( i \). The equilibrium bids are,

\[
\begin{align*}
    b_3 &= v_3, \\
    b_2 &= \frac{x_1 - x_2}{x_1} v_2 + \frac{x_2}{x_1} e_3 v_3.
\end{align*}
\]

Similarly with the basic model, we assume that \( e_A v_A > e_1 r_1, e_2 r_2 \). We analyze the retailers and outside advertiser’s bid as well as the manufacturer’s choice of participation rate given the positions of all advertisers. Then we compare the manufacturer’s profits among six all position configurations to identify the equilibrium.

We first consider the three position configurations where retailer 1 gets a higher position than retailer 2.

- Consider the position configuration \((R_1, R_2)\).

This happens when,

\[
\frac{e_1 \theta_1 r_1}{1 - \alpha_1} \geq \frac{e_2 \theta_2 r_2}{1 - \alpha_2} \geq e_A v_A.
\]

The manufacturer’s profit is,

\[
\pi_M(\alpha_1, \alpha_2) = x_1 e_1 \left[ \theta_1 m_1 - \alpha_1 \frac{e_2}{e_1} \left( \frac{x_1 - x_2}{x_1} \frac{\theta_2 r_2}{1 - \alpha_2} + \frac{x_2 e_A v_A}{x_1 e_2} \right) \right] + x_2 e_2 \left( \theta_2 m_2 - \alpha_2 \frac{e_A}{e_2} v_A \right),
\]

which decreases in \( \alpha_1, \alpha_2 \). Therefore the manufacturer will choose the smallest \( \alpha_1, \alpha_2 \) that satisfy (4.A46). The optimal participation rates are,

\[
\alpha_i^* = 1 - \frac{e_i \theta_i r_i}{e_A v_A}, \ i = 1, 2.
\]

187
Correspondingly, the manufacturer’s maximum profit is,

$$\pi_M(\alpha_1^*, \alpha_2^*) = x_1[e_1\theta_1(m_1 + r_1) - e_A v_A] + x_2[e_2\theta_2(m_2 + r_2) - e_A v_A].$$

- Consider the position configuration \((R_1, A)\).

This happens when,

$$\frac{e_1\theta_1 r_1}{1 - \alpha_1} \geq e_A v_A > \frac{e_2\theta_2 r_2}{1 - \alpha_2}. \tag{4.A47}$$

The manufacturer’s profit is,

$$\pi_M(\alpha_1, \alpha_2) = x_1 e_1 \left[ \theta_1 m_1 - \alpha_1 \frac{e_A}{e_1} \left( \frac{x_1 - x_2}{x_1} v_A + \frac{x_2}{x_1} \frac{\theta_2 r_2}{e_A 1 - \alpha_2} \right) \right],$$

which decreases in \(\alpha_1, \alpha_2\). Therefore the manufacturer will choose the smallest \(\alpha_1, \alpha_2\) that satisfy (4.A47). The optimal participation rates are,

$$\alpha_1^* = 1 - \frac{e_1\theta_1 r_1}{e_A v_A},$$

$$\alpha_2^* = 0.$$

Correspondingly, the manufacturer’s profit is

$$\pi_M(\alpha_1^*, \alpha_2^*) = x_1 [e_1\theta_1(m_1 + r_1) - e_A v_A] + x_2 \left[ e_A v_A - e_1\theta_1 r_1 - e_2\theta_2 r_2 + \frac{e_1\theta_1 r_1 \cdot e_2\theta_2 r_2}{e_A v_A} \right].$$

- Consider the position configuration \((A, R_1)\).

This happens when,

$$e_A v_A > \frac{e_1\theta_1 r_1}{1 - \alpha_1} \geq \frac{e_2\theta_2 r_2}{1 - \alpha_2}. \tag{4.A48}$$

The manufacturer’s profit is,

$$\pi_M(\alpha_1, \alpha_2) = x_2 e_1 \left[ \theta_1 m_1 - \alpha_1 \frac{e_2}{e_1} \frac{\theta_2 r_2}{1 - \alpha_2} \right],$$

which decreases in \(\alpha_1, \alpha_2\). Therefore the manufacturer will choose the smallest \(\alpha_1, \alpha_2\)
that satisfy (4.A48). The optimal participation rates are,

\[
\alpha_1^* = \frac{\max\{e_1\theta_1 r_1, e_2\theta_2 r_2\} - e_1\theta_1 r_1}{e_2\theta_2 r_2}, \\
\alpha_2^* = 0.
\]

Correspondingly, the manufacturer’s profit is

\[
\pi_M(\alpha_1^*, \alpha_2^*) = x_2 [e_1\theta_1 (m_1 + r_1) - \max\{e_1\theta_1 r_1, e_2\theta_2 r_2\}].
\]

The other three configurations can be solved by symmetry. The remaining analysis is straightforward and the same with basic model, thus omitted.

4.A.7 Proof of Theorem 5:

Step 1:

We have analyzed the case where the manufacturer sponsors both retailers to get displayed. The manufacturer’s profit under the optimal participation rates is

\[\pi_{M(2)}^* = \theta(w + t) - 2v_A,\]

where the subscript (2) denotes “under the case that the manufacturer sponsors two retailers”.

Step 2:

Now we consider the case in which the manufacturer sponsors only one retailer to get displayed.

Since the two retailers are ex ante symmetric, we assume that retailer 1 gets displayed without loss of generalizability. One outside advertiser will take the second position. \(x_1\) of the consumers on the Hotelling line will purchase from retailer 1, where \(x_1\) solves \(v - tx_1 - p_1 = 0\), or \(x_1 = 1\) if \(v - t - p_1 \geq 0\). The price per click for retailer 1 is \(v_A\).

Retailer 1’s profit is \(\pi_1 = \hat{\theta} \min\{\frac{v - p_1}{t}, 1\} (p_1 - w) - (1 - \alpha_1)v_A\). The optimal retail price for retailer 1 is \(p_1^* = \max\{\frac{v + w}{2t}, v - t\}\), and it will cover \(x^* = \min\{\frac{v - w}{2t}, 1\}\) of
consumers on the Hotelling line. Retailer 1’s profit is

\[
\pi_1^* = \begin{cases} 
\frac{\theta}{4t}(v - w)^2 - (1 - \alpha_1)v_A & \text{if } t > \frac{v-w}{2}, \\
\theta(v - w - t) - (1 - \alpha_1)v_A & \text{if } t \leq \frac{v-w}{2}.
\end{cases}
\]

The manufacturer’s profit given his participation rates \( \alpha_1 \) is

\[
\pi_M^{(1)} = \begin{cases} 
\frac{\theta}{2t}\frac{v-w}{2w}w - \alpha_1v_A & \text{if } t > \frac{v-w}{2}, \\
\theta w - \alpha_1v_A & \text{if } t \leq \frac{v-w}{2}.
\end{cases}
\]

which decreases in \( \alpha_1 \). Similarly, we use the subscript \((1)\) denotes the case where the manufacturer sponsors one retailer. Therefore, the manufacturer will choose the lowest participation rate \( \alpha_1 \) that can help retailer 1 get displayed, i.e., \( \pi_1^* \geq 0 \). This leads to

\[
\alpha_1^* = \begin{cases} 
1 - \frac{\theta}{4tv_A}(v - w)^2 & \text{if } t \geq \frac{v-w}{2}, \\
1 - \frac{\theta(v-w-t)}{v_A} & \text{if } t \leq \frac{v-w}{2}.
\end{cases}
\]

The manufacturer’s profit under the optimal participation rate \( \alpha_1^* \) is

\[
\pi_M^{(1)} = \begin{cases} 
\frac{\theta}{4t}(v^2 - w^2) - v_A & \text{if } t \geq \frac{v-w}{2}, \\
\theta(v - t) - v_A & \text{if } t \leq \frac{v-w}{2}.
\end{cases}
\]

**Step 3:**

Suppose the manufacturer does not sponsor either retailer. The two outside advertisers will take both positions, and the manufacturer’s profit will be zero.

**Step 4:**

Before we compare the manufacturer’s profit under the cases of sponsoring two, one, zero retailers, we need to pin down the restrictions on parameter space based on our assumptions: (1) neither retailer can get displayed without support from the manufacturer, and (2) the Hotelling line is covered with two retailers.

Assumption (1) implies that \( \alpha_i^* > 0 \), because otherwise the manufacturer provide zero participation rate and the retailer(s) can still get displayed. For the case of
manufacturer sponsoring two retailers, \( \alpha_i^* > 0, i = 1, 2 \) is equivalent to \( t < 2v_A/\theta \).

For the case of manufacturer sponsoring only one retailer, we need to have \( \alpha_1^* > 0 \). Notice that \( \alpha_1^* \) increases in \( t \), and thus \( \alpha_1^* > 0 \) is equivalent to \( t > v - w - \frac{v_A}{\theta} \) when \( t \leq \frac{v-w}{2} \), and equivalent to \( t > \frac{\theta}{4v_A}(v-w)^2 \) when \( t > \frac{v-w}{2} \). By combining the two cases, we have that \( 2v_A/\theta \geq t \geq \max \left\{ v - w - \frac{v_A}{\theta}, \frac{\theta}{4v_A}(v-w)^2 \right\} \).

It is straightforward to show that assumptions (1) and (2) together imply the restriction: \( t \in (B_0, \frac{2}{3}(v-w)] \), where

\[
B_0 = \left\{ \begin{array}{ll}
\frac{\theta}{4v_A}(v-w)^2, & \text{if } \frac{3}{8}(v-w) < \frac{v_A}{\theta} < \frac{v-w}{2} \\
v - w - \frac{v_A}{\theta}, & \text{if } \frac{v_A}{\theta} \geq \frac{v-w}{2}.
\end{array} \right.
\]

Notice that assumption (1) also guarantees that when the manufacturer only sponsors retailer 1 to get displayed, retailer 2 has no incentive to deviate. Suppose retailer 2 deviates to take a position. For any price \( p_2 \), if the market is covered, retailer 2’s profit will be less than the profit when retailer 2 is the monopoly, charging the same price. If the market is not covered, retailer 2 can be regarded as a monopoly. In both cases, retailer 2’s profit cannot be higher than \( \pi_2^* \) since this is the highest possible profit that a monopoly can get on the Hotelling line. Our parameter restrictions ensure that \( \pi_2^* \) is less than \( v_A \), which is the cost per click to get displayed. Therefore, retailer 2 has not incentive to get displayed without support from the manufacturer.

**Step 5:**

We first compare the manufacturer’s optimal profits under the cases of sponsoring two versus one retailer.

For \( t \geq \max \{ \frac{v-w}{2}, \frac{\theta}{4v_A}(v-w)^2 \} \), it is more profitable for the manufacturer to sponsor one retailer if and only if \( \frac{\theta}{4v_A}(v-w)^2 + \frac{\sqrt{(v_A-\theta w)^2 + \theta^2(v-w)^2}}{2\theta} > \frac{v-w}{2} \). Notice that \( v_A - \frac{\theta}{2\theta} + \frac{\sqrt{(v_A-\theta w)^2 + \theta^2(v-w)^2}}{2\theta} > \frac{v-w}{2} \) is equivalent to \( v - w > 0 \), which always holds by assumption. Also \( \frac{v_A - \theta w}{2\theta} + \frac{\sqrt{(v_A-\theta w)^2 + \theta^2(v-w)^2}}{2\theta} > \frac{2v_A}{\theta} \) is equivalent to \( \frac{v_A - w}{2} > wv_A \), which holds given \( v_A < \frac{v-w}{2} \) and \( w < \frac{v+w}{2} \). We have proven that when \( \frac{3}{8}(v-w) < \frac{v_A}{\theta} < \frac{v-w}{2}, \frac{2v_A}{\theta} > \frac{\theta}{4v_A}(v-w)^2 \).
Therefore, when \( \frac{3}{5}(v - w) < \frac{v_A}{\theta} < \frac{v - w}{2} \), for any \( t \in (\frac{\theta(v - w)^2}{4v_A}, \frac{2}{3}(v - w)] \), sponsoring one retailer is more profitable than sponsoring two retailers and this set is non-empty. When \( \frac{v_A}{\theta} \geq \frac{v - w}{2} \), for \( t \in \left[ \frac{v - w}{2}, \frac{v_A - \theta w}{2\theta} + \frac{\sqrt{(v_A - \theta w)^2 + \theta(v^2 - w^2)}}{2\theta} \right] \), sponsoring one retailer is more profitable than sponsoring two retailers and this set is non-empty.

For \( t < \frac{v - w}{2} \), it is more profitable for the manufacturer to sponsor one retailer if and only if \( t < \frac{1}{2}(\frac{v_A}{\theta} + v - w) \), which is naturally satisfied given that \( t < \frac{v - w}{2} \). Therefore, when \( t < \frac{v - w}{2} \), it is always more profitable for the manufacturer to sponsor only one retailer than to sponsor two retailers. This result is intuitive, since when \( t < \frac{v - w}{2} \), one retailer can already covers the market. The manufacturer cannot get more demand by sponsoring two retailers, but will incur higher sponsoring cost. Thus when \( t < \frac{v - w}{2} \), it is always more profitable to sponsor one retailer than to sponsor both retailers.

Summarizing the two cases and combining with the parameter restrictions we get in Step 4: denote

\[
B_1 = \frac{v_A - \theta w}{2\theta} + \frac{\sqrt{(v_A - \theta w)^2 + \theta(v^2 - w^2)}}{2\theta},
\]

\[
B_2 = \frac{2}{3}(v - w).
\]

- For \( t \in (B_0, \text{min}\{B_1, B_2\}] \), it is more profitable for the manufacture to sponsor one retailer rather than to sponsor two retailers. Notice that this interval is always non-empty given our assumptions.

- For \( t \in (\text{min}\{B_1, B_2\}, B_2 \] \), it is more profitable for the manufacturer to sponsor two retailers rather than to sponsor one retailer. (If \( B_1 \leq B_2 \), then this interval is empty and there does not exist \( t \) such that it is more profitable to sponsor two retailers.)

**Step 6:**

Then we need to figure out whether the manufacturer would rather sponsor neither retailer for some \( t \), i.e., to check whether \( \text{max}\{\pi_{M(1)}^*, \pi_{M(2)}^*\} \) is negative. We can classify \( t \) into three intervals, as listed below.

(1) When \( \frac{v_A}{\theta} \geq \frac{v - w}{2} \), there exists \( t \in (v - w - \frac{v_A}{\theta}, \frac{v - w}{2}] \), and for those \( t \), \( \pi_{M(1)}^* \)
\[ \pi^{*}_{M(2)} \]. Comparing \( \pi^{*}_{M(1)} = \theta(v - t) - v_A \) with 0, \( \pi^{*}_{M(1)} \geq 0 \) is equivalent to \( t \leq v - \frac{v_A}{\theta} \).

- If \( v - \frac{v_A}{\theta} \leq \frac{v-w}{2} \) (which is equivalent to \( \frac{v_A}{\theta} \geq \frac{1}{2}(v + w) \)), then the manufacturer should sponsor one retailer when \( t \in \left( v - w - \frac{v_A}{\theta}, v - \frac{v_A}{\theta} \right] \), and should sponsor zero retailer when \( t \in \left( v - \frac{v_A}{\theta}, \frac{v-w}{2} \right] \). Notice that the former interval is non-empty if and only if \( v > \frac{v_A}{\theta} \).

- If \( v - \frac{v_A}{\theta} > \frac{v-w}{2} \) (which is equivalent to \( \frac{v_A}{\theta} < \frac{1}{2}(v + w) \)), then the manufacturer should sponsor one retailer for any \( t \in \left( v - w - \frac{v_A}{\theta}, \frac{v-w}{2} \right] \).

(2) For \( t \in \left( \max\{\frac{v-w}{2}, \frac{\theta}{4v_A} (v - w)^2\}, \min\{B_1, B_2\} \right] \), \( \pi^{*}_{M(1)} > \pi^{*}_{M(2)} \). Comparing \( \pi^{*}_{M(1)} = \frac{\theta}{4t} (v^2 - w^2) - v_A \) with 0.

- If \( \frac{v_A}{\theta} \geq \frac{1}{2}(v+w) \), then for any \( t > \frac{v-w}{2}, \frac{\theta}{4t} (v^2-w^2) - v_A = \frac{v-w}{2} \theta - v_A < \frac{v+w}{2} \theta - v_A \leq 0 \), which means that for any \( t \in \left( \max\{\frac{v-w}{2}, \frac{\theta}{4v_A} (v - w)^2\}, \min\{B_1, B_2\} \right] \), the manufacturer should sponsor one retailer

- If \( \frac{v_A}{\theta} < \frac{1}{2}(v + w) \), then \( \frac{\theta}{4t} (v^2 - w^2) - v_A \geq 0 \) when \( t \leq \frac{\theta(v^2-w^2)}{4v_A} \). Thus the manufacturer should sponsor one retailer for

\[
\begin{aligned}
& t \in \left( \max\{\frac{v-w}{2}, \frac{\theta}{4v_A} (v - w)^2\}, \min\{\frac{\theta(v^2-w^2)}{4v_A}, B_1, B_2\} \right]. \\
\end{aligned}
\]

Notice that \( \frac{\theta(v^2-w^2)}{4v_A} > \frac{v-w}{2}, \frac{\theta(v^2-w^2)}{4v_A} > \frac{\theta}{4v_A} (v - w)^2 \). We have also shown that \( B_1, B_2 > \frac{v-w}{2} \), and \( B_1, B_2 > \frac{\theta}{4v_A} (v - w)^2 \) when \( \frac{3}{8} (v - w) < \frac{v_A}{\theta} < \frac{v-w}{2} \). Thus the interval \( \left( \max\{\frac{v-w}{2}, \frac{\theta}{4v_A} (v - w)^2\}, \min\{\frac{\theta(v^2-w^2)}{4v_A}, B_1, B_2\} \right] \) is non-empty. For \( t \in \left( \min\{\frac{\theta(v^2-w^2)}{4v_A}, B_1, B_2\}, \min\{B_1, B_2\} \right] \), the manufacturer should sponsor zero retailer.

(3) When \( B_1 < B_2 \), there exists \( t \in (B_1, B_2) \) such that \( \pi^{*}_{M(1)} < \pi^{*}_{M(2)} \). Comparing \( \pi^{*}_{M(2)} = \theta(w + t) - 2v_A \) to 0. \( \pi^{*}_{M(2)} > 0 \) is equivalent to \( t > 2\frac{v_A}{\theta} - w \). Thus the manufacturer should sponsor two retailers when \( t \in (\max\{B_1, 2\frac{v_A}{\theta} - w\}, B_2) \).

Notice that \( 2\frac{v_A}{\theta} - w < B_2 \equiv \frac{2}{3} (v - w) \) is equivalent to \( \frac{v_A}{\theta} < \frac{1}{3} v + \frac{1}{6} w \). \( B_1 < B_2 \) is equivalent to \( \frac{v_A}{\theta} > \frac{7}{24} v - \frac{1}{24} w \).
Thus when \( \frac{7}{24}v - \frac{1}{24}w < \frac{v_A}{\theta} < \frac{1}{3}v + \frac{1}{6}w \), the manufacturer should sponsor zero retailer for \( t \in (B_1, \max\{B_1, 2\frac{v_A}{\theta} - w\}) \) and should sponsor two retailers for \( t \in (\max\{B_1, 2\frac{v_A}{\theta} - w\}, B_2) \). When \( \frac{v_A}{\theta} \geq \frac{1}{3}v + \frac{1}{6}w \), the manufacturer should sponsor zero retailer for any \( t \in (B_1, B_2) \).

Summarizing the three intervals, we get the following result:

- If \( v_A \geq \theta v \), the manufacturer should sponsor neither retailer for any \( t \) under our parameter restrictions.

- If \( \frac{\theta}{2}(v + w) \leq v_A < \theta v \), the manufacturer should sponsor one retailer for \( t \in (v - w - \frac{v_A}{\theta}, v - \frac{v_A}{\theta}) \), and should sponsor neither retailer for \( t \in (v - \frac{v_A}{\theta}, B_2) \).

- If \( \frac{\theta}{8}(v - w) < v_A < \frac{\theta}{2}(v + w) \), the manufacturer should sponsor one retailer for

\[
t \in \left( B_0, \min \left\{ \frac{\theta(v^2 - w^2)}{4v_A}, B_1, B_2 \right\} \right),
\]

and this interval is non-empty given the parameter assumptions. The manufacturer should sponsor two retailers for any

\[
t \in \left[ \max \left\{ 2\frac{v_A}{\theta} - w, B_1 \right\}, B_2 \right],
\]

and this interval is non-empty only when \( \frac{7}{24}v - \frac{1}{24}w < \frac{v_A}{\theta} < \frac{1}{3}v + \frac{1}{6}w \). The manufacturer should sponsor neither retailer for any

\[
t \in \left[ \min \left\{ \frac{\theta(v^2 - w^2)}{4v_A}, B_1, B_2 \right\}, \min \left\{ \max \left\{ 2\frac{v_A}{\theta} - w, B_1 \right\}, B_2 \right\} \right).
\]

### 4.A.8 Endogenous Wholesale Contracts and Retail Prices:

#### Proof of Lemma 3:

**Proof.** Suppose \( v_1^* \geq v_2^* \), we first prove that retailer 1 will have a higher position than retailer 2 in equilibrium. We prove by contradiction. Suppose in equilibrium retailer 1 has a lower position than retailer 2, then we must have \( v_1 < v_2 \). If retailer 1 is in position 3, it can deviate by setting \( v_1 \) equal to \( v_1^* \), and earn positive profit at position...
Thus, it must be that retailer 1 takes position 2, and retailer 2 takes position 1. We have $v_1^* \geq v_2^* \geq v_A$. In this case, retailer 1’s profit $\pi_{R_1} = d_2 [v_1 - (1 - \alpha_1)v_A]$, which increases with $v_1$, so retailer 1 will choose the largest possible $v_1$ given its position. We have $v_1 = v_2^*$, and $\pi_{R_1} = d_2 [v_2^* - (1 - \alpha_1)v_A]$. Now, if retailer 1 deviates by setting $v_1 = v_1^*$, it will take the first position, and retailer 2 will take the second position. Similarly, given its position, retailer 2 will set $v_2 = v_2^*$ to maximize its profit. As a result, retailer 1 will earn,

$$\pi_{R_1}^* = d_1 \left[ v_1^* - (1 - \alpha_1) \left( \frac{d_1 - d_2}{d_1} v_2^* + \frac{d_2}{d_1} v_A \right) \right]$$

$$> d_1 \left[ v_1 - (1 - \alpha_1) \left( \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} v_A \right) \right]$$

$$= \pi_{R_1}.$$ 

This implies that it is not an equilibrium for retailer 1 to take the second position while retailer 2 is taking the first position. To summarize, we have proved that in equilibrium, when $v_1^* \geq v_2^*$, retailer 1 has a higher position than retailer 2.

Next, suppose $v_1^* \geq v_2^*$, and we need to prove that (i) when $v_A > v_1^*$, the outside advertiser will take the first position; (ii) when $v_1^* \geq v_A > v_2^*$, the outside advertiser will take the second position; and (iii) when $v_2^* \geq v_A$, the outside advertiser will take the third position.

(i) is obvious: given $v_A > v_1^* \geq v_1$, the outsider advertiser must take a higher position than retailer 1.

(ii): When $v_1^* \geq v_A > v_2^*$, we have $v_A > v_2^* \geq v_2$, so the outside advertiser will always take a higher position than retailer 2. Suppose it takes position 1, then we must have $v_A > v_1$. In this case, retailer 1’s profit is $\pi_{R_1} = d_2 [v_1 - (1 - \alpha_1)v_2]$. Similarly as above, we can show that retailer 1 can earn a higher profit by setting $v_1$ as $v_1^*$ and take the first position instead. Then (ii) is proved.

(iii) is straightforward to prove by contradiction. Suppose $v_2^* \geq v_A$ and the outside advertiser takes the second position. In this case, retailer 2 will take position 3 and earn zero profit, and it can deviate by setting $v_2$ at $v_2^*$ and taking the second position instead, which will give its positive profit.
Given that the positions are determined by the rank of \( v_i^*, v_2^*, \) and \( v_A \), we know that each retailer \( i \) profit function increases with \( v_i \), so in equilibrium, retailer \( i \) will set retail price \( p_i^* = (1 + w_i)/2 \), under which \( v_i \) takes the maximum value \( v_i^* \).

**Proof of Theorem 6**

Without loss of generalizability, we assume that retailer 1 gets a higher position than retailer 2 in equilibrium. There are three possible position configurations then.

- Consider the position configuration \((R_1, R_2)\).

The manufacturer's optimization problem is,

\[
\max_{\alpha_i, w_i} \quad d_1 \left[ \frac{1 - w_i}{2} (w_1 - c) - \alpha_1 \left( \frac{d_1 - d_2}{d_1} \tilde{\theta} (1 - w_2)^2 + \frac{d_2}{d_1} v_A \right) \right] \\
+ d_2 \left[ \frac{1 - w_2}{2} (w_2 - c) - \alpha_2 v_A \right]
\]

s.t. \( \frac{\tilde{\theta} (1 - w_1)^2}{4(1 - \alpha_1)} \geq \frac{\tilde{\theta} (1 - w_2)^2}{4(1 - \alpha_2)} \geq v_A \),

\[ 0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2. \]

The optimal solution is,

\[
\begin{cases}
    w_1^* = w_2^* = c, \quad \alpha_1^* = \alpha_2^* = 1 - \frac{\tilde{\theta} (1 - c)^2}{4 v_A}, & \text{when } v_A \geq \frac{\tilde{\theta} (1 - c)^2}{4} \\
    w_1^* = w_2^* = 1 - 2 \sqrt{\frac{v_A}{\tilde{\theta}}}, \quad \alpha_1^* = \alpha_2^* = 0, & \text{when } \frac{\tilde{\theta} (1 - c)^2}{16} \leq v_A \leq \frac{\tilde{\theta} (1 - c)^2}{4} \\
    w_1^* = w_2^* = \frac{1 + c}{2}, \quad \alpha_1^* = \alpha_2^* = 0, & \text{when } v_A \leq \frac{\tilde{\theta} (1 - c)^2}{16}
\end{cases}
\]

Correspondingly, the manufacturer's profit is,

\[
\pi_M^* = \begin{cases}
    (d_1 + d_2) \left[ \frac{\tilde{\theta} (1 - c)^2}{4} - v_A \right] & \text{when } v_A \geq \frac{\tilde{\theta} (1 - c)^2}{4} \\
    (d_1 + d_2) \left[ (1 - c) \sqrt{v_A \tilde{\theta}} - 2 v_A \right] & \text{when } \frac{\tilde{\theta} (1 - c)^2}{16} \leq v_A \leq \frac{\tilde{\theta} (1 - c)^2}{4} \\
    (d_1 + d_2) \frac{\tilde{\theta} (1 - c)^2}{8} & \text{when } v_A \leq \frac{\tilde{\theta} (1 - c)^2}{16}
\end{cases}
\]

To summarize, in this case, the manufacturer sells to both retailers. The wholesale prices and participation rates for the two retailers are "symmetric", i.e., the manu-
manufacturer will only provide a marginally lower wholesale price or a marginally higher participation rate to retailer 1 in order to let it get a higher position than retailer 2. When the outside advertiser’s profit per click is low, the manufacturer sets the monopolistic wholesale prices and provides zero participation rates to both retailers. When the outside advertiser’s profit per click is medium, the manufacturer lowers wholesale prices but still provides zero participation rates. When the outside advertiser’s profit per click is relatively high, the manufacturer only charges marginal production cost as wholesale prices, and provides positive participation rates to both retailers, so as to keep them in the first two positions.

- Consider the position configuration \((R_1, A)\).

The manufacturer’s optimization problem is,

\[
\max_{\alpha_i, w_i} \quad d_1 \left[ \frac{1 - \alpha_1}{2} (w_1 - c) - \alpha_1 \left( \frac{d_1 - d_2}{d_1} v_A + \frac{d_2}{d_1} \frac{\bar{\theta}(1 - w_2)^2}{4(1 - \alpha_2)} \right) \right]
\]

\[
s.t. \quad \frac{\bar{\theta}(1 - w_1)^2}{4(1 - \alpha_1)} \geq v_A > \frac{\bar{\theta}(1 - w_2)^2}{4(1 - \alpha_2)},
\]

\[
0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2.
\]

The optimal solution is,

\[
\begin{align*}
  w_1^* &= \frac{d_1 c + d_2}{d_1 + d_2}, \quad w_2^* = 1, \quad \alpha_1^* = 1 - \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)^2 v_A}, \quad \alpha_2^* \in [0, 1), \quad v_A \geq \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)^2}, \\
  w_1^* &= 1 - \sqrt{\frac{4v_A}{\bar{\theta}}}, \quad w_2^* = 1, \quad \alpha_1^* = 0, \quad \alpha_2^* \in [0, 1), \quad \frac{\bar{\theta}(1 - c)^2}{16} \leq v_A \leq \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)^2}, \\
  w_1^* &= \frac{1 + c}{2}, \quad w_2^* = 1, \quad \alpha_1^* = 0, \quad \alpha_2^* \in [0, 1), \quad v_A \leq \frac{\bar{\theta}(1 - c)^2}{16}.
\end{align*}
\]

Correspondingly, the manufacturer’s profit is,

\[
\pi_M^* = \begin{cases} 
  \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)} - (d_1 - d_2) v_A & \text{when } v_A \geq \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)^2} \\
  d_1 \left[ (1 - c) \sqrt{v_A \bar{\theta}} - 2v_A \right] & \text{when } \frac{\bar{\theta}(1 - c)^2}{16} \leq v_A \leq \frac{\bar{\theta}(1 - c)^2 d_2^2}{4(d_1 + d_2)^2} \\
  d_1 \frac{\bar{\theta}(1 - c)^2}{8} & \text{when } v_A \leq \frac{\bar{\theta}(1 - c)^2}{16}
\end{cases}
\]

To summarize, in this case, the manufacturer essentially only sells to retailer 1. When the outside advertiser’s profit per click is relatively low, the manufacturer sets the
monopolistic wholesale price and provides zero participation rate to the retailer at the same time. When the outside advertiser's profit per click is medium, the manufacturer lowers the wholesale price but still provides zero participation rate. A lower wholesale price leaves more profit margin to the retailer thus incentivizes its to bid higher so as to keep the first position. When the outside advertiser's profit per click is relatively high, the manufacturer will set a low wholesale price and provide positive participation rate at the same time so as to keep the retailer in the first position.

- Consider the position configuration \( (A, R_1) \).

The manufacturer's optimization problem is,

\[
\begin{align*}
\max_{\alpha_i, w_i} & \quad d_2 \left[ \frac{1}{2} (w_1 - c) - \frac{\theta (1 - w_2)^2}{4(1 - \alpha_2)} \right] \\
\text{s.t.} & \quad v_A > \frac{\theta (1 - w_1)^2}{4(1 - \alpha_1)} \geq \frac{\theta (1 - w_2)^2}{4(1 - \alpha_2)}, \\
& \quad 0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2.
\end{align*}
\]

The optimal solution is,

\[
\begin{cases}
&w_1^* = \frac{c+1}{2}, w_2^* = 1, \alpha_1^* = 0, \alpha_2^* \in [0, 1), \quad v_A \geq \frac{\theta (1-c)^2}{16} \\
&w_1^* = 1 - 2 \sqrt{\frac{v_A}{\theta}}, w_2^* = 1, \alpha_1^* = 0, \alpha_2^* \in [0, 1), \quad v_A \leq \frac{\theta (1-c)^2}{16}.
\end{cases}
\]

Correspondingly, the manufacturer's profit is,

\[
\pi^*_M = \begin{cases}
&d_2 \frac{\theta (1-c)^2}{8}, \quad v_A \geq \frac{\theta (1-c)^2}{16} \\
&\left[ (1 - c) \sqrt{v_A} \frac{\theta}{2} - 2 v_A \right], \quad v_A \leq \frac{\theta (1-c)^2}{16}.
\end{cases}
\]

To summarize, in this case, the manufacturer essentially only sells to retailer 1, and provides a zero participation rate. The manufacturer sets the monopolistic wholesale price when the outside advertiser's profit per click is relatively high; otherwise, it increases the wholesale price thus decreases the retailer's profit margin and its incentive to bid when the outside advertiser's profit per click is relatively low.
Proof of Theorem 7

The equilibrium analysis here parallels with that for the linear contracts above. For each of the three possible position configurations, we will formulate and then solve the manufacturer's channel profit maximization problem.

- Consider the position configuration \((R_1, R_2)\).

The manufacturer's optimization problem is,

\[
\max_{\alpha_i, w_i} d_1 \left[ \theta \frac{1 - w_1}{2} \left( \frac{1 + w_1}{2} - c \right) - \left( \frac{d_1 - d_2}{d_1} \frac{\bar{\theta}(1 - w_2)^2}{4(1 - \alpha_2)} + \frac{d_2}{d_1} v_A \right) \right] + d_2 \left[ \theta \frac{1 - w_2}{2} \left( \frac{1 + w_2}{2} - c \right) - v_A \right]
\]

s.t.

\[
\frac{\bar{\theta}(1 - w_1)^2}{4(1 - \alpha_1)} \geq \frac{\bar{\theta}(1 - w_2)^2}{4(1 - \alpha_2)} \geq v_A,
\]

\[0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2.\]

The optimal solution is,

\[
\begin{cases}
  w_1^* = c, w_2^* = 1 - \frac{d_2}{d_1} (1 - c), \alpha_1^* \in [0, 1), \alpha_2^* = 0, & v_A \leq \frac{d_2^2 \bar{\theta}(1-c)^2}{d_1^2} \\
  w_1^* = c, w_2^* = 1 - 2 \sqrt{\frac{v_A}{\bar{\theta}}} , \alpha_1^* \in [0, 1), \alpha_2^* = 0, & \frac{d_2^2 \bar{\theta}(1-c)^2}{d_1^2} \leq v_A \leq \frac{\bar{\theta}(1-c)^2}{4} . \\
  w_1^* = c, w_2^* = c, \alpha_1^* \in [1 - \frac{\bar{\theta}(1-c)^2}{4v_A}, 1), \alpha_2^* = 1 - \frac{\bar{\theta}(1-c)^2}{4v_A}, & v_A \geq \frac{\bar{\theta}(1-c)^2}{4} .
\end{cases}
\]

Correspondingly, the channel profit is,

\[
\pi_C^* = \begin{cases}
  \frac{d_1^2 + d_2^2 \bar{\theta}(1-c)^2}{4} - 2d_2 v_A, & v_A \leq \frac{d_2^2 \bar{\theta}(1-c)^2}{d_1^2} \\
  d_1 \frac{\bar{\theta}(1-c)^2}{4} + d_2 (1 - c) \sqrt{v_A \bar{\theta}} - (d_1 + 2d_2)v_A, & \frac{d_2^2 \bar{\theta}(1-c)^2}{d_1^2} \leq v_A \leq \frac{\bar{\theta}(1-c)^2}{4} . \\
  (d_1 + d_2) \frac{\bar{\theta}(1-c)^2}{4} - (d_1 + d_2)v_A, & v_A \geq \frac{\bar{\theta}(1-c)^2}{4} .
\end{cases}
\]

To summarize, in this case, the manufacturer sells to both retailers. It sets the wholesale price as the production cost and provides high enough participation rate for retailer 1 to ensure it gets the first position. When \(v_A\) is relatively low, it sets the wholesale price higher than marginal production cost and provides zero participation rate for retailer 2; when \(v_A\) is relatively high, it sets the wholesale price at the marginal
production cost and provides positive participation rate for retailer 2.

- Consider the position configuration \((R_1, A)\).

The manufacturer's optimization problem is,

\[
\max_{\alpha_i, w_i} d_1 \left[ \frac{1-w_1}{2} \left( \frac{1+w_1}{2} - c \right) - \left( \frac{d_1-d_2}{d_1} v_A + \frac{d_2}{d_1} \frac{\theta(1-w_2)^2}{4(1-\alpha_2)} \right) \right]
\]

\[
\text{s.t.} \quad \frac{\theta(1-w_1)^2}{4(1-\alpha_1)} \geq v_A > \frac{\theta(1-w_2)^2}{4(1-\alpha_2)}
\]

\[
0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2.
\]

The optimal solution is,

\[
\begin{cases}
    w_1^* = c, w_2^* = 1, \alpha_1^* \in [1 - \frac{\theta(1-c)^2}{4v_A}, 1), \alpha_2^* \in (0, 1), & v_A \geq \frac{\theta(1-c)^2}{4} \\
    w_1^* = c, w_2^* = 1, \alpha_1^* \in [0, 1), \alpha_2^* \in [0, 1), & v_A \leq \frac{\theta(1-c)^2}{4}.
\end{cases}
\]

Under both cases, the channel profit is,

\[
\pi^*_C = d_1 \frac{\theta(1-c)^2}{4} - (d_1 - d_2)v_A.
\]

To summarize, in this case, the manufacturer essentially only sells to retailer 1. It sets the wholesale price as the marginal production cost, and provides the participation rate high enough to help retailer 1 outbid the outside advertiser. The participate rate does not influence the channel profit.

- Consider the position configuration \((A, R_1)\).

The manufacturer's optimization problem is,

\[
\max_{\alpha_i, w_i} d_2 \left[ \frac{1-w_1}{2} \left( \frac{1+w_1}{2} - c \right) - \frac{\theta(1-w_2)^2}{4(1-\alpha_2)} \right]
\]

\[
\text{s.t.} \quad v_A > \frac{\theta(1-w_1)^2}{4(1-\alpha_1)} \geq \frac{\theta(1-w_2)^2}{4(1-\alpha_2)},
\]

\[
0 \leq \alpha_i < 1, 0 \leq w_i \leq 1, \text{ for } i = 1, 2.
\]
The optimal solution is,

\[
\begin{cases}
    w_1^* = c, w_2^* = 1, \alpha_1^* = 0, \alpha_2^* \in [0, 1), & v_A \geq \frac{\delta(1-c)^2}{4} \\
    w_1^* = 1 - 2\sqrt{\frac{v_A}{\theta}}, w_2^* = 1, \alpha_1^* = 0, \alpha_2^* \in [0, 1), & v_A \leq \frac{\delta(1-c)^2}{4}.
\end{cases}
\]

Correspondingly, the channel profit is,

\[
\pi^*_C = \begin{cases}
    d_2 \frac{\delta(1-c)^2}{4}, & v_A \geq \frac{\delta(1-c)^2}{4} \\
    d_2 \left[ (1-c)\sqrt{v_A\theta} - v_A \right], & v_A \leq \frac{\delta(1-c)^2}{4}.
\end{cases}
\]

To summarize, in this case, the manufacturer essentially only sells to retailer 1, and provides a zero participation rate. The manufacturer sets the wholesale price as the marginal production cost when the outside advertiser’s profit per click is relatively high; otherwise, it increases the wholesale price thus decreases the retailer’s profit margin and its incentive to bid when the outside advertiser’s profit per click is relatively low.
Bibliography


Shiyang Gong, Juanjuan Zhang, Ping Zhao, and Xuping Jiang. Tweets and sales. *Available at SSRN 2461370*, 2014.


Yufeng Huang and Bart J Bronnenberg. Pennies for your thoughts: Costly product consideration and purchase quantity thresholds. 2015.


