THREE ESSAYS ON DIGITAL BUSINESS: PRESCRIPTIVE ANALYTICS FOR NOVEL OPERATIONAL AND STRATEGIC CHALLENGES

by

Mingwen Yang

APPROVED BY SUPERVISORY COMMITTEE:

Vijay Mookerjee, Co-Chair

Zhiqiang (Eric) Zheng, Co-Chair

Varghese Jacob

Harpreet Singh

Feng Zhao

To my family

THREE ESSAYS ON DIGITAL BUSINESS: PRESCRIPTIVE ANALYTICS FOR NOVEL OPERATIONAL AND STRATEGIC CHALLENGES

by

MINGWEN YANG, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

MANAGEMENT SCIENCE

THE UNIVERSITY OF TEXAS AT DALLAS

May 2019

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisors, Dr. Vijay Mookerjee and Dr. Zhiqiang (Eric) Zheng, for their advice, support, and encouragement throughout my years in the PhD program at The University of Texas at Dallas. They were always available whenever I needed help with my research, and always welcomed my questions and concerns. Dr. Mookerjee's depth, broadness, and innovation inspired to keep exploring high-quality research. Dr. Zheng's innovation and vision benefit me to choose research topics always close to industry and practice. I am especially indebted to Dr. Zheng for encouraging me to join the Jindal PhD program.

I am also deeply grateful to have worked with Dr. Varghese Jacob and Dr. Srinivasan Raghunathan. I learned a lot from our research discussions throughout my entire PhD journey. I appreciate the support, comments, and directions they have given me. I am especially grateful for Dr. Raghu's excellent guide on our research on cloud security using game theory. He always provided responsive and extremely valuable advice.

My special thanks Dr. Radha Mookerjee providing valuable feedback on my research, advice on my teaching, excellent coaching during my job search period, and much more.

I'm also thankful to Dr. Harpreet Singh and Dr. Feng Zhao for serving in my committee and for providing feedback on my research. Especially for the social trading chapter, I am indebted to Dr. Zhao's insights and feedback from the finance perspective.

I am sincerely thankful to Drs. Indranil Bardhan, Norris Bruce, Jianqing Chen, Xianjun Geng, Syam Menon, and Sumit Sarkar for their excellent seminar courses, where I learned a lot from their broad knowledge in specific domains and how to apply the methodology I learned from textbook to solve research problems. I also appreciate their valuable feedback on my research and their advice on my job search.

I was lucky to be Research Assistant for one summer semester working with Dr. Sumit Sarkar and Dr. Young Ryu. I learned a lot from our regular research discussions. Especially Dr. Sarkar's rigorous and highly organized attitude toward research and work benefited me a lot.

I thank Drs. Mehmet Ayvaci, Huseyin Cavusoglu, Kutsal Dogan, Atanu Lahiri, Amit Mehra, Shaojie Tang, Yingjie Zhang, Zhe Zhang, Milind Dawande, Ganesh Janakiraman, Nanda Kumar, and Ying Xie for providing valuable feedback on my research and advice on my job search.

I appreciate all of the support from our excellent staff Jamie Selander, Marjorie Wu, Aylin Calisir, Vicki Smith, and Abeer Sumeir during my PhD journey.

I would also like to thank Dr. Hongyu Chen and Dr. Cheng Nie for their great help in the data collection. I also thank all my friends and fellow students in our department as well as in other departments at UT Dallas.

I am grateful to my advisor Dr. Guohua Wan at Shanghai Jiao Tong University, China, where I pursued a masters degree prior to this PhD journey. He was one of the first people who influenced me to pursue a PhD. His attitude toward research and teaching motivated me to apply a PhD program and pursue an academic life.

Last but not least, I would like to express my deepest gratitude to my family for their love and support.

April 2019

THREE ESSAYS ON DIGITAL BUSINESS: PRESCRIPTIVE ANALYTICS FOR NOVEL OPERATIONAL AND STRATEGIC CHALLENGES

Mingwen Yang, PhD
The University of Texas at Dallas, 2019

Supervising Professors: Vijay Mookerjee, Co-Chair
Zhiqiang (Eric) Zheng, Co-Chair

My dissertation provides prescriptive solutions and managerial implications for three novel operational and strategic challenges faced by firms or platforms in online business.

The first problem arises from the need to manage online customer opinions. Online review platforms such as Expedia.com and Tripadvisor.com allow firms to respond to customer complaints. However firms need to carefully decide when to respond to negative reviews. To unravel the underlying mechanics of the problem, I develop a stochastic differential equation model (SDE) that describes the evolution of review ratings over time for a given response strategy employed by the firm. This model is validated using data on online customer reviews and firm responses from two of the world's largest online travel agents. My approach is not just predictive, but more importantly one that can be used in a prescriptive sense, namely, to prescribe a response strategy that controls review ratings in a desired manner. I operationalize the theoretical response strategy in the stochastic model to an operational prescription that a firm can implement and show the applicability of the approach for different business objectives, such as Mean control, Mean-Variance control, and Service-Level control. Finally, I demonstrate the flexibility of the SDE model by extending it to encompass multiple state variables.

The second problem extends the idea of online reputation management to competitive settings. I consider a market consisting of competing firms that participate in a platform such as Expedia or Yelp. Each firm exerts effort to improve its ratings, but in doing so, also influences the mean market rating. The sales of a firm are influenced by its own ratings and the mean rating of the firms in the market. An equilibrium analysis of the mean market rating reveals several insights. A more heterogeneous market (one where the parameters of the firms are very different) leads to a lower mean market rating and higher total profit of the firms in the market. The results can inform platforms to target certain firms to join: Growing the middle of the market (firms with average ratings) is the best option considering the goals of the platform (increase total profit of the firms) and the other stakeholders, namely, incumbents and consumers. For firms, I find that a firm's profit can increase from an adverse event (such as, a reduction in sales margin, or an increase in the cost of control) depending on how other firms in the market are affected by the event. The findings are particularly significant for platform owners who could benefit from growing the platform in a strategic manner.

The third problem addresses a novel Financial Technology (Fintech) phenomenon in social trading. In social trading, less experienced investors (followers) are allowed to copy the trades of experts (traders) in real-time after paying a following fee. This raises the transparency-revenue tension: a dilemma between the need to release trading information transparently versus the risk of followers free riding on such information. I demonstrate the tension using data from a leading social trading platform operating in the Foreign Exchange market. An optimization model is developed to maximize information transparency while respecting a money-at-risk constraint. The performances of three information release policies are compared. Finally, I optimize platform revenue using an optimal release policy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

My dissertation provides prescriptive solutions and managerial implications for three novel operational and strategic challenges faced by firms or platforms in online business.

My second chapter considers a problem that arises from the need to manage customer opinions expressed in online reviews. Given the prevalence and importance of online reviews in todays businesses, managing customer opinions expressed in online reviews is of paramount significance. Online review platforms such as Expedia.com and Tripadvisor.com allow firms to respond to customer complaints. However firms need to carefully decide when and how to respond to negative reviews. On the one hand, inadequate responses can damage the firms reputation. On the other hand, responding too frequently, can dilute the benefit of responding and worse, may inadvertently encourage customers to be opportunistic. For example, customers may intentionally exaggerate their negative experience to take advantage of a firms compensatory rewards. Therefore, the natural question arises: *What is an optimal response strategy for firms to manage their online ratings in a desired manner*? To unravel the underlying mechanics of the problem, I develop a stochastic differential equation model that describes the evolution of review ratings over time for a given response strategy employed by the firm. This model is validated using data on online customer reviews and firm responses from two of the world's largest online travel agents. My approach is not just predictive, but more importantly one that can be used in a prescriptive sense, namely, to prescribe a response strategy that controls review ratings in a desired manner. I operationalize the theoretical response strategy in the stochastic model to an operational prescription that a firm can implement and show the applicability of the approach for different business objectives, such as Mean control, Mean-Variance control, and Service-Level control. Finally, I demonstrate the flexibility of the SDE model by extending it to encompass multiple state variables.

The third chapter extends the idea of online reputation management to competitive settings. An important aspect of online reputation is that it is not fully controllable, owing to its stochastic nature. I consider a market consisting of firms (participating in a platform such as Expedia or Yelp) that compete for sales that are influenced by their own ratings relative to an anchor, e.g., the mean rating of the firms in the market. Each firm exerts effort to improve its ratings, but in doing so, also influences the mean market rating. Given the importance of online reputation competition, there are several challenges firms and platforms continue to face:

- *How much effort should a firm exert to effectively manage online reputation (that evolves stochastically over time) and how does the effort level of competing firms affect this choice?*
- *From the platform's perspective, how should it target new firms to join, without compromising the objectives of incumbent firms and consumers?*
- *For a given market size, how should the platform balance the different kinds of firms in the platform so as to maximize platform goals?*

An equilibrium analysis of the mean market rating reveals several insights. A more heterogeneous market (one where the parameters of the firms are very different) leads to a lower mean market rating and higher total profit of the firms in the market. Our results can inform platforms to target certain firms to join: Growing the middle of the market (firms with average ratings) is the best option considering the goals of the platform (increase total profit of the firms) and the other stakeholders, namely, incumbents and consumers. For firms, I find that a firm's profit can increase from an adverse event (such as, a reduction in sales margin, or an increase in the cost of control) depending on how other firms in the market are affected by the event. Our findings are particularly significant for platform owners who

could benefit from growing the platform in a strategic manner. I model each firm's decision problem as a stochastic control problem where the objective is to maximize discounted profit over a planning horizon. These control problems are connected through a common market belief that represents the mean rating of the firms in the market. The joint actions of the firms generate a mean market equilibrium. I prove that such an equilibrium exists, is unique, and use a simple algorithm to compute its value.

The fourth chapter addresses a novel Financial Technology (Fintech) phenomenon in social trading – an emerging paradigm in the spirit of the sharing economy – enables a trader to share her trading wisdom with other investors. A special type of social trading is copy trading, where less experienced investors (followers) are allowed to copy the trades of experts (traders) in real-time after paying a fee. Such a copy trading mechanism often runs into a transparency-revenue conundrum. On the one hand, social trading platforms need to release traders' trades as transparently as possible to allow followers to evaluate traders. On the other hand, complete transparency may undercut the platform's revenue since followers could free ride. That is, followers could manually copy the delayed trades of a trader to circumvent paying following fees. This study addresses this simple, but fundamental problem by determining the optimal policy for releasing trade information. Of key interest here is the real time value of trade information, which measures using the concept of profit-gap, i.e., the would-be profit difference between the real time and a delayed execution of a trade. Our data shows that profit-gap increases with delay in a concave manner. To provide empirical evidence of the conundrum, I explore how the amount following is affected by delay and users' evaluation activity on the trader (measured by the number of views received by the trader's profile page). Surprisingly, when the number of views is high, a larger delay is associated with a lower amount following.

I propose the notion – money-at-risk – to quantify the possible loss of existing followers (amount following) as a result of releasing delayed trades. The tradeoff between transparency

and money-at-risk is cast as an optimization problem that attempts to maximize information transparency while respecting a money-at-risk constraint. Three information release policies are compared: (1) Uniform Release Policy, (2) Customized Indifference Policy, and (3) Customized Money-at-Risk Release Policy . I demonstrate the performance of the above policies using data from a leading social trading platform operating in the Foreign Exchange market. I also study a Stochastic control formulation that directly optimizes platform revenue. The control is the delay that is calculated as a function of the current amount of money following a trader and the number of views received by the trader's profile page. Besides, the calculated revenue can be incorporated into the ranking algorithm to provide a systematic way to infuse the platform's goals into the ranking of the traders.

CHAPTER 2

PRESCRIBING RESPONSE STRATEGIES TO MANAGE CUSTOMER OPINIONS: A STOCHASTIC DIFFERENTIAL EQUATION APPROACH [1]

2.1 Introduction

In today's online economy, customers are increasingly relying on online reviews, social media and other forms of word of mouth to form opinions on a product or service they intend to purchase. Of these sources, online reviews have been found to have a big impact on a firm's reputation and revenue (Jabr and Zheng, 2014). Marketers have observed that online reviews influence 90% of consumers in their perception and adoption of a product or service (Gesenhues, 2013). The monetary consequences of review ratings on corporate reputation and profitability are significant. For example, Luca (2011) observes that a drop of one-star rating on Yelp translates to a 5% to 9% decrease of a restaurant's revenue. On the positive side, an extra half-star rating causes restaurants to sell out during peak hours 49% more frequently, with larger impacts when alternate information is scarcer (Anderson and Magruder, 2012). In addition, Anderson and Han (2016) find that a one-star improvement in online rating on TripAdvisor increases hotel revenue by 39%.

Negative reviews in particular, play a more prominent role (Chevalier and Mayzlin, 2006). Firms, however, continue to struggle with methods to mitigate or counter the influence of negative reviews. One approach focuses on soliciting positive reviews, for example, by sending emails to encourage positive customers to post reviews after transactions. Since recent reviews are often displayed first, and most customers only read the first few pages of online reviews, negative reviews get crowded out quickly among a large number of positive ones. Some review platforms allow firms to talk to customers off-line to resolve their complaints so

[1] ©2019. Reprinted with permission from INFORMS. Mingwen Yang, Zhiqiang (Eric) Zheng, Vijay Mookerjee, Prescribing Response Strategies to Manage Customer Opinions: A Stochastic Differential Equation Approach, Information Systems Research, 2019.

that they may revise their negative opinions (Banjo, 2012). There are also some companies that actively discourage customers from posting negative reviews. However, as a hotel in New York found out, such a policy could backfire: negative sentiments flooded in afterwards, causing the hotel's rating to plummet down to 1 star with over one hundred negative reviews emerging on Yelp shortly after implementing the policy (DeMers, 2014).

Companies also try to manipulate reviews by fabricating fake ones. It is reported that 10%–15% of online reviews are manipulated or fake, sponsored by companies (Gartner, 2012). Mayzlin et al. (2014) found evidence that some firms manufacture positive reviews for themselves and negative reviews for their competitors. Although manipulated reviews could reward a firm in the short term, manipulation runs the risk of being seen through by consumers with dire consequences for the firm's long term reputation. The U.S. government has begun to stipulate tighter policies to regulate the online review environment (Rushe, 2013). Firms that fabricate fake reviews will face severe penalties in the form of government fines, possibility of lawsuits, and loss of reputation. Online review platforms such as Yelp ensure the integrity of its reviews by filtering out illegitimate, fake reviews.

Many savvy firms have found that user opinions can be managed more effectively by publicly responding to reviews. Evidence suggests that a calibrated firm response is often an effective way to tackle negative reviews. It is reported that after seeing a carefully crafted response to a negative review, 71% of customers change their perception of the brand (Meyer, 2013). Proserpio and Zervas (2017) find that hotels that actively respond to reviews receive an average increase of 0.12 star rating in TripAdvisor. Gu and Ye (2014) show that publicly responding to negative reviews not only addresses a particular customer's concerns, but also helps limit the spillover of negative opinions to other customers.

Firm responses can affect future reviews in several ways. People are more likely to write negative reviews when their experience of a product or service is different (worse) than what they have experienced (Moe and Schweidel, 2012; Ho et al., 2017). Moe and Schweidel

(2012) observe sequential dependency among reviews, that occurs because a review writer considers what other customers have experienced and reported on a matter before writing a review. Once a complaint has been reported by other customers and acknowledged by the firm, repeating the same complaint does not have much information value. Importantly, if customers were to voice a complaint on an issue that has already been raised and responded to, they would feel obliged to express their complaint differently – referred to as the "beg to differ" effect by Moe and Schweidel (2012).

Responding to reviews is also likely to decrease the occurrence of future negative reviews because consumers feel that their reviews will be closely scrutinized (Proserpio and Zervas, 2017). That is, responding increases the "cost" of leaving a negative review, especially a flaky or unfounded one. For example, consumers with negative opinions need to justify their negative posts by providing sufficient details of their complaints, thus increasing the length of negative reviews (Proserpio and Zervas, 2017). We will elaborate further on how management responses influence future reviews in Section 3.1 and demonstrate it empirically in Section 4.2.

Our focus in this study is on the design of a response strategy to manage user opinions. A full response strategy – one that responds to every review – will likely be either too costly, or ineffective if the responses are not adequate. Thus firms should respond in a selective manner, but when they do, the response must be adequate to resolve the customer's problem. To unravel the mechanics of how review ratings evolve over time, we develop a Stochastic Differential Equation (SDE) method to model the underlying review rating generation process. This model captures how average ratings react to the arrival of new reviews as well as the firm's response strategy.

In the SDE model, the stochastic variable of interest (or the *state*), is a measure of customer value proposition, i.e., a customer's perception of quality of a product or service

at any given point in time.[2] The state is operationalized as the average of a certain number of recent ratings provided in customer reviews. The change in the state during a small time interval is decomposed into a deterministic component and a stochastic component. The deterministic component combines the firm's response strategy and the influence of newly arrived reviews. The stochastic component consists of a random noise term that cannot be explicitly observed or explained. With some modifications, our proposed SDE model can be transformed into a form similar to the Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985).

We solve our proposed SDE model to estimate the stochastic process of the state as a function of time and other primitives, including the firm's response strategy. This stochastic process is validated empirically using data on online customer reviews and firm responses from two of the world largest travel agents. Compared with traditional time series forecasting methods such as Autoregressive Moving Average (ARMA) (Box et al., 2015), Generalized Autoregressive Conditional Heteroscedasticity (GARCH) (Engle, 1982), Moving Average (MA) (Brown, 2004), Exponential Smoothing (ES) (Brown, 2004), and Naive Method (NM), our approach achieves on-par or superior predictive performance.

However, the key strength of our approach is that it is able to recover the distribution of future review ratings as a function of the response strategy used by the firm. Our approach therefore, is not just predictive, but also one that can be used in a prescriptive sense, namely, to prescribe a response strategy that controls the rating generation process in a desired manner. We map the theoretical control in the stochastic model to an operational prescription that a real firm can implement. We next demonstrate the use of our operational prescription to achieve different control objectives that managers may have to influence customer opinions as they evolve over time. Finally, we show how the SDE approach can be extended to encompass multi-state variables. This further enhances the model's prescriptive ability

[2]We hereafter use consumers' perception of quality interchangeably with the notion of customer value proposition in the marketing literature (Anderson et al., 2006), that describes the overall experience that potential consumers could have after purchasing the product.

and adds new practical insights because it enables a firm's response strategy to react to a change in multiple conditions (states).

2.2 Literature Review

We review literature in the problem domain being studied (customer opinion management) as well as the main methodology being used (stochastic differential equations).

2.2.1 Customer Opinion Management

There is abundant literature on customer opinions expressed through online reviews and the impact on product sales (Chevalier and Mayzlin, 2006; Forman et al., 2008). However, there is relatively less attention paid on how customer opinions can be managed over time. Among the few studies on the subject of managing opinions, some researchers have investigated how customer opinions *spread*, i.e., how one customer's opinion could influence the opinion of subsequent customers (Moe and Schweidel, 2012; Ho et al., 2017). Other researchers address the problem of online review manipulation, where firms act as customers and fabricate reviews to inflate their own reputation while damaging those of their competitors'. Dellarocas (2006) and Mayzlin (2006) analytically model manipulation strategies used by firms. Mayzlin (2006) studies a firm's incentives to generate anonymous promotional messages, yielding a unique equilibrium wherein firms with low quality products engage in more promotional chat. Manipulation is documented in Mayzlin et al. (2014), Anderson and Simester (2014), and Luca and Zervas (2016). Mayzlin et al. (2014) compare ratings of 3,082 U.S. hotels on TripAdvisor.com with those on Expedia.com. They find that magnitude of manipulation on Expedia.com is much lower than that on TripAdvisor.com due to the fact that Expedia, unlike Tripadvisor, requires users to have had an actual purchase before posting a review. Within the scope of opinion management, a small but growing stream of research has begun to examine how firms can mitigate the reputational harm of negative reviews, such as by

inducing buyers to revoke negative feedback (Ye et al., 2014) or by instituting a reward-for-feedback mechanism (Li et al., 2016).

Our study considers a firm response mechanism that is different from the extant literature on review manipulation. In our setting, firms choose to directly respond to some customer reviews; the response is public and visible to everyone. Some papers attempt to address this phenomenon, with emphases on studying whether a firm's response affects subsequent review ratings (Gu and Ye, 2014; Proserpio and Zervas, 2017; Wang and Chaudhry, 2018). Proserpio and Zervas (2017) examine the impact of responses on customer reviews by contrasting two hotel review platforms: one regularly responds but the other does not. They find that the responding hotel reaped an average increase of 0.12 star in the rating after starting to respond. Gu and Ye (2014) measure the impact of management responses on customer satisfaction and find that online management response is highly effective among low-satisfaction customers but has limited influence on others. Wang and Chaudhry (2018) conduct a natural-experiment to examine the influences of managers' responses to negative reviews and observe an increase in customers' stated satisfaction after receiving responses. A recent study finds that responding to reviews could decrease future ratings since it might encourage more negative reviews (Chevalier et al., 2018). In such cases, the benefit of responding would be to satisfy the customer who wrote the review, but the downside of responding would be that it could attract new negative reviews. In such situations, i.e., management response has two-sided effects; cost is not the only reason to limit the frequency of responding to reviews. Additionally, Gunarathne et al. (2017) finds that how promptly a firm responds to a review is affected by the customer's social influence (popularity) and sentiment toward the firm. Ma et al. (2015) analyze how a firm's response (service intervention) changes individual customers' decision to voice out (post a message on Twitter).

Some recent papers focus on teasing out the effect of management response on a firm's financial performance (Lee et al., 2016; Kumar et al., 2017). Lee et al. (2016) observe

that online WOM metrics, e.g. valence and volume, moderate the effect of management response on a hotel's revenue; their findings suggest that hotels' response strategy should factor in the level of online WOM metrics. Kumar et al. (2017) document the significant role of management response on the performance of the focal firm's business, as well as the performance of the nearby competitors' businesses (i.e. the spillover effect).

Unlike the emphasis of previous work, our study focuses on the best response strategy to manage online ratings. We focus on what review should a firm respond to, thus addressing a gap raised in the literature (e.g. Proserpio and Zervas (2017)). Our study is also related to the complaint management literature in Marketing because responding to customer reviews can be regarded as a special type of a complaint management (or defensive marketing) strategy. Fornell and Wernerfelt (1988) analyze incentives to manage complaints and characterize industries where complaint management is likely to be used. They show that complaint management can lower the total marketing expenditure by substantially reducing the cost of advertising (offensive marketing).

To summarize, our study differs from the past research both in terms of the problem being solved, and the research methodology being used. To our knowledge, this study represents one of the first attempts to prescribe an appropriate response strategy to manage user opinions. From a methodological perspective, we develop a stochastic differential equation model to prescribe the review data generation process. In contrast, extant literature has either resorted to analytical modeling (e.g. Dellarocas (2006)) or reduced form empirical analysis (e.g. Gu and Ye (2014), Proserpio and Zervas (2017), Wang and Chaudhry (2018)). A key distinction of our stochastic differential equation model is that we model the stochastic nature of the review data generating process, namely, how review ratings evolve over time after accounting for the response strategy used by the firm.

2.2.2 Stochastic Differential Equations

In our chapter, we develop a Stochastic Differential Equation (SDE) model of the dynamics of review ratings in the presence of a response strategy used by the firm. SDE has been applied in finance to model the time series of stock price movements, where randomness is inevitable. The well-known Black-Scholes equation (Black and Scholes, 1973), modeling the price of a European call option, uses the SDE methodology. A stochastic process, $X_t, t \geq 0$, models a random variable of interest, that varies continuously and stochastically through time. We model the stochastic rating process as a Markov process, where the probability distribution of the future value depends only on its current value, that subsumes the effect of past values of the process. Randomness is captured by a Wiener process (the continuous limit of random walk), a fundamental building block for randomness in stochastic processes. A stochastic process, $W_t, t \geq 0$, is defined to be a Wiener Process if $W_0 = 0, W_t$ has stationary and independent increments, and W_t is normally distributed with mean zero and variance dt for every t (Ross, 2014).

Geometric Brownian Motion (GBM) is one of the classic SDE frameworks to model price movement in finance (Dixit et al., 1994). A stochastic process X_t is said to follow a GBM if it satisfies $dX_t = b_1 X_t dt + b_2 X_t dW_t$, where dX_t is the incremental change in X_t and W_t is a Wiener process. In the GBM equation, the first term is used to model deterministic trends, referred to as the drift process with the parameter b_1; the second term models randomness, referred to as the diffusion process with the parameter b_2. Often there are special characteristics of interest in time series data such as the mean reversion property, which allows the state variable to fluctuate around one specific level (Dixit et al., 1994), e.g. in the movement of stock prices (Poterba and Summers, 1988). One of the simplest mean-reverting processes, called the Ornstein-Uhlenbeck process, follows $dX_t = \alpha(\mu - X_t)dt + \sigma dW_t$, where α is the speed of reversion, and μ represents the mean level of X_t. If X_t is greater (less) than μ, it is more likely to fall (rise) over the next short interval of time. The discrete format for

the continuous mean reverting process is equivalent to the first-order autoregressive process AR(1) (Dixit et al., 1994). We use a stochastic process similar to the Ornstein-Uhlenbeck process in this study.

We next develop and operationalize the proposed SDE model of how review ratings evolve over time.

2.3 Stochastic Model of Review Ratings

In this section, we first provide a theoretical background for the impact of responses on reviews. Next, we present the SDE model for review ratings.

2.3.1 Theoretical Background

There is considerable evidence that responses help boost ratings, as reported in the popular press and industry reports. For example, a TripAdvisor survey shows that 84% users feel that appropriate management responses to negative reviews improve their impression of the hotel (TripAdvisor, 2012). Another study by a global customer experience management leader reports that hotels with the highest responsiveness to social media outperform competitors in their overall social reputation (Medallia, 2015).

From a theoretical point of view, there are several ways that responses could impact ratings. In general, people are more likely to write negative reviews when their experience of a product or service is different (in this case, worse) than what they have experienced (Moe and Schweidel, 2012; Ho et al., 2017). As shown in Moe and Schweidel (2012), a review writer often considers what other customers have reported on the matter before writing a review. Once a complaint has been reported by other customers (and acknowledged by the firm), there does not appear much information value in repeating the same complaint in the same manner. If customers were to voice a complaint on an issue that has already been raised and responded to, they would feel obliged to express their complaint differently than what

previous reviewers have done. Additionally, customers would also feel the need to describe the complaint in more detail. We empirically verify the above assertions in Section 4.2.

Thus, when a negative review arrives and gets responded to successfully, the impetus for a reviewer to write on the (same) negative issue is lowered or suppressed, leading to improved future ratings. We believe, therefore, that a response prevents, or at least reduces, negative spillovers. A firm's response to a negative review reflects the firm's willingness and intention to solve the problem, thus creating goodwill among other customers who might otherwise have written a negative review. Responses also enable the firm to clarify misunderstandings, apologize for mistakes, and explain causes of sub-par service, thus curbing the spread of a negative event. Finally, responses allow the firm to resolve problems that other customers might encounter. Thus, responding helps curb future negative reviews and hence boosts ratings.

2.3.2 Model Description

We consider a typical online review setting, where reviews and ratings arrive in a chronological sequence forming time series data. At any point of time, we capture the notion of the *state* (consumers' perception of product quality) as the moving average of the last n most recent ratings.[3] To model the stochastic nature of the ratings process, we use the moving average of recent ratings as the state variable of interest. Another choice could have been the cumulative average. However, the cumulative average hardly changes over time, making it a poor candidate to be a state variable. Consumers also deem recent reviews to be more relevant: according to a recent industry report (BrightLocal, 2015), 44% of consumers maintain that a review must be written within 1 month to be relevant. Academic research

[3]In this study, we set the sliding window size $n = 20$. This is a reasonable choice for our data given the fact that there were 10 reviews displayed in a page and that most users do not browse more than two pages of reviews (Pavlou and Dimoka, 2006). We also experimented with different window sizes (e.g. $n = 10$) and the results remained qualitatively the same, as shown in Appendix A.

also finds that the most recent reviews play a more important role on sales than cumulative ratings (Duan et al., 2008). For these reasons, we construct our state variable using the moving average of the most recent ratings.[4]

In a time series data, a moving average is commonly used as a measure of the state (De Gooijer and Hyndman, 2006). For example, a moving average of price measures how a stock is trending, whereas in economics, moving average is commonly used to study gross domestic product, employment rates, and other macroeconomic variables (Brown, 2004). In our setting, the state of the system is affected by the arrival of new reviews as well as responses the firm may provide to improve customer value proposition.

The goal in this section is to devise a model that captures the evolution, over time, of the state of the system (x_t). To model the data generating process, we start with a simplified micro-structure: the change of the state (dx_t) in a small time interval from t to $t + dt$. The change in the state consists of a deterministic component and a random component.

The deterministic component models the expected change (drift) in the state, $\mathbb{E}(dx_t)$, as a function of the arrival of reviews and firm responses, if any. We assume that the arrival of reviews follow a Poisson process with rate λ.[5] During a small time interval, the probability of arrival of one review is λdt. A review can be either negative or positive. Let p (or $1 - p$) be the probability that a review is negative (positive). The impact of a negative review on the state depends on a damage parameter β and the current state of the system x_t. The negative impact is larger for higher values of x_t; or conversely, when customers already had a low opinion, a negative review causes less damage. In the extreme case when $x_t = 0$, there is no further damage. On the other hand, the impact of a positive review depends on a boost parameter ρ and $(b - x_t)$, the difference between the highest possible quality perception b and the current state x_t. When the perception of quality is already high, a positive review

[4]The ordering by time is the default setting in the review page.

[5]We will validate our Poisson Process assumption and other modeling assumptions in Section 5.

does not boost it as much. In the extreme case when $x_t = b$, there is no further gain possible of a positive review. The basic idea is that consumers' perception of quality changes more when "new" information arrives. For a product or service that already has low quality in the eyes of consumers, another negative review will not diminish the perception by much; by the same token when the quality perception is already high, it will not increase much if a new positive review arrives. For example, when a discount airline receives another complaint about its poor service, it will likely have less of an impact than when a premium airline is reported to have a case of bad service.

Finally, let the firm use a damage control effort of α, associated with responding to reviews. The impact of this effort on the state is $\alpha(b - x_t)$. When the perception of quality is high, the impact of damage control diminishes. Note that it is possible that management responses may incur a negative impact on future review ratings. That is the sign of α could possibly be negative, and when it happens, this indicates that management response does not necessarily improve future ratings (in which case the firm should simply stop responding). Taking together the impact of all the three driving forces (the impact of positive review, the impact of negative review, and the impact of damage control in the form of management responses) on x_t, we can write

$$\mathbb{E}(dx_t) = (\lambda(1-p)\rho(b - x_t) - \lambda p\beta x_t + \alpha(b - x_t))dt$$

Collecting the terms and rewriting we get the form,

$$\mathbb{E}(dx_t) = k_1(k_2 - x_t)dt$$

where, $k_1 k_2 = \rho\lambda(1-p)b + \alpha b$ and $k_1 = \alpha + \beta\lambda p + \rho\lambda(1-p)$.

The stochastic component (diffusion) of the change in perception of quality (dx_t) is modeled as $\sigma\sqrt{b - x_t}dW_t$, where dW_t is the Wiener process used to capture white noise or randomness; $dW_t \sim N(0, dt)$. The parameter σ influences the magnitude of the random

component, and the term $\sqrt{b - x_t}$ ensures the state variable not to exceed the upper bound, b. When the state variable touches b, the diffusion term dissolves. Square root processes are commonly used to model stochastic movements (Brown, 2004). As $(b - x_t)$ becomes very small, the square root term diminishes slower than a linear structure, implying that the stochastic component continues to have a material impact even as x_t approaches b. Also, the square root structure is one among the few SDE structures that lends itself to closed form solutions. Table 2.1 lists the parameters in our model. To summarize, we model the change in state using the following SDE:

$$dx_t = k_1(k_2 - x_t)dt + \sigma\sqrt{b - x_t}dW_t \tag{2.1}$$

Table 2.1. Notation

Parameter	Definition
λ	Review Arrival Rate
p	Probability of Negative Review
β	Damage Parameter
ρ	Boost Parameter
α	Damage Control Effort
σ	Magnitude of the Random Component

Regarding equation (2.1), in the long run (as $t \to \infty$), the steady state mean is obtained by setting and solving $\mathbb{E}(dx_t) = 0$. Thus, the steady state (long run) mean of perception of quality (k_2) is given by

$$\mathbb{E}(x_t) = k_2 = \frac{b}{1 + \frac{\beta\lambda p}{\alpha + \rho\lambda(1-p)}} \tag{2.2}$$

It is clear from the above that the steady state mean increases with the damage control effort α. Also, if no damage control effort is exerted by the firm to counter negative reviews ($\alpha = 0$), the steady state mean is likely to go to zero when reviews are predominantly negative. That is, the steady state mean will tend to zero if there is no damage control effort ($\alpha = 0$) and the impact of negative reviews is much larger than that of positive ones ($\beta p \gg \rho(1 - p)$). This

negative trend in posted ratings over time is consistent with observations in the literature (Li and Hitt, 2008; Moe and Schweidel, 2012). Generally speaking, firms would like to keep the perception of quality at a high level, and perhaps would also like its fluctuations to remain within a relatively small interval.

The SDE model in equation (2.1) can be transformed into a standard form using the linear transformation $y_t = b - x_t$. Then using Ito's Lemma on the function $F(x_t, t) = b - x_t$, we get

$$\begin{aligned} dy_t &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dx_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(dx_t)^2 \\ &= -dx_t \end{aligned}$$

Hence,

$$\mathbb{E}(dy_t) = -\mathbb{E}(dx_t) = -k_1(k_2 - x_t)dt = -k_1(k_2 + y_t - b)dt = k_1(k_3 - y_t)dt \tag{2.3}$$

In equation (2.3), $k_3 = b - k_2$. Thus the transformed SDE model becomes:

$$dy_t = k_1(k_3 - y_t)dt + \sigma\sqrt{y_t}dW_t \tag{2.4}$$

We can see that the structure of the SDE model in equation (2.4) is in the form of the CIR model. Cox et al. (1985) noted that the distribution of y_t given y_u for some $u < t$ is, up to a scale factor, a non-central chi-square distribution. The expectation and variance for y_t, given the initial value y_0 are

$$\mathbb{E}(y_t|y_0) = y_0 e^{-k_1 t} + k_3(1 - e^{-k_1 t}) \tag{2.5}$$

$$\mathbb{V}(y_t|y_0) = y_0\frac{\sigma^2}{k_1}(e^{-k_1 t} - e^{-2k_1 t}) + \frac{k_3\sigma^2}{2k_1}(1 - e^{-k_1 t})^2 \tag{2.6}$$

The CIR process can also be represented as a sum of squared Ornstein-Uhlenbeck processes. This provides one way to derive the transition density of the CIR process. Readers can refer to Chalasani and Jha (1997) for more details. In the Model Estimation section, we provide an explicit form of the transition density of the CIR process.

2.4 Exploratory Investigation of Data

In this section, we describe our data and the results of an exploratory investigation to provide supporting evidence of the positive effect of management response on future review ratings.

2.4.1 Data

Our main data source comes from Ctrip.com, the leading online travel agent aggregator in China, accounting for more than half of the market share in the online travel market. Ctrip offers a variety of tours (products), in which the travel agent provides tourism services including itinerary planning, hotel accommodation, transportation, guided tour service, etc. Customers book these tours and then post reviews at Ctrip.com. A snapshot of the page for reviewers to write reviews is presented in Appendix A, where the review rating is on the scale of 1 to 5. These tours are offered by different travel agents where Ctrip itself is the largest travel agent on Ctrip.com. A firm (travel agent) can only respond to customer reviews of its own tours. Figure 2.1 provides a browser (Chrome) translated snapshot[6] of a sample customer review page from a Ctrip tour with responses to these reviews.

We obtained the data for 117 random tours from April 2012 to June 2014. The average number of reviews per tour is 528 with a minimum value of 92 and a maximum value of 7,035. Review ratings for most tours lie within a narrow range. For example, among the 117 tours in the Ctrip data, more than 90% tours have an average rating between 4.2 and 4.8. Thus, in this data, a small numerical difference can be significant, e.g., a 0.1 difference corresponds to a difference of more than 16% of the range. For each tour, we collected customer review information for each posting, including a unique identifier for the reviewer, review date, review rating (from 1 to 5), and review text. If the review received a response

[6]The URL is http://vacations.ctrip.com/grouptravel/p93390s0-comment.html (Last accessed on Oct 20th, 2017).

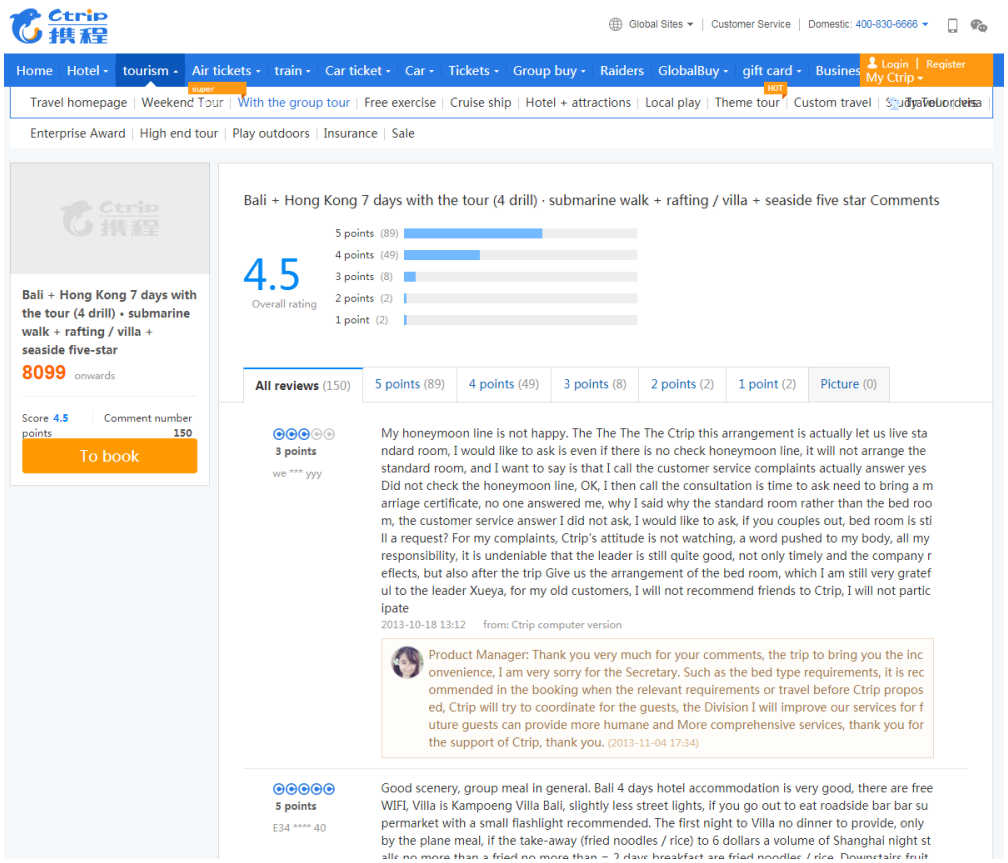

Figure 2.1. Translated Snapshot of a Firm's Response to Customer Reviews on Ctrip.com

from the firm, we record the date of response and the text of the response. Table 2.2 presents the frequency distribution of review ratings.

Table 2.2. Distribution of Customer Review Ratings and the Fraction of Reviews with Responses

Review Rating	No. of Reviews	Percentage	No. of Responses	Fraction with Response
1	538	0.87%	242	44.98%
2	356	0.58%	165	46.35%
3	2,588	4.18%	433	16.73%
4	15,129	24.46%	749	4.95%
5	43,247	69.91%	336	0.78%
Total	61,858	100.00%	1,925	3.11%

The last column in Table 2.2 presents the fraction of customer reviews which received responses by the star-rating of the tours. As can be clearly seen, there is more response activity when the ratings are low, indicating that firms are selective with their use of a response. The last row in Table 2.2 provides the total number of responses, the total number of reviews, and the proportion of reviews receiving responses.

Next, we label a review as negative or positive in a manner described below. If the rating associated with a newly arrived review is above a certain threshold, the review is considered positive, and otherwise negative. Since the review ratings in this industry are generally high (above 4), we choose a relatively high threshold: a rating is considered to be negative if it is less than or equal to 4, and positive otherwise.[7]

Prior studies show that besides the review rating, potential customers heavily rely on the actual review text when making their purchase decisions (Chevalier and Mayzlin, 2006). In Appendix A, we conducted a robustness check by mining the review text using sentiment analysis as an alternative way to identify positive and negative reviews. We obtain similar estimation results compared to those presented in Section 5.

2.4.2 Exploratory Study

To study the impact of management response on future ratings, we start with a model-free analysis to examine how future ratings evolve with versus without a response. First, suppose a review (say, r_j) does not receive a response. The average of the next n (window size) review ratings is

$$\frac{1}{n}\sum_{i=j+1}^{j+n} r_i.$$

[7]Alternatively, if we were to re-label a rating of 4 as positive (instead of negative) and ratings of 1, 2, or 3 as negative, the estimates for the various parameters will be affected. However, the predictive performance of the SDE approach and the response strategy will remain largely unaffected.

To exclude any possible pre-existing trend, we calculate the difference of the mean review ratings between the n pre- and post- reviews with respect to review j as

$$\Delta r_j = \frac{1}{n}\sum_{i=j+1}^{j+n} r_i - \frac{1}{n}\sum_{i=j-n}^{j-1} r_i.$$

We repeat the above calculation for each review with no response, and then compute the average Δr_j. The grand mean of Δr_j across all the tours is denoted as $\mathbb{E}(\Delta r_j)$. We split r_j to negative review ratings (with ratings less than or equal to 4) and positive ones (with ratings equal to 5) and calculate $\mathbb{E}(\Delta r_j)$ under each situation.

Following the same calculation, we then compute $\mathbb{E}(\Delta r_j)$ for the reviews with a response. In both cases (with and without response), we only consider those reviews where the pre- and post- reviews did not contain a response to expunge possible confounds.

Table 2.3 tabulates the values of $\mathbb{E}(\Delta r_j)$ for the two cases by varying n as 20, 10, 3, and 1. To empirically verify whether the mean values of the cases - with versus without response - are statistically different or not, we conduct t-test for each n value with p value shown in Table 2.3. In sum, Table 2.3 demonstrates that the impact of responding is more when the current rating is low (rating of 1, 2, 3, 4). However, there is no significant impact of a response when the rating is high (rating of 5). The effect of responding is quite consistent across different window sizes n, but the effect in the case of window size of 1 is only marginally significant at the 0.1 level. Therefore, a response may not have an immediate effect (e.g., it may not improve the very next rating).

Table 2.3. Improvement of Ratings after Response

Window Size	Without Res. to Neg. Rating	With Res. to Neg. Rating	p value	Without Res. to Pos. Rating	With Res. to Pos. Rating	p value
20	0.045 (4,077)	0.079 (211)	0.013	0.048 (15,519)	0.041 (65)	0.742
10	0.064 (7,297)	0.116 (495)	0.000	0.065 (24,825)	0.071 (126)	0.807
3	0.065 (12,225)	0.099 (1,060)	0.042	0.079 (38,015)	0.106 (232)	0.392
1	0.057 (14,516)	0.099 (1,385)	0.093	0.089 (42,121)	0.166 (289)	0.102

Notes. The number of observations is reported in parenthesis.

Having empirically demonstrated that there is indeed a positive impact of responding to reviews, we probe deeper to shed light on *how* responses may alter the way reviewers write reviews. As we discussed earlier in Section 3.1, if customers were to voice a complaint on an issue that has already been raised and responded to, they would feel obliged to express their complaint differently than what previous reviewers have done. They would also feel the need to describe the complaints in more detail, i.e., better "justify" their complaints. This surfaces two, empirically testable, questions:

1. *Does a response affect the content of negative reviews after the response?*

   We found that when there is a response (treatment group), the content similarity of negative reviews before and after the response is lower than the similarity when there is no response (control group).[8] To empirically verify whether the mean values of the treatment and control group are statistically different or not, we conduct t-test for each window size. These results are presented in Table 2.4. For example, in Table 2.4, for a window size of 20, the average similarity significantly decreased ($p = 0.001$) from 0.528 (without response) to 0.451 (with response). We therefore believe that review writers face the additional burden of explaining the point differently, if they were to write a negative review after a response.

Table 2.4. Average Similarity Scores of Negative Reviews

Window Size	Without Response	With Response	p value
20	0.528 (0.184; 19,596)	0.451 (0.199; 276)	0.001
10	0.378 (0.173; 32,122)	0.347 (0.158; 621)	0.003

Notes. The standard deviation and the number of observations are reported in parenthesis.

2. *Does a response affect the length of negative reviews after the response?*

   We found that when there is a response (treatment group), the length of negative reviews after the response increases compared with that of the no response case (control

[8] In Appendix A, we describe how to calculate content similarity from the review texts.

group). To empirically verify whether the mean values of the treatment and control group are statistically different or not, we conduct t-test for each window size. These results are presented in Table 2.5. For a window size of 20, the average length significantly increased ($p < 0.001$) from 41.2 words (without response) to 45.3 words (with response). We believe that this increased length reflects the additional burden negative review writers have when writing a negative review after a response.

Table 2.5. Average Length of Negative Reviews

Window Size	Without Response	With Response	p value
20	41.2 words (19,596)	45.3 words (276)	0.000
10	42.5 words (32,122)	47.3 words (621)	0.000
3	43.9 words (50,240)	47.9 words (1,292)	0.002

Notes. The number of observations is reported in parenthesis.

To further mitigate the concern whether responses really matter, we establish a more causal link between a firm's response and future ratings. A direct, field experiment with Ctrip or Expedia would have been most helpful. However, given the unavailability of such unique data, the next best alternative was to obtain casual evidence through a quasi-experimental analysis. We use difference-in-differences (DID) analysis in combination with propensity score matching (PSM) to establish the link between the firm's response and future review ratings. Once again, the analysis shows that responses indeed boost future ratings (Please refer to Appendix A for details on the quasi-experiment).

2.5 Model Estimation

We apply the Maximum Likelihood Estimation (MLE) procedure to recover the parameters in our SDE model. The parameters need to be estimated are the response effort (α), damage (β), boost (ρ), arrival rate (λ), negative review probability (p) and magnitude of the stochastic component (σ). The value of the upper bound for the rating (b) is fixed to be 5 since

this is the value of the highest rating allowed by the review system. The unit of analysis is a specific tour, i.e., we estimate the above parameters for each tour. To apply MLE, we first need to specify the probability density function of y_t, which has been originally derived in Feller (1951). For a given value of y_t at time t, the density of y_{t+s} at time $t+s$ is

$$p(y_{t+s}|y_t;\alpha,\rho,\beta,\sigma,\lambda,p,s) = ce^{-u-v}(\frac{v}{u})^{\frac{q}{2}} I_q(2\sqrt{uv}) \tag{2.7}$$

where

$$c = \frac{2k_1}{\sigma^2(1-e^{-k_1 s})},$$
$$u = cy_t e^{-k_1 s},$$
$$v = cy_{t+s},$$
$$q = \frac{2k_1 k_3}{\sigma^2} - 1,$$
$$k_1 = \alpha + \rho(1-p)\lambda + \beta p\lambda,$$
$$k_3 = \frac{5\beta p\lambda}{\alpha + \rho(1-p)\lambda + \beta p\lambda},$$

and

$$I_q(2\sqrt{uv}),$$

is the modified Bessel function (Lizorkin, 2001) of order q. k_1 measures the speed of reversion, and k_3 is the steady state mean of y_t. The stochastic quantity $2cy_{t+s}$ follows a non-central chi-squared distribution with the non-centrality parameter $2u$ and degrees of freedom $2q+2$ (Cairns, 2004).[9] For simplicity, we define $\theta \equiv (\alpha,\rho,\beta,p,\lambda,\sigma)$. The log-likelihood function for y_t with N observations is

$$\ln L(\theta) = \sum_{i=1}^{N-1} \ln p(y_{t_{i+1}}|y_{t_i};\theta,\Delta t) \tag{2.8}$$

[9]Later in Section 7 where we discuss the Service-Level Control, we map $P\{x_i \geq d\}$ to $P\{2cy_i \leq 2c(5-d)\}$, so that we are able to directly use the formula (CDF) of non-central chi-square distribution to calculate the probability.

Plugging equation (B.7) into equation (B.8) yields

$$\ln L(\theta) = (N-1)\ln c + \sum_{i=1}^{N-1}[-u_{t_i} - v_{t_{i+1}} + 0.5q \ln \frac{v_{t_{i+1}}}{u_{t_i}} + \ln I_q(2\sqrt{u_{t_i} v_{t_{i+1}}})] \tag{2.9}$$

where $u_{t_i} = cy_{t_i}e^{-k_1\Delta t}$ and $v_{t_{i+1}} = cy_{t_{i+1}}$. The Maximum Likelihood estimate $\hat{\theta}$ is solved by maximizing the log-likelihood function described in equation (2.9) over its parameter space:

$$\hat{\theta} = (\hat{\alpha}, \hat{\rho}, \hat{\beta}, \hat{p}, \hat{\lambda}, \hat{\sigma}) = \arg\max_{\theta} \ln L(\theta).$$

We then describe how the values of the review arrival rate λ and the negative review probability p are estimated from the data. For each tour, we observe the timestamp of every review. Taking the first difference along two consecutive reviews yields the inter-arrival time of reviews. The reciprocal of its average is the expected review arrival rate λ. We empirically verify with our data that the inter-arrival time of reviews approaches an exponential distribution, indicating reviews arrive according to a Poisson Process with rate λ. The expected probability of negative review p is inferred by counting the number of negative reviews over the total number of reviews for each tour. We also empirically verify that negative reviews arrive according to a Poisson Process with rate $p\lambda$, consistent with the theory that the split of a Poisson Process is also a Poisson Process.

Next we validate our model assumption using a representative tour with id 5106. Figure 2.2(a) plots the histogram of the inter-arrival time (days) of reviews. Figure 2.2(b) presents an empirical Q-Q plot of the observed empirical quantiles of the inter-arrival time versus the theoretical quantiles of the exponential distribution with mean value $\frac{1}{\lambda} = 0.627$ (1 review every 0.627 days). Figure 2.2(b) shows that the dotted line approximates the (straight) solid line, indicating that the review inter-arrival time fits well with an exponential distribution with mean 0.627.[10] Thus our assumption of Poisson review arrival with rate $\frac{1}{0.627} = 1.595$ (1.595 reviews per day) is reasonable.

[10]Note that the density is heavily left concentrated with a long tail on the right end. So the relative misfit on the long tail part carries less weight.

Since we have the timestamp of every negative review, we are able to compute the inter-arrival time of negative reviews as well. Figure 2.2(c) depicts the histogram of the inter-arrival time of negative reviews, and Figure 2.2(d) delineates the Q-Q plot for the case of negative reviews with a mean value 2.121 for the theoretical exponential distribution. The agreement between the two lines indicates that the inter-arrival time of negative reviews approximates the exponential distribution with mean 2.121. Therefore, negative reviews arrive according to Poisson Process with rate $\frac{1}{2.121} = 0.4714$ (0.4714 negative reviews per day). Note that the probability of negative reviews for tour id 5106 is $p = 0.298$, and the estimated negative reviews arrival rate from the review arrival rate and the probability of negative reviews equals to $p\lambda = \frac{0.298}{0.627} = 0.4753$ (0.4753 negative reviews per day), which is very close to 0.4714.

We further calculate the Theil's U index (Theil, 1966) to test whether it is a satisfactory fit (less than the critical value 0.1) statistically. The Theil's U index is 0.06 for the case of inter-arrival time, and is 0.05 for the case of negative inter-arrival time, indicating a satisfactory fit.[11] The results show that our assumption of Poisson arrivals is valid for this representative tour. We similarly examined each tour yielding different values of λ and p and validated the Poisson arrival assumption for all the tours.

In equation (2.1), the defining feature behind our SDE model is a mean-reverting property (x_t fluctuating around k_2). We further empirically verify whether the time-series data used in this study exhibit a mean-reverting characteristic by calculating the Hurst Exponent H (Hurst, 1951). When $0 < H < 0.5$, the series display a mean-reverting characteristic, and the strength of the mean reverting behavior increases as the Hurst Exponent H approaches zero. The estimated Hurst Exponent H value of our data is 0.36, implying that our time series data indeed follows a mean reverting process. Interestingly, such a mean-reverting phenomenon in

[11] As a contrast, the Theil's U index for the normal distribution is 0.18 for the case of inter-arrival time, and is 0.1336 for the case of negative inter-arrival time.

ratings is explained in the work of Moe and Schweidel (2012) where they find that reviewers tend to "beg to differ", i.e., say something different from previous reviewers. Reviewers tend to post negative (positive) reviews when earlier reviews are more positive (negative), hence fostering mean-reversion.

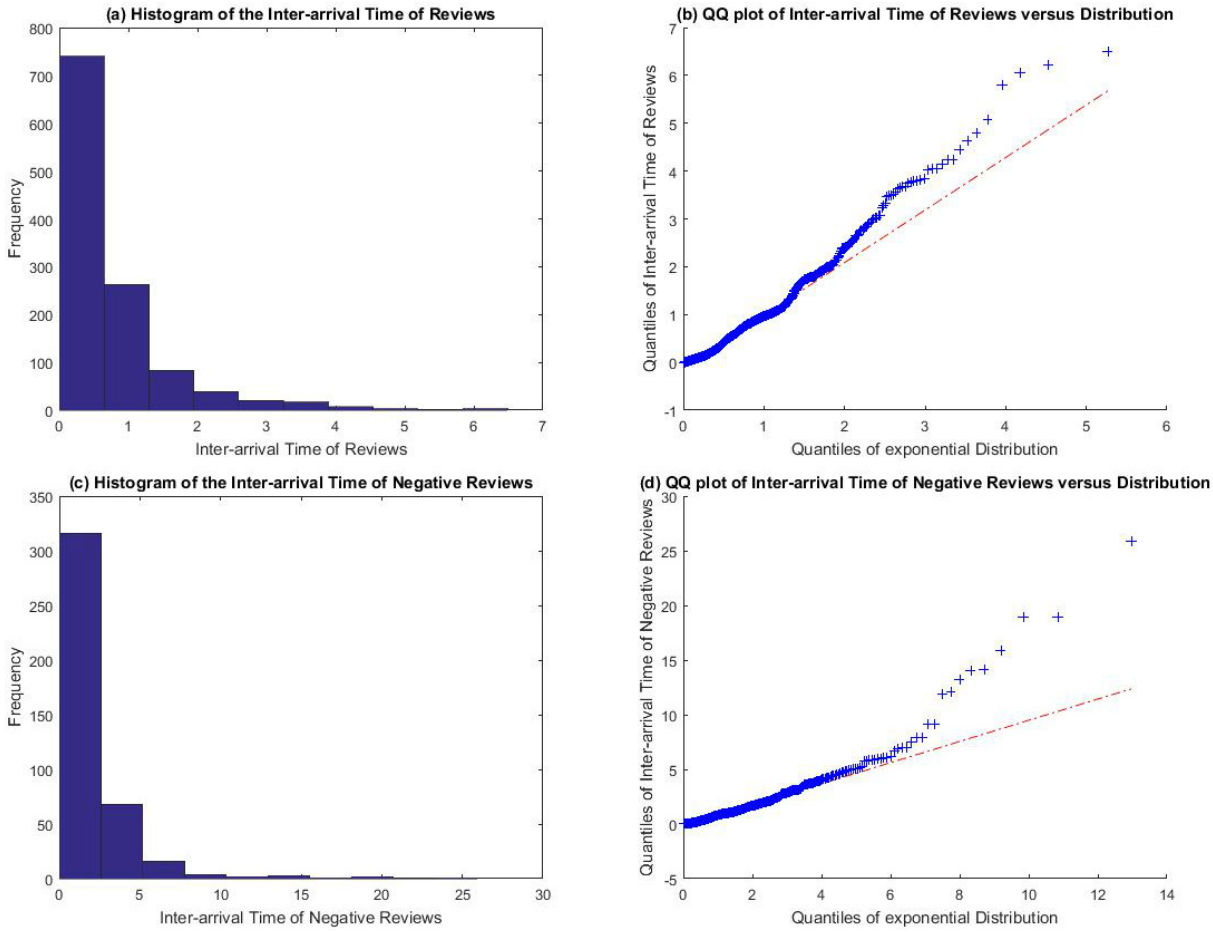

Figure 2.2. Inter-arrival Time of Reviews and Negative Reviews for Tour id 5106

We use the Nelder-Mead simplex algorithm to numerically obtain the MLE estimates. For quick convergence, good initial values of the variables are crucial. In Appendix A, we describe how good initial values for the control effort parameter, the boost parameter, the damage parameter, and the magnitude of the stochastic component are chosen. We implemented the SDE estimation procedure in MATLAB.

Table 2.6. Parameter Estimates for 10 random tours (out of 117 tours)

Tour ID	$\hat{\lambda}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\sigma}$	γ_1	γ_2
5106	1.595***(0.259)	0.298***(0.031)	0.032***(0.007)	0.019***(0.007)	0.009***(0.001)	0.092***(0.002)	1.55	0.17
72528	1.533***(0.333)	0.42***(0.052)	0.031***(0.008)	0.019** (0.009)	0.008***(0.001)	0.084***(0.003)	1.83	0.26
73154	2.22***(0.281)	0.358***(0.029)	0.059***(0.009)	0.020***(0.006)	0.011***(0.001)	0.115***(0.003)	2.06	0.18
80961	6.377***(0.817)	0.193***(0.024)	0.031 (0.027)	0.019***(0.005)	0.005***(0.000)	0.232***(0.003)	0.32	0.19
29336	1.941***(0.644)	0.266***(0.063)	0.013* (0.010)	0.010* (0.007)	0.003***(0.001)	0.112***(0.003)	0.92	0.19
23222	1.633***(0.269)	0.248***(0.029)	0.052***(0.011)	0.019** (0.009)	0.011***(0.001)	0.098***(0.003)	2.23	0.10
30938	1.54***(0.489)	0.279***(0.053)	0.014* (0.010)	0.020** (0.009)	0.006***(0.001)	0.091***(0.003)	0.63	0.24
71480	3.428***(0.471)	0.261***(0.026)	0.030** (0.013)	0.020***(0.005)	0.006***(0.001)	0.156***(0.003)	0.59	0.22
88292	3.711***(0.539)	0.216***(0.027)	0.030* (0.023)	0.017** (0.008)	0.006***(0.001)	0.198***(0.004)	0.60	0.17
1618693	19.72***(1.547)	0.163***(0.013)	0.248***(0.063)	0.187***(0.004)	0.045***(0.001)	0.410***(0.002)	0.08	0.18

Notes. The estimation is based on the in-sample data for each tour.
*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

As an illustration, Table 2.6 presents estimation results for only 10 tours (due to space limitation) with the standard error in parenthesis.[12] The estimation results for the rest 107 tours are in Appendix A. In Table 2.6, for each tour, we estimate the corresponding review arrival rate, the probability of negative review, the control effort parameter, the boost parameter, the damage parameter, and the magnitude of the stochastic component. Table 2.6 shows that there is considerable variation across all tours with respect to all these parameters, and most estimates are significant.

In understanding the estimation results across tours, comparing the estimated values of the parameter α and ρ provides us less insights, since the impacts of the control part and positive arriving reviews to dx_t depend on $\alpha(b - x_t)$ and $\lambda(1 - p)\rho(b - x_t)$ respectively. In order to compare the impacts from the two forces more intuitively, we introduce $\gamma_1 = \frac{\alpha}{\lambda(1-p)\rho}$, which measures the normalized boosting impact of the firm's control relative to the boosting impact of a positive review. In this sense, γ_1 is similar to the notion of odds ratio, where the marginal impact of firm response is measured against that of a positive review. $\gamma_1 > 1$ represents that firms' control exerts more influence in boosting the state than a positive review does on average. The eighth column in Table 2.6 presents the value of γ_1 for each tour. We find there is a large variance in γ_1 across the whole 117 tours (the average value of

[12]For each tour after the first response, we use the first 70% as in-sample and the remaining data as out-of-sample.

γ_1 is 1.73 with a high variance 24.06), indicating that the response strategies across tours are quite different. In Section 7, we will provide an operational interpretation of these response strategies and develop procedures to optimize them under different business objectives.

In a similar manner, we define $\gamma_2 = \frac{\rho(5-\mathbb{E}(x_t))}{\beta\mathbb{E}(x_t)}$, to measure the relative impact of a positive review against the impact of a negative review. The last column in Table 2.6 shows that the relative impact of positive reviews with respect to negative ones is about the same across the whole 117 tours (the average value of γ_2 is 0.38 with a small variance 0.12). Given $\gamma_2 < 1$ from our results, the negative review has bigger influence (on the state) than positive review, which is consistent with prior study (Chevalier and Mayzlin, 2006). We next validate our SDE model with extensive model comparisons.

2.6 Model Validation

The best way to validate a structural model is to predict the outcomes of quasi-experiments that the world presents to us, in which a policy change occurs and the data before and after the change is available. However, such opportunities are rare. Keane (2010) advocates alternatively validating a structural model (against a reduced form model) by resting primarily on how well the model performs in validation exercises, i.e, by examining whether the model does a reasonable job of fitting the historical data and whether the model does a reasonable job at out-of-sample prediction. In our setting, since we could not apply falsification tests to validate (or invalidate) our prescriptive model that requires conducting field experiment with the firm, we rely on comparing different models using the predictive performance. Specifically, we examine whether our proposed SDE model performs well on the basis of predictive performance as compared with conventional reduced form time-series models, including the representative ARMA, GARCH, Moving Average (MA), Exponential Smoothing (ES), and Naive Method (NM).

Autoregressive moving average (ARMA) is a classic method to model time series data. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The autoregressive part is a function of lagged dependent variable while the moving average component is a function of lagged error terms. The model normally takes the form of $ARMA(p, q)$ where p is the order of the autoregressive part and q is the order of the moving average part. After specifying p and q, ARMA models can be estimated by least square regressions. The generalized autoregressive conditional heteroscedasticity (GARCH) model is an extension of the Engle's autoregressive conditional heteroscedasticity (ARCH) model for variance heteroscedasticity. The $GARCH(p, q)$ model specifies p GARCH coefficients associated with lagged variances, and q ARCH coefficients associated with lagged squared innovations. A simple Moving Average (MA) uses the un-weighted mean of the previous m observations to forecast the next data point (Brown, 2004); while the simple Exponential Smoothing (ES) weights the past data in an exponentially decreasing manner, analogous to the discounting of cash flows over time (Brown, 2004). We choose m equal to 5 in our operationalization of MA and ES.[13] The Naive Method (NM) myopically uses the historical mean of the in-sample as prediction of the future state.

For each tour, we use the first 70% portion as in-sample and the remaining as out-of-sample. We use in-sample to calibrate all the models considered.[14] For both the ARMA and GARCH, the choices of p and q are empirically determined from the data on the basis of Bayesian Information Criterion (BIC) values for each tour. We then use the trained models to predict the out-of-sample. For SDE, y_t follows the data generating process specified in equation (2.4). Given an initial point of y_0, the conditional expectation at time t is $\hat{y}_t = \mathbb{E}(y_t|y_0) = (y_0 - k_3)e^{-k_1 t} + k_3$ and $\hat{x}_t = 5 - \hat{y}_t$. The out-of-sample predictive performance is measured using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE),

[13]We conduct robustness checks by choosing m equal to 3 and 10, and the results are qualitatively the same.

[14]The coefficient estimates for the ARMA and GARCH models are presented in Appendix A

and Symmetric Mean Absolute Percentage Error (SMAPE). RMSE is calculated following $RMSE = \sqrt{\sum_{t=1}^{n} \frac{(\hat{x_t}-x_t)^2}{n}}$, MAE is computed as $MAE = \sum_{t=1}^{n} \frac{|\hat{x_t}-x_t|}{n}$, and SMAPE takes the form of $SMAPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|\hat{x_t}-x_t|}{|\hat{x_t}|+|x_t|}$.

2.6.1 In-Sample Performance

We first validate our proposed data generating process of review ratings by comparing the estimated steady state mean (as t goes to infinity) of the state variable x_t with the observed steady state mean. The observed mean μ is derived from the data directly according to $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$. The estimated steady state mean of x_t using SDE is calculated as $\hat{\mu}_{SDE} = 5 - \hat{k}_3$. The values of $\hat{\mu}_{ARMA}$, $\hat{\mu}_{GARCH}$, $\hat{\mu}_{MA}$, and $\hat{\mu}_{ES}$ represent the estimated steady state means of review ratings using ARMA, GARCH, MA, and ES respectively.[15]

We compare the steady state mean of the SDE approach with the benchmark methods based on the RMSE criterion. The RMSE values of all the methods for each tour are calculated. The mean RMSE values (in increasing order) are 0.011 (SDE), 0.033 (ARMA), 0.061 (GARCH), 0.110 (MA), and 0.111 (ES). Note that SDE has the lowest mean RMSE value. We conduct a paired t-test between SDE and the other benchmark methods to examine whether the two means are statistically different. These results are reported in Table 2.7. For robustness, we also conduct the non-parametric Wilcoxon rank sum test and find that the results of the paired t-test and the Wilcoxon rank sum test are consistent. The statistical tests confirm that the differences in the mean RMSE values for SDE and the benchmark methods are statistically significant. To summarize, the SDE method outperforms the benchmark methods in terms of its ability to predict the steady state mean.

Besides the steady state mean, we also present RMSE, MAE, and SMAPE for all the methods to gauge the in-sample inference performance in Table 2.8 (with standard deviation

[15]The detailed information about the estimated and observed means for each tour as well as detailed in-sample and out-of-sample RMSE, MAE, and SMAPE are reported in Appendix A.

Table 2.7. Paired T-Test Results: Mean RMSE of Steady State Mean

Comparison	t-statistic (p-value)
SDE versus ARMA	-5.3145*** (0.000)
SDE versus GARCH	-7.6493*** (0.000)
SDE versus MA	-10.7183*** (0.000)
SDE versus ES	-10.7762*** (0.000)

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

in parenthesis) across the 117 tours. Table 2.9 presents the t-statistics and corresponding p-values for the statistical comparisons pertaining to in-sample metrics. As Table 2.9 shows, the SDE method outperforms GARCH, MA, and ES, but is comparable with ARMA. Note that the NM performance is trivial in the sense that it is true by construction: it uses the historical mean to predict the historical mean.

Table 2.8. Comparative Performance Results: In-Sample and Out-of-Sample

	Metric	SDE	ARMA	GARCH	MA	ES	NM
In-Sample	RMSE	0.1457 (0.0444)	0.1431 (0.0392)	0.1952 (0.0948)	0.2107 (0.0858)	0.2102 (0.0859)	N/A
	MAE	0.1182 (0.0353)	0.1163 (0.0316)	0.1414 (0.0565)	0.175 (0.0791)	0.1745 (0.0791)	N/A
	SMAPE	0.1569 (0.0572)	0.1561 (0.0574)	0.1826 (0.0758)	0.2372 (0.1095)	0.2367 (0.1103)	N/A
Out-of-Sample Long Term	RMSE	0.1878 (0.0868)	0.1937 (0.0861)	0.2505 (0.1383)	0.2267 (0.1024)	0.2267 (0.1029)	0.183 (0.0612)
	MAE	0.1537 (0.078)	0.1582 (0.0758)	0.1914 (0.0962)	0.1922 (0.0942)	0.1921 (0.0947)	0.1488 (0.0507)
	SMAPE	0.1793 (0.0811)	0.1886 (0.0975)	0.2223 (0.1048)	0.2234 (0.1136)	0.2231 (0.1137)	0.1761 (0.0625)
Out-of-Sample Short Term	RMSE	0.1353 (0.0618)	0.1547 (0.076)	0.6251 (0.7716)	0.1648 (0.0768)	0.1634 (0.076)	0.1642 (0.0785)
	MAE	0.1129 (0.054)	0.133 (0.0701)	0.4722 (0.5508)	0.1433 (0.069)	0.1415 (0.068)	0.1448 (0.0731)
	SMAPE	0.1377 (0.0722)	0.1615 (0.0849)	0.3103 (0.2243)	0.1734 (0.0926)	0.1715 (0.0916)	0.1755 (0.0938)

2.6.2 Out-of-Sample Performance

We then compare the predictive performance of the SDE model with ARMA, GARCH, MA, ES, and NM models on out-of-sample data. Here we conduct two types of predictive analyses: long term and short term. Long-term prediction refers to the ability to predict the value of x_t at the end of the out-of-sample period, whereas short-term prediction refers to the ability to predict the value of x_t at the end of the next 20 observations.[16] Table

[16] To make the measurement more robust, we consider multiple (randomly picked) starting states and find an average measure of predictive performance across these starting states.

Table 2.9. Paired T-Test Results: In-Sample and Out-of-Sample

	Comparison	In-Sample	Out-of-Sample Long Term	Out-of-Sample Short Term
RMSE	SDE versus ARMA	0.4752 (0.3175)	-0.5241 (0.3003)	-2.1332**(0.017)
	SDE versus GARCH	-5.1091***(0.000)	-4.155***(0.000)	-6.8437***(0.000)
	SDE versus MA	-7.2795***(0.000)	-3.1381***(0.001)	-3.2357***(0.0007)
	SDE versus ES	-7.2183***(0.000)	-3.1262***(0.001)	-3.0932***(0.0011)
	SDE versus NM	N/A	0.4798 (0.3159)	-3.1252***(0.001)
MAE	SDE versus ARMA	0.4434 (0.3289)	-0.4517 (0.3259)	-2.4514***(0.0075)
	SDE versus GARCH	-3.7716***(0.0001)	-3.2924***(0.0006)	-7.0213***(0.000)
	SDE versus MA	-7.093***(0.000)	-3.411***(0.0004)	-3.742***(0.0001)
	SDE versus ES	-7.0335***(0.000)	-3.3833***(0.0004)	-3.5603***(0.0002)
	SDE versus NM	N/A	0.5635 (0.2868)	-3.7892***(0.0001)
SMAPE	SDE versus ARMA	0.1153 (0.4541)	-0.7889 (0.2155)	-2.3143***(0.0108)
	SDE versus GARCH	-2.9251***(0.0019)	-3.5086***(0.0003)	-7.9258***(0.000)
	SDE versus MA	-7.0269***(0.000)	-3.4182***(0.0004)	-3.2901***(0.0006)
	SDE versus ES	-6.9412***(0.000)	-3.391***(0.0004)	-3.1329***(0.001)
	SDE versus NM	N/A	0.3432 (0.3659)	-3.4512***(0.0003)

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

2.8 presents the average of the out-of-sample metrics including RMSE, MAE, and SMAPE (in terms of long-term and short-term) across the 117 tours for all the methods. Table 2.9 shows the t-statistics and corresponding p-values pertaining to out-of-sample metrics.[17] Among these methods, GARCH performs the worst. We applied the Engle test for residual heteroscedasticity (Engle, 1982) and the results indicate that the autoregressive conditional heteroscedasticity effect is not significant, meaning the GARCH model does not fit our time series data well. The SDE method achieves comparable performance with ARMA and NM for the long-term out-of-sample prediction, but outperforms ARMA and NM in terms of the short-term prediction. As a whole, the SDE method performs better than GARCH, MA, and ES for both long-term and short-term out-of-sample prediction.

As summarized in Table 2.9, SDE is superior (or at least comparable) to all the benchmark methods, no matter what performance measure is used (RMSE, MAE, and SMAPE).

[17]We also conduct the (non-parametric) Wilcoxon rank sum test that does not require any distribution assumptions, and the results are consistent.

The few cases where SDE does not show superiority in long-term predictive performance is when compared against ARMA and NM. Both these methods typically perform well when predicting the long-term out-of-sample.

These results do not undermine the superiority of the SDE model. First, the short-term predictive performances of NM and ARMA are rather poor. From a control perspective, knowing the future in the short term is more important, because a control is a short-term intervention device, rather than something like a strategic plan that has medium to long-term implications. Further, only the SDE model is capable of prescribing a response strategy that best achieves the control objectives of the firm.

The key advantage of the SDE approach lies in its ability to perform counter-factual analyses that managers can use to anticipate the impact (on ratings) of a change in the control parameter (response strategy). That is, SDE is not just a predictive tool, but is also a prescriptive one. On the other hand, all the benchmark methods are all solely predictive in nature with no prescriptive capabilities. We will further demonstrate the prescriptive value of SDE in section 7.2.

2.7 Probabilistic Response Strategy and Applications

To demonstrate the prescriptive use of our model, this section maps the control parameter (α) to a probabilistic response strategy (policy) under different business objectives, thus providing an operational interpretation for the control. We further provide some practical applications of how the probabilistic response strategy could be implemented to achieve a certain managerial goal. This is only made possible through our structural SDE model (e.g. as opposed to ARMA, GARCH, etc.).

2.7.1 Mapping to a Probabilistic Control

A theoretical control parameter is difficult to interpret in practice, and raises the natural question: what does it mean for a firm to use a particular value under the control? One possible interpretation of our control parameter α is in terms of staffing (or sales force management), where the maximum value of the control corresponds to the maximum number of personnel that can be allocated to the job of responding to reviews. A second, more direct interpretation is that the firm could choose to respond to a certain percentage of reviews. It would be reasonable to expect this percentage to increase as x_t decreases, implying that the firm would respond more actively when the current perception of quality is low.

Based on the above rationale, we propose a probabilistic response strategy where the firm responds to a review with a probability of $p(x_t)$. In other words, the probability of providing response to the arriving review (r_t) at time t depends on the current state x_t. Using equation (2.5), the expected value of the transformed perception of quality variable after the time increment Δt is

$$\mathbb{E}(y_{t+\Delta t}|y_t) = y_t e^{-k_1 \Delta t} + k_3(1 - e^{-k_1 \Delta t}),$$

where, y_t is the initial value of the variable at time t.

Because $y_t = b - x_t$, we get,

$$\mathbb{E}(x_{t+\Delta t}|x_t) = b - (b - x_t)e^{-k_1 \Delta t} - k_3(1 - e^{-k_1 \Delta t})$$

where, $k_1 k_2 = \rho\lambda(1-p)b + \alpha b$, $k_1 = \alpha + \beta\lambda p + \rho\lambda(1-p)$, and $k_3 = b - k_2$. The coefficient k_1 measures the speed of reversion; k_2 and k_3 are the steady state mean of x_t and y_t respectively. Theoretically, the expected change in the state variable over time span Δt given the initial value x_t at time t is

$$\Delta x = \mathbb{E}(x_{t+\Delta t}|x_t) - x_t = (b - x_t - k_3)(1 - e^{-k_1 \Delta t}).$$

Next, using the data for each tour, we calculate the average increment in the perception of quality following a response. To calculate this average, suppose we have M responses given the state x_t in a tour. For each response i, we calculate the increment of the state variable over Δt, denoted as δ_i. Then the average increment given the state x_t (denoted as δ) is

$$\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i.$$

Over the time span Δt, there are an expected $\lambda \Delta t$ number of arriving reviews, of which $p(x_t)\lambda\Delta t$ would, on average, receive a response. The empirical average increment over the time span Δt equals to $p(x_t)\Delta t\lambda\delta$. Here, $\lambda\Delta t$ is the probability of the arrival of a review in time Δt. If we multiply this probability by $p(x_t)$, we get the probability of an increase or decrease during Δt. Note that δ represents the empirical average increase or decrease in ratings following a response, after considering all effects. Equating the empirical average increment to the theoretically expected increment and setting Δt close to zero, we have:

$$p(x_t)\lambda\delta = k_1(k_2 - x_t) \tag{2.10}$$

The left-hand side of equation (2.10) reflects the (empirically found) net increase or decrease in the ratings per unit time. The right-hand side reflects the theoretical increase or decrease per unit time. Equation (2.10) directly solves for $p(x_t)$. Using this relationship, we can represent the control parameter α in terms of a probabilistic response strategy. The probability of providing a response is clearly a decreasing function of x_t. Also as the level of the control (α) increases, the probability of providing a response increases. The above relationship for $p(x_t)$ shows that the probabilistic response strategy depends on the parameters for the tour: $\delta, \alpha, \lambda, \rho, \beta$, and p. Of these parameters, the value of α can be chosen by the decision maker to achieve a pre-specified goal. Having chosen an appropriate value of α (to be discussed next), the decision maker can implement this control using the probabilistic response strategy that corresponds to this choice of α.

The interpretation of the process can be visualized from the data. For each tour, we split the data into in-sample and out-of-sample. We use in-sample to estimate $\hat{\alpha}$, δ as well as other parameters, and further calculate $\hat{p}(x_t)$ according to equation (2.10) on the out-of-sample. For the same tour, we obtain the observed $p(x_t)$ from out-of-sample data. This is done by using the fraction of observations where the perception of quality is x_t and a review resulted in a response. Then we compare the observed $p(x_t)$ with our fitted $\hat{p}(x_t)$. We use tour 5106 to illustrate. In Figure 2.3(a), the solid line represents the observed probability of response and the dotted line delineates $\hat{p}(x_t)$ derived from equation (2.10). Figure 2.3(b) depicts the corresponding Q-Q plots of the observed response probability and the fitted probability. The points all are located fairly close to the reference line in general, indicating that the predicted response strategy fits well with the observed response strategy.[18]

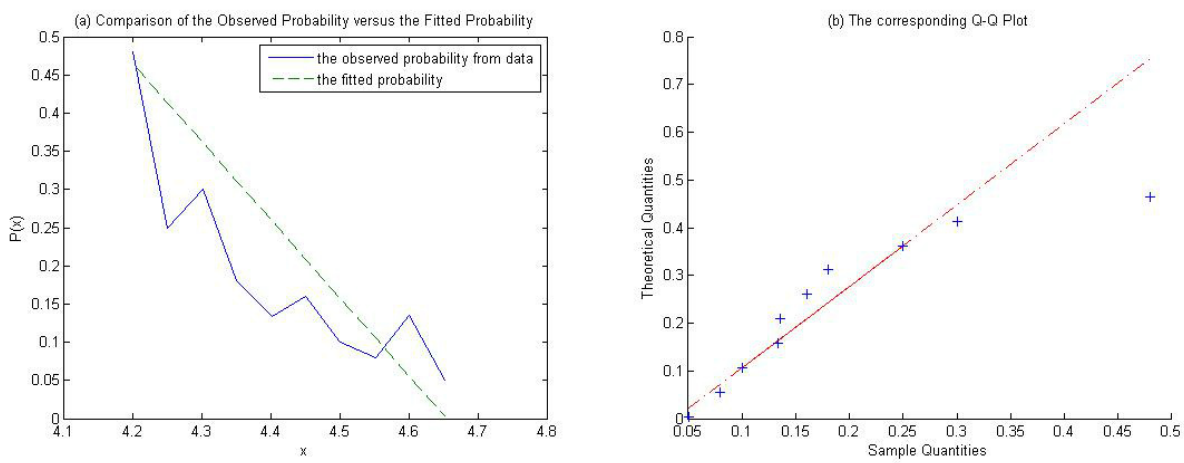

Figure 2.3. Comparison of the Observed $p(x_t)$ versus the Fitted $\hat{p}(x_t)$

[18]We conducted more formal analysis to test the fit. The correlation between the observed probability and fitted probability is 0.859. Then we run a regression with observed probabilities as the dependent variable and fitted probability as the independent variable. This results in a coefficient of 0.71 (significant at the 0.01 level). The R square is 0.74 with a F statistic 22.46 ($p = 0.0015$). All these statistics show a reasonably good fit, although the observed probability does not visually appear to match the fitted probability very well.

2.7.2 Policy Recommendations and Applications

A key advantage of our proposed SDE approach lies in its ability to not only predict consumers' perception of quality, but also influence it by responding to customer opinions in a prescriptive sense. The firm could have different objectives concerning the manner in which it desires to influence the perception of quality. Our SDE approach enables the firm to determine how a probabilistic response strategy (policy) shall be chosen to achieve such a goal. Clearly, a purely predictive approach (especially, one that does not model the underlying data generating process) will not be of much help when it comes to intervene the evolution of consumers' perception of quality over time. We discuss below how our approach can be applied to three different control objectives with corresponding policy recommendations.

Mean Control

A natural goal that firms might want to achieve is a target value of the average consumers' perception of quality, μ_{goal}. We have,

$$\mu_{goal} = \frac{5}{1 + \frac{\beta\lambda p}{\alpha + \rho\lambda(1-p)}},$$

hence, $\alpha = \frac{\beta\lambda p \mu_{goal}}{5-\mu_{goal}} - \rho\lambda(1-p)$. Using equation (2.10), this choice of α can then be mapped to a corresponding probabilistic response strategy.

Mean-Variance Control

A firm may want to influence not only the mean perception of quality but also its variance since large fluctuations in consumers' perception of quality would likely be viewed by customers as a sign of an unreliable product or service. For example, one kind of mean-variance control would be to achieve a specified value of the coefficient of variation of the state (c_g, say). By solving

$$\frac{\sqrt{\mathbb{V}(x_t)}}{\mu_x} = \frac{\sqrt{\frac{k_3\sigma^2}{2k_1}}}{5-k_3} = \frac{\sqrt{\left(\frac{5\beta\lambda p\sigma^2}{2(\alpha+\beta\lambda p+\rho\lambda(1-p))^2}\right)}}{\frac{5(\alpha+\rho\lambda(1-p))}{\alpha+\beta\lambda p+\rho\lambda(1-p)}} = c_g,$$

we can determine the value of the control (α) and then map this value to a corresponding probabilistic response strategy $p(x_t)$.

On similar lines, one may wish to set a lower limit (say, m_l) for the mean-variance expression $\mu - \gamma\sqrt{\mathbb{V}(x_t)}$, where γ is a penalty for variation. Note that the mean and variance of x_t are both functions of α. Hence, one can determine the smallest value of the control parameter α to achieve m_l, and then map this value of α to a corresponding probabilistic response strategy.

Service-Level Control

Rather than a mean-variance objective, firms may want to ensure that a certain percentage of the perception of quality is greater than a desired level d. This is especially important in service management since a high probability of poor performance (state falling below a certain level), despite a reasonable mean, indicates poor service level. In the context of control applications, this phenomenon is referred to as "out of control" (Merchant, 1982). Put differently, the objective is to provide a probabilistic guarantee that the state will not fall below a specified level over a given planning horizon. We call this *Service-Level Control* because it is similar to provide a service-level guarantee if the ratings system (and the responsibility to respond to negative reviews) was offered as a service by a vendor.

The probability density function of $2cy_i$ is a non-central chi-squared distribution with $2q+2$ degrees of freedom and non-centrality parameter $2u$. We have

$$P\{x_i \geq d\} = P\{y_i \leq 5-d\} = P\{2cy_i \leq 2c(5-d)\} = p_s,$$

and further

$$F_y(2c(5-d); 2q+2, 2u) = g(\alpha) = p_s.$$

Then the corresponding control parameter α is calculated through $\alpha = g^{-1}(p_s)$.

The above three examples illustrate how firms' potential objectives can be achieved by utilizing our SDE approach to manage an on-going process proactively, rather than merely predict its behavior reactively. Of course, firms could set different objectives and prescribe the optimal response strategy to achieve their specific objective in practice.

A Numerical Illustration

We again illustrate the above policies with tour 5106. The estimated parameters for this tour are $\hat{\lambda} = 1.595$, $\hat{p} = 0.298$, $\hat{\rho} = 0.019$, $\hat{\beta} = 0.009$, $\hat{\sigma} = 0.092$, $\hat{\delta} = 0.0329$, and $\hat{\alpha} = 0.032$ and the mean of the review ratings is $\hat{\mu} = 4.63$. We calculate its coefficient of variation which is equal to 0.036, and the probability of the perception of quality greater than 4.7 is equal to 0.05. In Mean-Control, the firm sets the target mean review rating to be 4.7, and the corresponding control parameter is $\hat{\alpha}_M = 0.046$. In Mean-Variance control, the target coefficient of variation of the review ratings is set to be 0.05, and the corresponding control parameter is $\hat{\alpha}_V = 0.017$. In Service-Level control, the firm would like to ensure that the probability that the perception of quality is greater than 4.7 is 0.8. Here, the corresponding control parameter is $\hat{\alpha}_S = 0.054$. Using equation (2.10), we map these values of the control to the probabilistic response strategy. When the state variable is greater than k_2 (the target mean), the probability of providing a response is zero.

Table 2.10 tabulates the recommended probability of providing a response given the state x_t under different control objectives discussed above. Also shown (in the second column) is the value of the control (α) that achieves the objective (mean, mean-variance, and service-level). As the control (α) magnitude increases, the probability of providing response becomes higher for the same level of the state variable.

Table 2.10. Probabilistic Response Strategy for Different Control Objectives

Objective	α	x_t							
		4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7
Mean	0.046	0.96	0.82	0.68	0.55	0.41	0.27	0.14	0
Mean-Variance	0.017	0.40	0.32	0.24	0.16	0.08	0	0	0
Service-Level	0.054	1	0.96	0.81	0.65	0.50	0.35	0.20	0.05

2.8 Robustness Checks

We conduct several robustness checks in this section.

2.8.1 Robustness to Potential Review Manipulation

One may argue that as an alternative to responding to negative reviews, a firm may opt to engage in review manipulation, e.g. by posting fake reviews to self-promote its products or services (Mayzlin et al., 2014). If this occurs, it could pose an identification challenge because the increase in ratings could be confounded under the influence of two forces: management response and review manipulation.

We take two measures to address this concern.[19] First we conduct a falsification test. The occasion when the firm responds to consumer complaints would indicate a moment in time where the management thinks it needs to take some form of corrective action. If indeed the firm engages in manipulating reviews, this is the time we would expect to see more manipulated reviews in the form of fake positive reviews. It is reasonable to assume that such firm manufactured reviews will be different from customer generated ones. If we are able to verify that the positive reviews following a management response do not systematically differ from the rest of positive reviews, we falsify the significant presence of review manipulation (because otherwise we should see a significant difference between these two groups of positive reviews). This can be done through text analytics.

[19]We thank the anonymous associate editor and reviewers for this suggestion.

Specifically, we constructed a corpus consisting of all the positive reviews and applied Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to extract topics from the corpus. We set the number of topics (T) to be 10. We then split all the positive reviews into two groups: those right before a response (Group I) and those right after a response (Group II). Group I consists of the positive reviews that are within the 10 preceding reviews that are not subject to another response; Group II consists of the positive reviews among the 10 reviews after a response, that are not subject to the impact of another response. Contrasting these two groups enables us to conduct an event-study type of analysis where the responses can be regarded as events. We next examine the topic composition (i.e. what topic keywords appeared in a review) in Group I and Group II separately, and then calculate the cosine similarity between the two vectors of keywords. The cosine similarity between positive reviews before and after response is 0.851, indicating high similarity between them. We also varied the number of topics to be 20, and the cosine similarity between positive reviews before and after response is 0.824, still suggesting high similarity. This analysis alleviates the concern that the Ctrip data is contaminated by review manipulation.

Alternatively, we calculate another cosine similarity score based on the term frequency-inverse document frequency (TF-IDF) values (Manning et al., 2008), with window sizes of 10 and 20 in Table 2.11. The t tests show that the mean similarity scores from those two groups are not different. The main take-away of this analysis is that the content of positive reviews do not get significantly affected by a response.

Table 2.11. Average Similarity Scores of Positive Reviews

Window Size	Without Response	With Response	p value
20	0.719 (0.106)	0.720 (0.112)	0.947
10	0.550 (0.142)	0.548 (0.120)	0.852

Notes. The standard deviation is reported in parenthesis.

Second, to further address the potential identification concern, we incorporate into our analysis a second dataset from Expedia, where review manipulation is less of a concern.

Expedia's review platform requires authentication of reviewers and only those customers who have stayed in a hotel are invited to post a review for the hotel. As such, Expedia's hotel reviews are widely perceived to be free of manipulation (e.g. Mayzlin et al. (2014)). We retrieved all the reviews in Expedia for the hotels in the Dallas metropolitan area, provided that these hotels have at least 20 reviews and at least 1 management response to eliminate outliers. This process left us with 136 hotels.

The Expedia dataset consists of 58,235 reviews from the period of January 2010 to May 2016. Similar to the Ctrip dataset, we retrieved the star rating, review text, review time stamp and management response if any. Table 2.12 reports the summary statistics for the Expedia data. The grand mean of reviews across all the hotels is 4.09.

Table 2.12. Summary Statistics for the Expedia Data

Review Rating	No. of Reviews	Percentage	No. of Responses	Fraction with Response
1	2,641	4.54%	423	16.02%
2	3,525	6.05%	465	13.19%
3	7,093	12.18%	846	11.93%
4	17,856	30.66%	1,501	8.41%
5	27,120	46.57%	2,405	8.87%
Total	58,235	100.00%	5,640	9.68%

We use the most recent 20 review ratings to construct the state variable x_t, and estimate the SDE model by hotel.[20] The estimation results show that most of the estimates for the control parameter α are significant, suggesting that the SDE framework is effective in a setting that is free of manipulation as well. Out of 136 hotels, we observe that 38 of them have negative estimated values of the control parameter. This is intriguing as it suggests that the firm's responses can have an adverse effect on future review ratings.

Our findings concerning the negative impact of responding to reviews are similar to those of Wang and Chaudhry (2018) who find that responding too often to a positive customer

[20]The detailed parameter estimation results for each of the 136 hotels are available upon request.

review could backfire. We further drilled down hotel characteristics to examine what types of hotels tend to experience a negative impact of management response. We extracted the following features for a hotel: hotel star rating, hotel affiliation (dummy variable with 1 for chain affiliated and 0 for independent hotels), mean review rating, mean response time, the positive response fraction (the number of responses to positive reviews divided by the number of responses to all reviews), and response ratio (the number of responses divided by the number of reviews) for each hotel. We then ran a simple logistic regression with the sign of α as the dependent variable (label $\alpha > 0$ as 1, and $\alpha \leq 0$ as 0) and hotel characteristics as independent variables. From Table 2.13, while a higher response ratio in general has a positive impact, a higher fraction of responses to positive reviews is harmful. This is intuitive: holding the number of responses constant, if we direct our efforts more toward positive reviews (rather than negative ones, where a response is perhaps, more appropriate), it will likely lower ratings.

Table 2.13. The Logistic Regression Result

Sign of α	Coef.	Std. Err.	z value	p value
Response Ratio	1.716	0.873	1.97	0.049
Hotel Star Rating	0.538	0.323	1.66	0.096
Hotel Affiliation	0.087	0.634	0.14	0.890
Mean Review Rating	-0.329	0.295	-1.12	0.265
Mean Response Time	0.005	0.005	1.07	0.285
Positive Response Fraction	-1.419	0.659	-2.15	0.031

We then calculate the measure γ_1 to gauge the impact of a response relative to that of a positive review. We categorize the estimated values of γ_1 by hotel features as shown in Table 2.14. We observe that the ratio of γ_1 is larger for non-chain, 2-star or lower hotels with average rating less than or equal to 4. This reveals that responses tend to be more effective for low-end, independent hotels that have relatively low review ratings.

We also replicated the predictive performance analysis on the Expedia data. The results, as measured with RMSE, MAE, and SMAPE, are shown in Table 2.15 and 2.16. The results

Table 2.14. Relative Impact of Management Response (Expedia Data)

Category	Group	γ_1
Hotel Affiliation	Non-chain (Independent)	10.9
	Chain	4.87
Hotel Star Rating	2-star or lower	10.31
	3-star or higher	2.53
Mean Rating	Less than or equal to 4	9.70
	Greater than 4	3.44

are qualitatively the same with those of the Ctrip data. As a whole, the SDE method is comparable with ARMA and NM (long-term), but outperforms GARCH, MA, ES, and NM (short-term).

Table 2.15. Comparative Performance Results (Expedia Data): In-Sample and Out-of-Sample

	Metric	SDE	ARMA	GARCH	MA	ES	NM
In-Sample	RMSE	0.242 (0.0941)	0.244 (0.0985)	0.2645 (0.1033)	0.3435 (0.155)	0.3434 (0.1553)	N/A
	MAE	0.1961 (0.0801)	0.1974 (0.0831)	0.2024 (0.0842)	0.2849 (0.1302)	0.2849 (0.1301)	N/A
	SMAPE	0.1401 (0.0416)	0.1413 (0.0431)	0.1443 (0.047)	0.203 (0.0886)	0.2052 (0.0948)	N/A
Out-of-Sample Long Term	RMSE	0.2848 (0.1866)	0.2912 (0.1965)	0.4165 (0.3506)	0.2753 (0.0975)	0.2771 (0.0986)	0.3107 (0.2142)
	MAE	0.234 (0.1635)	0.2405 (0.1742)	0.3145 (0.2564)	0.2309 (0.0881)	0.2325 (0.0893)	0.2635 (0.1964)
	SMAPE	0.1438 (0.0569)	0.1472 (0.0588)	0.1755 (0.0755)	0.1606 (0.1031)	0.1614 (0.1033)	0.1547 (0.051)
Out-of-Sample Short Term	RMSE	0.1523 (0.0639)	0.1618 (0.087)	0.7016 (1.2117)	0.1788 (0.0922)	0.1756 (0.0881)	0.2501 (0.2011)
	MAE	0.1263 (0.0538)	0.1363 (0.0791)	0.6117 (1.0361)	0.1546 (0.0829)	0.1509 (0.0792)	0.2319 (0.2014)
	SMAPE	0.0856 (0.0482)	0.1016 (0.0865)	0.2273 (0.2266)	0.1084 (0.0771)	0.107 (0.0768)	0.1247 (0.0553)

2.8.2 Additional Comparative Analyses

In our probabilistic response strategy, firms respond to a review with some probability depending on the state of recent review ratings. For higher values of the state, the probability of a response decreases. One might argue that this finding is non-surprising and the Probabilistic Strategy (PS) could easily be replaced by a simple Threshold Strategy (TS) that only responds to reviews where the rating is low (e.g., 1, 2, or 3), and never responds to moderate or high ratings (e.g., 4 or 5). This prompts us to probe deeper on what value our PS offers beyond the TS.

To make the analysis tractable, we begin with a simple case with a window size of 1 and analyze the steady state properties (e.g., mean, standard deviation, and steady state

Table 2.16. Paired t-Test Results (Expedia Data): In-Sample and Out-of-Sample

Metric	Comparison	In-Sample	Out-of-Sample Long Term	Out-of-Sample Short Term
RMSE	SDE versus ARMA	-0.162 (0.871)	-0.613 (0.54)	-1.018 (0.31)
	SDE versus GARCH	-3.914***(0.000)	-5.616***(0.000)	-6.769***(0.000)
	SDE versus MA	-6.812***(0.000)	0.684 (0.494)	-1.908*(0.057)
	SDE versus ES	-6.85***(0.000)	0.723 (0.47)	-1.623*(0.100)
	SDE versus NM	N/A	-1.07 (0.285)	-5.579***(0.000)
MAE	SDE versus ARMA	-0.081 (0.936)	-0.626 (0.532)	-1.269 (0.205)
	SDE versus GARCH	-2.263**(0.024)	-5.205***(0.000)	-6.984***(0.000)
	SDE versus MA	-7.129***(0.000)	0.981 (0.327)	-2.51**(0.013)
	SDE versus ES	-7.183***(0.000)	1.019 (0.309)	-2.148**(0.032)
	SDE versus NM	N/A	-1.433 (0.153)	-6.366***(0.000)
SMAPE	SDE versus ARMA	-0.359 (0.72)	-0.865 (0.388)	-1.501 (0.134)
	SDE versus GARCH	-2.024**(0.044)	-5.306***(0.000)	-8.961***(0.000)
	SDE versus MA	-6.627***(0.000)	-1.838*(0.067)	-1.952*(0.052)
	SDE versus ES	-6.661***(0.000)	-1.883*(0.061)	-1.645*(0.101)
	SDE versus NM	N/A	-1.407 (0.161)	-5.688***(0.000)

Note. *** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

response probability) under TS and PS. Before response, the review ratings follow the prior distribution F (p_i, $i = 1, 2, 3, 4, 5$), where p_i is the probability that the review rating equals to i. Once a review receives a response, it will affect the distribution of next arriving review ratings, denoted by the posterior distribution F' (p'_i, $i = 1, 2, 3, 4, 5$). The probability of responding at time t (q_t) is the weighted sum of two components: if there was a response at the previous time $(t-1)$, the review arriving at time t will follow the posterior distribution F', otherwise it will follow the prior distribution F. Thus we have,

$$q_t = q_{t-1}P^{'} + (1 - q_{t-1})P.$$

In the above, P (P') represents the probability of responding under prior (posterior) distribution. In steady state, $q_t = q_{t-1} = q$. After solving the equation, we get the steady state response probability,

$$q = \frac{P}{(1 + P - P')} \tag{2.11}$$

The steady state mean of ratings is calculated by:

$$\mathbb{E} = q\mu' + (1-q)\mu \tag{2.12}$$

where $\mu' = p'_1 + 2p'_2 + 3p'_3 + 4p'_4 + 5p'_5$ and $\mu = p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5$. The steady state variance of ratings is computed by:

$$\mathbb{V} = \sum_{r=1}^{5} qr^2 p'_r + \sum_{r=1}^{5} (1-q)r^2 p_r - (q\mu' + (1-q)\mu)^2 \tag{2.13}$$

Under PS, the firm adopts a trigger that smoothly decreases with the rating. That is, given the review rating r, we denote the probability of responding as $g(r)$, which is decreasing as rating magnitude increases. Here we consider a linear function form $g(r) = k * r + d$, where $k < 0$ and $d > 0$. To guarantee a meaningful probability, we set $g(r) = 0$ ($g(r) = 1$) when $g(r) < 0$ ($g(r) > 1$). Thus, we have

$$P'(PS) = \sum_{r=1}^{5} g(r)p'_r = g(1)p'_1 + g(2)p'_2 + g(3)p'_3 + g(4)p'_4 + g(5)p'_5,$$

and

$$P(PS) = \sum_{r=1}^{5} g(r)p_r = g(1)p_1 + g(2)p_2 + g(3)p_3 + g(4)p_4 + g(5)p_5.$$

TS follows a zero-one trigger (always respond when the rating is less than a threshold value and otherwise not respond). For example, the firm always responds when the arriving review rating is less than or equal to 2, and otherwise does not respond. Therefore, we have $P'(TS) = p'_1 + p'_2$, and $P(TS) = p_1 + p_2$.

Based on the above analysis, we conduct several comparisons. Under TS, the threshold can only take on values of 1, 2, 3, 4, or 5, since the ratings are discrete ranging from 1 to 5. However, PS is more flexible since the probability of responding $g(r)$ can accommodate any functional form. Next, we provide a numerical example to illustrate. We assume a prior distribution ($p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.2$, $p_4 = 0.4$, $p_5 = 0.1$) and a posterior distribution ($p'_1 = 0.02$, $p'_2 = 0.05$, $p'_3 = 0.08$, $p'_4 = 0.2$, $p'_5 = 0.65$). Let us assume that the firm wants

to achieve a target mean of 3.45. Under PS, the firm can use the response strategy given by $g(r_t) = -0.41 * r_t + 1.39$, corresponding (exactly) to the target mean of 3.45 (equation (2.12)). We can calculate the steady state variance of ratings to be 1.51 using equation (2.13) and the steady state response probability (q) to be 20.62% using equation (2.11). Under TS, the firm can only choose a threshold that achieves a steady state mean closest to the target mean. This threshold is 2; the value that achieves a steady state mean of 3.5 (closest to the target mean of 3.45). On the other hand, if a threshold value of 1 is chosen, the steady state mean is 3.31, further away from the target mean of 3.45. The threshold value of 2 yields a variance of 1.53 and a steady state response probability of 24.39%. In this example, we see that PS has lower variance and is more efficient (lower steady state response probability) than TS.

When the window size n is greater than 1 (e.g., 20), the analysis becomes more complicated, and thus simulation is used to investigate how the state evolves over time. The state variable x_t at time t is no longer equal to r_t, but will absorb the newly arriving review r_t at time t. We simulate $N = 10,000$ trajectories to smooth the randomness, and the time horizon is from $t = 1$ to $T = 2,000$ in each trajectory. The values of the prior and posterior distributions are the same as when the window size equals to 1. From $t = 1$ to $t = 20$, we draw r_t from the prior distribution to construct the initial state. From $t = 21$, the distribution of arriving review at time t depends on whether there was a response at previous time $(t - 1)$. If there was a response at time $(t - 1)$, we draw r_t from the posterior distribution and otherwise from the prior distribution.

Let us assume that the firm wants to achieve a target mean of 3.8. Under PS, the firm can use the response strategy given by $g(x_t) = -0.95 * x_t + 4.08$, corresponding (exactly) to the target mean of 3.8. Under PS, we observe the steady state variance of ratings to be 0.036 and the steady state response probability (q) to be 48.02%. Under TS, the firm can only choose a threshold that achieves a steady state mean closest to the target mean. This

threshold is 4; the value that achieves a steady state mean of 3.9 (closest to the target mean of 3.8). On the other hand, if a threshold value of 3 is chosen, the steady state mean is 3.65, further away from the target mean of 3.8. The threshold value of 4 yields a variance of 0.04 and a steady state response probability of 57.49%. Once again, we see that PS has lower variance and is more efficient (lower steady state response probability) than TS.

In practice, firms may often consider a simple response strategy (such as TS) to achieve their desired control objectives. However, as we have shown, a fundamental difference between PS and TS is that PS is a more flexible strategy and can precisely achieve the target set by the firm. This target could be the mean (mean control), or a more sophisticated target, such as the combination of the mean and variance (mean-variance control), or even a target that is stated in terms of the probability of the state variable (service-level control). In our simulation, we provided examples of cases where TS has a higher variance than PS (also a higher steady state response probability). This may not always hold for all parameter settings. The important take away regarding the issue of PS versus TS is that TS, because of its discrete nature, may not be able to precisely meet the control objectives of the firm.

2.9 Model Extension – Multi-Dimensional Strategy

We further extend our model by incorporating additional information (beyond review rating) that can guide the response strategy. This enriched model provides new practical insights that are more prescriptive.

Without loss of generality, we extend our one-state model to a composite, two-state model. To do so, we construct a composite variable that is a weighted combination of two states: the sentiment of a review together with the review rating. The sentiment in a review (S_t) is obtained by text mining the review text, which is elaborated in Appendix A. A composite state is constructed as, $Z_t = w_1 X_t + w_2 S_t$. The weights w_1 and w_2 are determined using logistic regression with the binary response event as the dependent variable. The

independent variables are the rating and the sentiment. We run the logistic regression for each tour and check whether ratings and sentiments are significant factors for evoking a response. If both coefficients are significant, we refer to the tour as a Z-strategy tour (treatment group). If only the ratings variable is significant, the tour is said to adopt a X-strategy (control group). The situation where only the sentiment is significant, can be handled in a similar manner; however, in our context such a situation rarely arose. Hence, we do not consider this situation.

Next, we investigate the impact of the response strategy (Z or X) on the review ratings and sentiments of tours. We first use PSM to match tours based on observed tour characteristics, which are the same with what we have defined in the quasi-experiment in Appendix A. We run a Probit model to match the treated and control tours based on their predicted propensity scores. The Probit model results are shown in Table 2.17. We use nearest neighbor (NN) matching with replacement; the PSM yields 54 tours in the treatment group and 63 tours in the control group. We focus on those pairs (matched tours) in which tour one adopts a Z-strategy (treatment) while the other adopts an X-strategy (control). We further check whether the covariates of the matched treated and control tours are balanced. Table 2.18 shows that after matching the tour characteristics are comparable. For each tour, we consider six outcomes of a response strategy: mean rating, standard deviation of rating, the proportion of very low ratings (equal to 1, the lowest possible rating), mean sentiment, the standard deviation of sentiment, and the proportion of very low sentiments (equal to 1, the lowest possible sentiment).[21]

Table 2.19 reports the average treatment effect for each outcome. We see that the X-strategy performs better than the Z-strategy based upon rating-related outcomes. Specifically, the X-strategy has a mean rating that is about 0.1 higher, or about 2% higher. This

[21] We normalize the sentiment score from 1 to 5, the same scale with ratings.

Table 2.17. Probit Regression of Response Strategies

Treatment	Coef.	Std. Err.	Z value	p value
Length	-0.105	0.145	-0.730	0.468
Flexibility	-0.487	0.570	-0.850	0.393
Departure	-0.481	0.487	-0.990	0.324
Destination	0.076	0.255	0.300	0.767
Agent	0.180	0.928	0.190	0.846
Transportation	1.048	0.644	1.630	0.104
Type	-0.184	0.298	-0.620	0.538
Hotel	-0.156	0.252	-0.620	0.535
Price	-0.077	0.294	-0.260	0.794
Age	2.207	0.890	2.480	0.013
Intercept	-70.043	38.188	-1.830	0.067
Number of obs	117			
LR chi2(11)	51.93			
Prob >chi2	0.000			
Log likelihood	-54.789			

Table 2.18. Balance Check Before and After Matching

	Mean Difference		p-value from t-test	
	Before Matching	After Matching	Before Matching	After Matching
Length	0.803	-1.093	0.046	0.171
Flexibility	-0.209	-0.167	0.012	0.202
Departure	0.061	-0.111	0.253	0.282
Destination	0.056	0.185	0.345	0.276
Agent	-0.074	-0.042	0.022	0.154
Transportation	0.188	-0.056	0.022	0.394
Type	-0.497	0.296	0.017	0.269
Hotel	0.690	0.083	0.016	0.452
Price	788.675	668.241	0.004	0.166
Age	0.138	-0.015	0.000	0.421

difference might seem small; however, industry reports on lodging indicate that small differences indeed matter to revenue. For example, according to industry reports on lodging performance, a 1% reputation improvement results in a 1.42% increase in revenue per available room (RevPAR) (Anderson, 2012a). On the other hand, the Z-strategy does better on sentiment related outcomes (almost the same mean sentiment, but lower standard deviation of the sentiment and smaller proportion of very low sentiments). Thus, neither strategy

dominates. Our model and analyses permit a firm to study various response strategies to choose the one that best suits management goals. Adding an additional variable in our SDE model has demonstrated the power and flexibility of the SDE approach. Furthermore, the specific two-dimensional example used in the demonstration has generated insights that are more prescriptive as it factors in the nature of the review text.

Table 2.19. Performance of X-strategy and Z-strategy

Variable	Z-Strategy	X-Strategy	Difference	Std. Err.	p value
Mean Rating	4.536	4.638	-0.102	0.044	0.012
Mean Std. Dev. (Rating)	0.697	0.649	0.049	0.031	0.060
Proportion of Ratings Equal to 1	1.082%	0.603%	0.005	0.003	0.043
Mean Sentiment	3.558	3.592	-0.034	0.024	0.085
Mean Std. Dev. (Sentiment)	0.451	0.491	-0.040	0.018	0.018
Proportion of Sentiments Equal to 1	0.011%	0.222%	-0.002	0.001	0.000
Response Ratio	0.068	0.069	-0.001	0.015	0.464

2.10 Conclusions

This chapter studies the problem of managing online customer opinions using management response strategy. Toward this end, we develop a stochastic differential equation model to study the evolution of user opinions over time. The model incorporates a *control* strategy into firm response to user reviews and investigates the impact of the responses on review ratings. The model is empirically estimated using data on firm responses and online customer reviews for two of the world largest travel agents. The model is validated along different dimensions of its performance. First, the predicted steady state mean obtained from the model is compared with the observed steady state mean, inferred directly from the data. We then demonstrate the superiority of our SDE model by examining its predictive performance compared with benchmark models including ARMA, GARCH, MA, ES, and NM. Our approach achieves superior or comparable predictive performance with those benchmark models. We further provide an operational interpretation of the control by mapping it to

an equivalent probabilistic response strategy. Then, we demonstrate the applicability of the probabilistic response policy under different control objectives, namely, mean control, mean-variance control, and service-level control.

Finally, we enrich our model by incorporating additional information (review sentiment) garnered from the review text. This enables the firm to embrace a multi-dimensional response strategy. It also provides practical insights that are more prescriptive as the firm can react to the change of environment along multiple dimensions.

The main contributions of the chapter are two-fold. Recent research shows that publicly responding to user comments can increase online reputation. However, while predictive models of review ratings exist, there are no prescriptive models that recommend the best response strategy to decision makers to achieve a specific managerial goal. We believe what we offer in this study is a significant first step towards a prescriptive response strategy. From a methodological perspective, the stochastic differential equation approach presented here, opens the black box of the data generating process underlying review data as opposed to reduced form models that essentially stop at estimation or prediction. Compared with other structural models, a key distinction of our approach is that we model the stochastic, time-series nature of the review data generating process. Most other structural approaches have emphasized a utility framework to understand the process of data generation.

From an operational (or practical) perspective, our study offers specific guidelines (via a probabilistic strategy) on managing user opinions through controlled responses. A full response strategy – one that responds to every review – will likely be too costly or ineffective, if the responses are not adequate. The cost of responding can be material, e.g. the personnel costs for the assigned staff to manually respond to reviews. It is also costly to train people to investigate and professionally respond to reviews and effectively manage customer relationships. However, even if the cost of responding is ignored, it may still be better not to respond to every review, because a full-response strategy may trigger opportunistic

behavior from some consumers. For example, it is common for firms to send out coupons to unsatisfied consumers. If coupons were sent out for every negative comment, this practice could encourage some opportunistic consumers to deliberately post negative reviews, even if they are in fact satisfied with the product or service. Thus, responding to every review may backfire and incur additional cost. Further as we demonstrated, our probabilistic response strategy is more flexible than the simple threshold strategy.

Using our model, managers can fine tune the response strategy to achieve a desired outcome such as mean control, mean-variance control and service-level control. Of these, the last two goals can only be achieved if the predictive model can recover the distribution of the review ratings as a function of the response strategy used by the firm.

This study has several limitations. The response strategy considered in this chapter is a first attempt at constructing a prescriptive model on this subject. In general, a response strategy could include the dimensions of what, when and how. The what aspect of a strategy answers the question: what review should the firm respond to? The when aspect of a strategy relates to the issue of delay, i.e., should the response be immediate or delayed. A delayed response is likely to be less effective but, at the same time, less costly to implement. Finally, the how aspect of a strategy is associated with how the actual response is constructed. Clearly, all dimensions of a response strategy are important. However, our study was limited to one dimension, namely, it provided insight into what review merited a response. We believe that the other dimensions of a response strategy (i.e., the when and the how) are fruitful directions for future work.

Another aspect of designing a response strategy is to allow the impact of responding to a positive review to be different from that of responding to a negative review. In our study, we did not distinguish between these impacts. However, the SDE framework lends itself well to models where different response strategies for positive and negative reviews can be jointly optimized. This is a promising direction for future work.

Our model of the impact of a negative review (see equation 2.1) is linear in nature. However, this impact may be non-linear in some situations. For example, one possible non-linear model is one where a negative review has less impact at low values and at high values of the state. That is, the maximum impact of a negative review occurs at some intermediate value of the state. In an inverted-U impact curve, firms would have little incentive to operate in the increasing portion of the curve. This is because the impact of a negative review at a point in the increasing portion of the curve is the same as the impact at a corresponding point in the decreasing portion of the curve where the rating is much higher. Further, if a firm could successfully drive its ratings to a very high value, it would have to do very little to maintain its ratings at this high value. Thus, in an inverted U-model of impact, one should expect to see many firms with ratings close to the maximum value of the range. In general, the manner in which the current state moderates the impact of a negative review, can profoundly affect the response strategy. A study of non-linear models of impact is therefore a useful direction for future work. Finally, competition is not modeled: the behavior or strategy of competing firms would likely to influence the focal firm's response strategy as well. We are currently exploring these possibilities.

CHAPTER 3

THE RACE FOR ONLINE REPUTATION: IMPLICATIONS FOR PLATFORMS, FIRMS AND CONSUMERS

3.1 Introduction

The Internet has empowered consumers to garner product or service information from the experience of other consumers through online user opinions (e.g. customer reviews). It is reported that 88% of consumers read online reviews to evaluate the quality of a local business (Anderson, 2014). Online reviews have been found to have a large impact on a firm's online reputation and revenue. For example, a 1% improvement in reputation can result in up to a 1.42% increase in revenue (Anderson, 2012b), and firms making into the list of top 25 reviews generate 5% more payments, compared to those which did not make it (Tadelis, 2016). In response, firms spend millions of dollars annually to manage user opinions to build and maintain brand awareness (Forbes, 2013).

In TripAdvisor's latest study, it is reported that 97% of business owners consider online reputation management to be important to their businesses (Erskine, 2018). Thus, online reputation has become a dominant marketing-mix variable for firms. For example, hotels have begun to redirect their advertising budget to appear in preferred AAA listings to manage their reputation and monitor guest comments etc. (Manley, 2017). As Doug Collins, president of DC Hospitality, said: "We teach our franchisees that you've got to monitor online comments, and if there is a bad comment you should address it and see if you can make lemonade out of lemons" (Manley, 2017).

A firm's online reputation (reflected in its ratings) fluctuates and is not fully controllable. It is this fact that makes it valuable and realistic (Erskine, 2018). Thus, while a firm's mean ratings can be influenced by its *control* efforts, there is usually a stochastic component that cannot be influenced by the firm. This control effort can be considered to be any form of

activity that a firm performs to improve its online reputation. These activities could include monitoring online customer reviews, investigating negative reviews, improving quality or customer service, writing a management response to customer reviews, etc.

A consumer's purchasing decision on a platform depends not only on the ratings of the focal firm but also on those of its competitors. As such, firms need to justify to customers why they should choose them over their competitors. To achieve this, firms often track social media key performance indicators (e.g. review ratings) about themselves as well as those of their competitors (Glassman, 2011). Our research setting considers a group of firms in a platform that compete for sales (such as Yelp or Expedia) by managing their reputation. These firms face the challenge of effectively managing online reputation in the presence of the ratings of competing firms (Kent, 2014). To this end, the literature has recently started examining the problem of reputation management through approaches such as firm response (Kumar et al., 2018; Yang et al., 2018) and review manipulation (Luca and Zervas, 2016). There is also a stream of research that examines the impact of online reviews of competing firms on a focal firm's sales (Jabr and Zheng, 2014; Kwark et al., 2018).

From the perspective of platform owners, managers often assume that the key to success is to grow the number of firms and consumers as quickly as possible, to reap the positive network effects, believing that size leads to competitive advantage (Cennamo and Santaló, 2013). However, a simple growth strategy, often referred to as "get big fast," could lead to a *platform trap* because it ignores competitive forces that arise among the firms as a result of the growth strategy (Cennamo and Santaló, 2015).

Given the importance of online reputation competition, there are several challenges firms and platforms continue to face:

- How much effort should a firm exert to effectively manage online reputation (that evolves stochastically over time) and how does the effort level of competing firms affect this choice?

- From the platform's perspective, how should it target new firms to join, without compromising the objectives of incumbent firms and consumers?

- For a given market size, how should the platform balance the different kinds of firms in the platform so as to maximize platform goals?

Most previous models on advertising competition (that can be considered similar to ratings competition) have considered a duopoly market to analyze equilibrium outcomes. A notable exception is (Naik et al., 2008) where competition among many firms is studied. However, unlike our study, Naik et al. (2008) consider a deterministic problem where the impact of the control (advertising) on sales does not have a stochastic component. While a duopoly analysis is often a good substitute for a market with many competing firms, in online review competition, not only can consumers observe the reviews received by all firms in the market, but they can also observe the responses that firms provide to customer reviews. Thus, in a competitive market, when all firms in the market attempt to manage their online reputations, the equilibrium ratings in the market are difficult to solve using a traditional stochastic differential game-theoretic model.

To address the above difficulty, we derive the market equilibrium using the following process. Because the rating of each individual firm is stochastic, the market rating (average rating across the firms) is also stochastic. However, in equilibrium, consumers and firms use the mean of the market rating as the basis for their actions. Each firm's sales (another stochastic process) are driven by its current ratings and this (mean) belief. This is analogous to the reflection problem identified in (Manski, 1993) where the group behavior affects individual behavior, while the group behavior is the aggregation of individual behavior.

An analysis of the mean market equilibrium reveals several insights. The mechanism underlying these insights is that a particular firm's actions not only directly affect its outcomes (such as ratings or profit), but also indirectly affect these outcomes via the equilibrium

mean market rating. We list below several results that impact the strategies of platforms and firms.

- For a given number of firms in the platform, a more heterogeneous market (one where the parameters of the firms are very different) leads to a lower mean market rating and a higher total profit of the firms in the market. This finding can benefit platform owners to choose the right mix of firms in the platform. A heterogeneous market can increase the profit of firms and the platform, but at the expense of customer experience.

- Our results can also inform platform owners to develop a targeted growth strategy, i.e., encourage certain kinds of firms to join the platform. Conversely, our results can also help an individual firm decide whether or not to join an existing platform.

  - Adding firms with average ratings (i.e., the middle of the market) is the safest option considering the goals of the platform (increasing total profit) and other stakeholders, namely, incumbent firms and consumers.

  - Growing the market at the bottom (i.e., adding firms with lower than average ratings) hurts consumers, but benefits both the platform and the incumbent firms.

  - Finally, growing at the top is good for the platform and consumers, but hurts the profit of incumbent firms.

- Our results also provide insights for firms to manage their online reputation in the presence of competition. A firm could benefit (hurt) from an increase in the cost of control effort (or a decrease in the sales margin). However, the above result could flip if another firm experiences a higher (lower) increase in the cost of control effort (or a greater or lower decrease in the sales margin). Thus, in a competitive market, adversity can be a friend.

The rest of the chapter has the following structure. Section 2 summarizes the relevant literature. Our model is presented in Section 3. In Section 4, we study the equilibrium. Section 5 describes the results and discuss the implications of this study. Section 6 concludes the study.

3.2 Literature Review

In this section, we discuss the related literature and how this chapter builds upon and extends various streams of related research.

3.2.1 Advertising Competition

Traditionally, in the advertising literature, many researchers have explored the optimal advertising strategy in monopoly, or competitive markets (Bass et al., 2005; Naik et al., 2008; Rubel et al., 2011). In this literature, firms invest in advertising over multiple periods to maximize their total profits, and the advertising expenditure (effort) affects both present and future demands of the product. The effect of advertising persists beyond the current period, but with diminishing returns (Liu et al., 2012), where the term *goodwill* is proposed to capture consumers' awareness of the product built through advertising. The competing brands seek to increase brand awareness to boost their sales. Hence, managers have to take into account the presence of multiple competitors in determining their best course of action. Each firm selects a level of advertising expenditure to maximize its discounted profit flow over time, where firms compete among themselves for goodwill. One well-known duopoly example is the strategic advertising competition between Coke and Pepsi, where Chintagunta and Vilcassim (1992) empirically validate the classic Lanchester model (Little, 1979) using the real data. When there are more than two rivals in the market, Naik et al. (2008) consider oligopolistic competition, in which each firm solves an optimal advertising strategy to maximize its profit in equilibrium. Instead of modeling deterministic sales-advertising

dynamics, Sethi (2004) adds a diffusion term capturing randomness in the evolution of sales in a duopoly setting.

Unlike previous studies, we consider multiple firms in a market that compete for sales by attempting to control their ratings. To the best of our knowledge, this combination of features, namely, more than two firms, stochastic state variables, and interaction: control effort → ratings → sales rate → profit, has not been studied earlier. We next discuss online ratings competition between firms.

3.2.2 Online Ratings Competition

Since online consumer reviews are observable to other consumers and competing firms in the market, consumers are able to form informed beliefs about the review ratings of competing firms in this market. These beliefs could act as an anchor when they make purchase decisions. Our chapter contributes to a growing body of literature related to online review ratings competition, such as those examining the impact of review ratings on product sales (Chevalier and Mayzlin, 2006; Duan et al., 2008; Zhu and Zhang, 2010). Within this literature, a recent stream of studies has examined the impact of the reviews received by competitors. It has been well documented that consumers make their purchase decisions based on the review ratings of the focal product and competitors'; there is a strong spillover effect from competitors' ratings to the focal product's sales (Jabr and Zheng, 2014; Kwark et al., 2018). For example, Jabr and Zheng (2014) empirically demonstrate that product sales of a focal firm drop with improvements in the reviews of a competing product. Using retail click-stream data, Kwark et al. (2018) show that the mean review rating of substitute products exerts a negative impact on the sales of the focal product.

Several analytical studies have attempted to unravel how competition emanating from online reviews influences a firm's decisions (Mayzlin, 2006; Li et al., 2011; Kwark et al., 2014). Mayzlin (2006) identifies a unique equilibrium where online reviews are persuasive despite

the promotional chat (e.g. review manipulation) activity of competitors. Contrary to the prior advertising literature, firms spend more resources promoting inferior products under such an equilibrium. Li et al. (2011) analyze the offsetting effects caused by competition for repeatedly purchased products, resulting in a "S-shaped" relationship between the quality of reviews and firm profits. Kwark et al. (2014) study the effect of online product reviews on upstream competition between manufactures in a channel structure. The quality and fit information provided by consumer reviews affect the upstream competition in different ways. There are some papers focusing on the strategic behaviors between firms and consumers with online feedback where agents can rate each other sequentially (Hui et al., 2014; Ye et al., 2014) or manipulating customer reviews in response to competition (Wu and Qiu, 2016), where they find that although forging customer reviews can improve the perceived quality, high-quality sellers do not do so due to higher marginal cost.

Different from prior studies, we investigate how competition impacts a firm's decision (control effort) to manage user opinions and future profit in a competitive market. Besides, we also study the impact on the total profit of the firms in the market, a measure that is in line with the goals of the platform.

3.2.3 Equilibrium Models with Many Players

Most previous analytical studies on online ratings competition have considered a few competing firms (often two firms in a duopoly setting). In this study, we consider a competitive market with a large number of firms that attempt to manage their online reputations. When the number of players is large, stochastic games become notoriously intractable. Lasry and Lions (2007) introduce mean field games to study Nash equilibria when the number of players is large, and the players interact symmetrically through the empirical distribution of the states of all players. Given such a distribution, each player typically solves a control problem. One example is the question: "What time does the meeting start? " (Guéant et al., 2011),

where a meeting scheduled for a certain time very often starts several minutes after the scheduled time. The actual time when the meeting starts (T) depends on the dynamics of the arrival of its participants. Each agent decides her (intended) arrival time by minimizing her expected total cost (if the agent arrives earlier than T, she incurs a waiting time cost; otherwise she suffers a cost of lateness due to loss of reputation.) with the assumption that T is known. T is the mean field, i.e. the exhaustive summary for each agent of the behavior of the others. T is a *priori* distribution but can be treated as deterministic due to the "law of large numbers". The equilibrium shows that individual optimization behavior, supposing T is known, fully generates the realization of this time T.

Given the stochastic nature our problem setting (both ratings and sales of each firm are stochastic processes), the equilibrium concept we use in this study has been inspired by the concept of a mean-field. However, we consider heterogeneous players, rather than identical players. Our equilibrium is also similar to rational expectations equilibrium (Muth, 1961) but is beyond this concept owing to the stochastic setting of our problem. In the next section, we present the model and study its properties.

3.3 Model Description

We model a firm's decision with regard to its control effort to manage its online reputation (review ratings or user opinions) assuming rational, profit maximizing behavior. Our discussion begins with a high-level, conceptual model to better understand its inner mechanism and key components.

3.3.1 Conceptual Model

The conceptual model is presented in Figure 3.1 where a focal firm's sales depend not only on its own rating, but also on the ratings of other firms in the market. That is, customers of the focal firm not only consider its ratings, but also the ratings of the other $N-1$ firms in

the market. In theory, therefore, the equilibrium ratings in the market would be the outcome of a $N \times N$ game, whose analysis would be intractable for most reasonable values of N.

To address the above problem, we consider the following equilibrium concept. In equilibrium, customers and firms have a common belief about the average ratings of the firms in the market (market rating). Customers and firms believe that the market rating follows a stochastic process with some mean (μ). That is, while the market rating could fluctuate over time, it has a steady-state mean. This mean market rating is used by customers and firms as the basis for their decisions.

In Figure 3.1, we depict each focal firm's sales as driven by the mean market rating and the rating of the focal firm. Because the mean market rating affects the evolution of a firm's ratings, it must also be considered in the choice of the control effort. Finally, each firm's ratings affect the market mean in equilibrium.

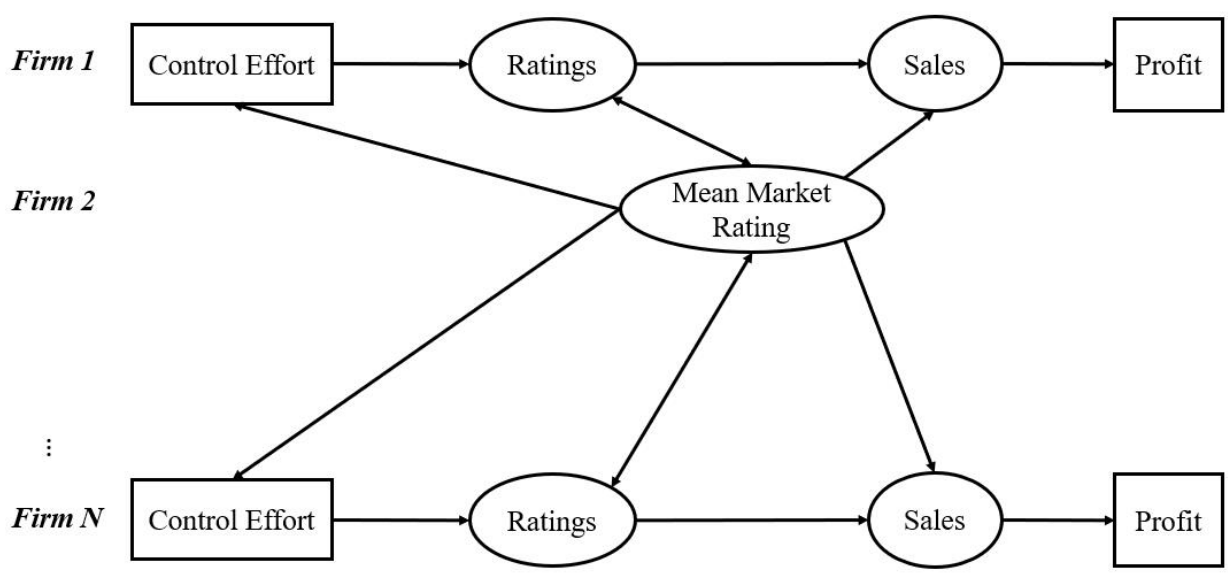

Figure 3.1. Conceptual Model

3.3.2 Mathematical Preliminaries

We first present a model of an individual firm in the platform that exerts effort to control its ratings so as to maximize a profit objective. This problem is solved for a given value of the mean market rating (μ). Later, this value is solved in equilibrium when we solve the N-firm model in the next section. There are two continuous-state, continuous-time stochastic processes corresponding to the two state variables in the firm's control problem: (1) Rating ($x(t)$), and (2) Sales Rate ($S(t)$).[1]

The state variable for ratings (online reputation) at time t is $x(t)$, which can be considered to be the moving average of the n most recent ratings at time t. A reasonable value of n could be the number of customer reviews typically displayed in a single page on the ratings website.

We first describe how the firm's review ratings evolve over time in reaction to its control. The change in the firm's ratings, $dx(t)$, in a small time interval from time t to $t + dt$ is modeled as

$$dx(t) = (u\sqrt{b - x(t)} - \mu x(t))dt + \zeta(x(t))dW(t) \tag{3.1}$$

where u denotes the control effort (u is not necessarily constant and we will solve for u later) exerted by a firm to manage its online reputation, and $\mu x(t)$ captures the potential fall in review ratings when the firm exerts no effort to manage ratings ($u = 0$). The ratings decay at a rate that is proportional to the current rating: if the rating is higher, the decay is faster. Furthermore, as the mean market rating μ increases, the ratings decay faster. This reflects the fact that customers are less tolerant of a firm's lack of effort in a market where the mean market rating is higher. The above model of decay is conceptually similar to the notion of *forgetting* in many traditional advertising models (Sethi, 2004). Finally, b is the upper bound of review ratings, e.g., in many review platforms, this value is 5. The boosting impact

[1]Because, we are first presenting a model for an individual firm, we will suppress a firm specific subscript to avoid clutter.

of the control on the ratings diminishes as the rating improves. This diminishing marginal effect, captured by the square root structure, has been used in many previous studies on the impact of advertising effort on market share or sales (Sethi, 2004; Prasad and Sethi, 2004; Bass et al., 2005). The square root form implies that it is relatively easier for firms to improve their ratings when they are low than when they are high. The stochastic term is a function of the state, $\zeta(x(t))dW(t)$, where $dW(t)$ is the increment of a Wiener process following a normal distribution with mean zero and variance dt. The stochastic term captures all the randomness influencing the ratings; $\zeta(x(t))$ is the magnitude of the stochastic component.

The change in the sales rate over a small time interval from time t to $t + dt$, denoted by $dS(t)$, is described below.

$$dS(t) = (\beta(x(t) - \mu) + \gamma)dt + \sigma(S(t))dZ(t) \tag{3.2}$$

This change is composed of a deterministic (drift) component and a stochastic (diffusion) component. The first term of the deterministic component is proportional to the difference between the firm's rating at time t ($x(t)$) and the mean market rating (μ). The parameter β represents the impact of ratings on the sales rate and can be interpreted as the customers' sensitivity to ratings. The second term of the deterministic component is a *trend* parameter (γ) that captures the impact of factors on the sales rate that are not related to ratings (e.g., the declining or increasing popularity of the cuisine served by a restaurant, a growing population of affluent customers of a tour operator, etc.).[2] The value of the trend parameter can be positive, negative, or zero. The stochastic term is a function of the state, $\sigma(S(t))dZ(t)$, where $dZ(t)$ is the increment of a Wiener process and follows a normal distribution with mean zero and variance dt. The stochastic term captures all the randomness influencing the sales rate; $\sigma(S(t))$ is the magnitude of the stochastic component.

[2]Note that our inclusion of a trend term is structurally similar to that of the Unobserved Component Model (UCM) in time series analysis where a time series is often decomposed into a trend component and other components (see e.g. https://www.amazon.com/Modelling-Unobserved-Components-Matteo-Pelagatti/dp/148222500X).

We do not include price as a control variable in the sales rate equation (3.2). This is consistent with many past studies on advertising competition that exclude price as a variable so as to focus on the role of advertising as a strategic variable. As Erickson (1985) points out, other marketing-mix variables (such as price) are omitted so that advertising as one very important variable can be the focus of the study. Similarly, other studies (Sorger, 1989; Erickson, 1985, 1995, 2009; Fruchter and Kalish, 1997; Fruchter, 1999; Wang and Wu, 2001; Prasad and Sethi, 2004; Bass et al., 2005; Naik et al., 2008) investigate firms' advertising decisions in competitive settings without considering other marketing-mix variables. Our study is a variant of advertising competition, that is, the effort to control ratings can be considered similar to advertising effort. Furthermore, we are considering situations where consumers compare the ratings of similar, competitively priced, products or service. For example, the prices of 4-star hotels in a particular neighborhood of a particular city should not vary much. A concrete example of such a situation can be found in Ctrip.com, the largest online travel platform in China. Controlling for tour characteristics (e.g., departure city, destination city, hotel class, tour guided or not, etc.), the prices of these tours do not vary much. This finding is consistent with prior studies that have shown that price dispersion declines as an Internet market matures and as more retailers join the platform (Venkatesan et al., 2007; Ghose and Yao, 2011).

3.3.3 Stochastic Optimal Control Problem

Equation (3.1) and (3.2) model how the focal firm's control u influences its rating and further sales rate over time. We consider a firm that wishes to maximize its total discounted profit over a planning horizon. The cost of the control (e.g., the cost of the effort needed to improve the online reputation) is assumed to be a convex and increasing function in the control effort: As the control effort increases, the marginal cost of control increases. We

formulate a stochastic optimal control problem as below

$$\begin{aligned} \max_{u} \quad & \mathbb{E}\left[\int_0^\infty e^{-\rho t}(\eta S(t) - cu^2)dt\right] \\ \text{subject to} \quad & dS(t) = (\beta(x(t) - \mu) + \gamma)dt + \sigma(S(t))dZ_(t) \\ & dx(t) = (u\sqrt{b - x(t)} - \mu x(t))dt + \zeta(x(t))dW(t) \end{aligned} \tag{3.3}$$

where η indicates the sales margin, c denotes the cost of control effort, and ρ represents the discount factor. In the objective, $(\eta S(t) - cu^2)$ represents the firm's profit rate at time t; $e^{-\rho t}$ is the continuous discount effect at time t. The total profit is calculated by integrating the profit function from time 0 to ∞. Since the sales rate $S(t)$ evolves stochastically over time, we take expectation of the total profit. Table 3.1 summarizes the parameters and variables of interest in the model.

Table 3.1. Definition of Variables and Parameters

Notation	Definition
$S(t)$	Sales Rate at time t
$x(t)$	Ratings at time t
u	Control Effort
ν	Mean Rating
μ	Mean Market Rating
β	Customer Sensitivity to Ratings
$\sigma(S(t))$	Magnitude of Volatility in Sales Rate
η	Sales Margin
c	Cost of Control Effort
γ	Sales Rate Trend parameter
$\zeta(x(t))$	Magnitude of Volatility in Ratings
ρ	Discount Factor

We solve the stochastic optimal control problem in equation (3.3) (see Appendix B) and obtain the optimal control as shown below.[3]

[3]The proofs for all the mathematical results in this chapter are provided in Appendix B.

Theorem 1. *The optimal control effort depends on the firm's rating $x(t)$ (i.e., a feedback control); the firm exerts more effort when its current rating is low.*

$$u^* = \alpha\sqrt{b - x(t)} \tag{3.4}$$

where $\alpha = \sqrt{(\rho+\mu)^2 + \frac{\beta\eta}{\rho c}} - (\rho+\mu)$.

The above control falls within the class of a *feedback* control, i.e., the control depends on the current state. In our context, it is reasonable that the optimal effort exerted by the firm is higher when the current rating is lower. Also, for a given market mean rating, the control increases with the sales margin rate (η) and the customer sensitivity parameter (β), but decreases with the cost of the control effort (c). This is also reasonable since the firm has a greater incentive to increase the rating when the sales margin rate or the customer sensitivity parameter is higher. On the other hand, there is less incentive to increase effort if the cost of control increases.

Substituting the optimal control u^* back into the ratings equation in (3.3) and rewriting, we get the trajectories of two controlled diffusion processes, i.e., processes that embed the optimal control strategy as shown below.

$$dS(t) = (\beta(x(t) - \mu) + \gamma)dt + \sigma(S(t))dZ(t) \tag{3.5a}$$

$$dx(t) = \lambda(\nu - x(t))dt + \zeta(x(t))dW(t) \tag{3.5b}$$

where $\nu = \frac{A-(\rho+\mu)}{A-\rho}b$, $\lambda = A - \rho$, and $A = \sqrt{(\rho+\mu)^2 + \frac{\eta\beta}{\rho c}}$.

The controlled rating process ($x(t)$) follows a mean reverting process where it fluctuates around steady state mean ν, with λ being the speed of reversion. The mean-reverting property is consistent with the "beg to differ" effect that customers prefer to express their complaint differently (Moe and Schweidel, 2012; Ho et al., 2017). The steady state mean rating of the firm (ν) decreases as the rating decay factor (μ) increases. On the other hand, the sales rate process is a more complex process (note the presence of the stochastic variable

$x(t)$ in the deterministic term) and does not have a well-behaved form. However, under certain structures of the stochastic term, it is possible to obtain a closed-form expression for the expected sales rate at time t, given the initial states S and x, denoted by $\mathbb{E}(S(t)|S,x)$ (See Appendix B).

The firm's value function $V(S,x)$ that is the total discounted profit earned by adopting the optimal control is shown below.

Theorem 2. *The optimal profit function depends on the various parameters of the firm, including the initial sales rate (S) and the initial rating (x).*

$$V(S,x) = \frac{\eta}{\rho}S + 2c(A-\rho-\mu)x - \frac{\eta\beta\mu}{\rho^2} + \frac{bc}{\rho}\left[\frac{\eta\beta}{\rho c} - 2(\rho+\mu)A + 2(\rho+\mu)^2\right] + \frac{\eta\gamma}{\rho^2} \quad (3.6)$$

3.4 Equilibrium Analysis

We now introduce a differential game where there are N firms engaged in ratings competition. Each firm is represented with parameters: β_i (customer sensitivity to ratings for firm i), η_i (sales margin for firm i), sales rate trend (γ_i), and c_i (cost of control for firm i). The parameter ρ) is considered common to all the firms in the market (platform).

After solving firm i's stochastic optimal control problem (for a given μ), its optimal control effort is

$$u_i^* = \alpha_i\sqrt{b - x_i(t)},$$

where $\alpha_i = \sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} - (\rho+\mu)$.

Substituting the optimal control u_i^* back into the state equations, we get the trajectories of two controlled diffusion processes for firm i as shown below.

$$dS_i(t) = (\beta_i(x_i(t)-\mu)+\gamma_i)dt + \sigma_i(S_i(t))dZ_i(t)$$

$$dx_i(t) = \lambda_i(\nu_i - x_i(t))dt + \zeta_i(x_i(t))dW_i(t)$$

where

$$\nu_i = \frac{A_i - (\rho + \mu)}{A_i - \rho} b,$$

$$\lambda_i = A_i - \rho,$$

and

$$A_i = \sqrt{(\rho + \mu)^2 + \frac{\eta_i \beta_i}{\rho c_i}}.$$

The profit of firm i, given the initial sales rate (S_i) and the initial rating (x_i), is

$$V_i(S_i, x_i) = \frac{\eta_i}{\rho} S_i + 2c_i(A_i - \rho - \mu)x_i - \frac{\eta_i \beta_i \mu}{\rho^2} + \frac{bc_i}{\rho}\left[\frac{\eta_i \beta_i}{\rho c_i} - 2(\rho + \mu)A_i + 2(\rho + \mu)^2\right] + \frac{\eta_i \gamma_i}{\rho^2}.$$

Firm i's steady state mean rating is ν_i. The market rating ($z(t)$) is a stochastic process given by $z(t) = \frac{\sum_{i=1}^{N} x_i(t)}{N}$. Because, the controlled diffusion process for $x_i(t)$ has a steady state mean, it follows that $z(t)$ also has a steady state mean. In equilibrium, we have

$$\frac{1}{N}\sum_{i=1}^{N} \nu_i = \mu,$$

Therefore we have,

$$\mu = \frac{b\sum_{i=1}^{N} \frac{A_i - (\rho+\mu)}{A_i - \rho}}{N}. \tag{3.7}$$

The equilibrium mean market rating should satisfy the equation (3.7).

3.4.1 Equilibrium Mean Market Rating

Theorem 3. *A unique equilibrium exists for the mean market rating* $\mu = g(\eta_i, \beta_i, c_i, \rho, N, b)$.

Note that although we do not provide a closed form expression for the mean market rating in equilibrium, we prove the existence and uniqueness of the mean market rating in equilibrium (see Appendix B). Furthermore, its value can easily be recovered using a simple search procedure. The mean market rating in equilibrium depends on the sales margin, cost of control, as well as customers sensitivity to each firm in the market, the total number of firms in the market, the discount factor, and the upper bound of ratings.

Intuitively speaking, the reason for the existence of an equilibrium is as follows. Assume that the equilibrium mean rating for a given market exists for $\mu = \mu^\star$. However, let us say we provide a value of $\mu^{'} > \mu^\star$ as a parameter for each firm to solve their stochastic control problems. Since the provided value is higher than the equilibrium value, each firm will respond with a control effort that is lower than the one that they would choose at equilibrium ($\frac{\partial u^\star}{\partial \mu} < 0$). This would, in turn, lower the mean ratings of each firm so the response from the firms would be such that the average of the ratings of the firms would be lower than $\mu^{'}$, say $\mu^{''}$. If $\mu^{''} > \mu^\star$, then the above process would repeat until the equilibrium value of $\mu^\star$ is reached. In a similar way, we can understand how the equilibrium value of $\mu^\star$ would be reached if the starting value of the market parameter was provided below the true equilibrium value. To summarize, the equilibrium can be reached by a sequence of adjustments, starting above or below the true value. This is because the responses of the firms as such that they naturally push the next term in the sequence closer to the true value: When the parameter for the mean is above the true value, the responses push this value down, otherwise the responses push this value up.

3.4.2 Example

We illustrate the market equilibrium for a simple case of $N = 3$ firms in the market.[4] The market rating ($z(t)$) is stochastic and is calculated as

$$z(t) = \frac{x_1(t) + x_2(t) + x_3(t)}{3},$$

the average of the ratings of the three firms. Figure 3.2 and 3.3 illustrate each firm's ratings and sales rates as they evolve over time. We see that both the ratings and the sales rates

[4]We choose $\eta_1 = 13.89$, $\eta_2 = 13.25$, $\eta_3 = 13.3$, $\beta_1 = 6.55$, $\beta_2 = 8.79$, $\beta_3 = 1.24$, $c_1 = 1.71$, $c_2 = 1.95$, $c_3 = 1.3$, $\rho = 1$, $b = 5$, $\sigma_1 = 0.5S_1(t)$, $\sigma_2 = 0.5S_2(t)$, $\sigma_3 = 0.5S_3(t)$, $\zeta_1 = 0.1\sqrt{b - x_1(t)}$, $\zeta_2 = 0.1\sqrt{b - x_2(t)}$, $\zeta_3 = 0.1\sqrt{b - x_3(t)}$, $\gamma_1 = \gamma_2 = \gamma_3 = 0$, $x_1(0) = x_2(0) = x_3(0) = 3$, $S_1(0) = S_2(0) = S_3(0) = 20$, $\Delta t = 0.01$, , and $N = 3$.

fluctuate over time. In each figure, we also plot the expected ratings and expected sales rates over time. The formulas of expected sales rate and ratings can be found in Appendix B. The mean ratings for the three firms are $\nu_1 = 3.13$, $\nu_2 = 3.22$, and $\nu_3 = 1.74$. Figure 3.4 shows how the market rating is also stochastic with a mean value, $\mu = 2.69$. Firm 1's (Firm 3's) rating on average is higher (lower) than the mean market rating, and therefore, the sales rate of firm 1 is higher than firm 3 on average. The control effort exerted by each firm is depicted in Figure 3.5. This figure shows that Firm 2 exerts more effort although its ratings are higher. This is because the effort coefficient (α) for Firm 1 is higher, owing to the higher sales margin for this firm ($\beta_1(8.79) > \beta_2(6.55)$).

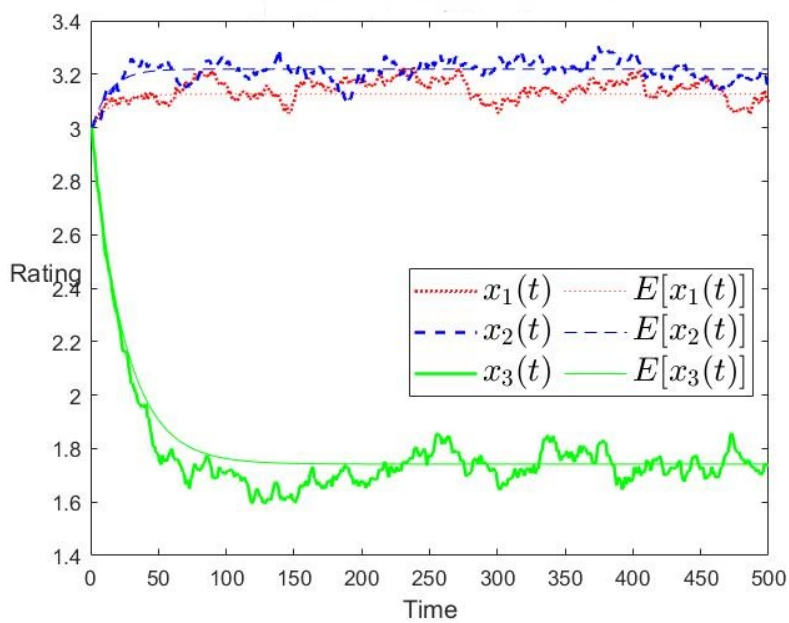

Figure 3.2. Rating Trajectory of each firm in the Market

Although the market rating is a stochastic process, in equilibrium, customers use the mean of this process as decay factors affecting the sales rate and the ratings. Also, in equilibrium, the mean of the market rating is the mean of the ratings of all firms, i.e., $2.69 = \frac{3.13+3.22+1.74}{3}$. We note that the equilibrium can be asymmetric, i.e., $\nu_1 \neq \nu_2 \neq \nu_3$. The profits of Firm 1 (V_1), Firm 2 (V_2), and Firm 3 (V_3) are 250.65, 239.51, and 246.26.

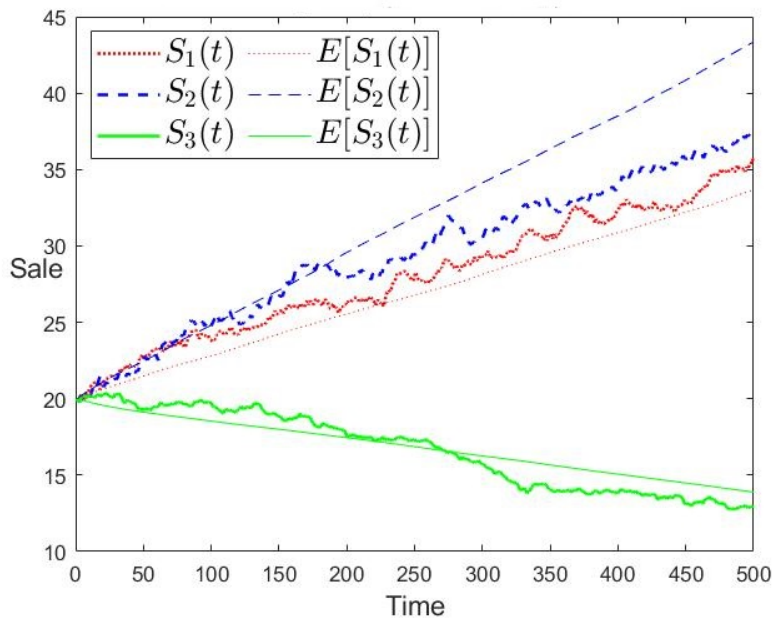

Figure 3.3. Sales Rate Trajectory of each firm in the Market

Note that a firm in the market can have a mean rating below the market mean, yet make positive profit (e.g., $\nu_3 < \mu$ but $V_3 = 246.26 > 0$).

3.5 Results and Discussion

We present and discuss several results relating to the impact of different parameters on the mean market rating in equilibrium and the profit of one firm in equilibrium. The impacts (on equilibrium mean market rating and profit) where the parameters of more than one firm change are also studied. Finally, several results relating to the market structure (heterogeneity among firms, entry and exit of firms) are studied.

3.5.1 Results

We begin with two propositions concerning the impact of various parameters on the mean market rating and the profit of a firm in equilibrium.

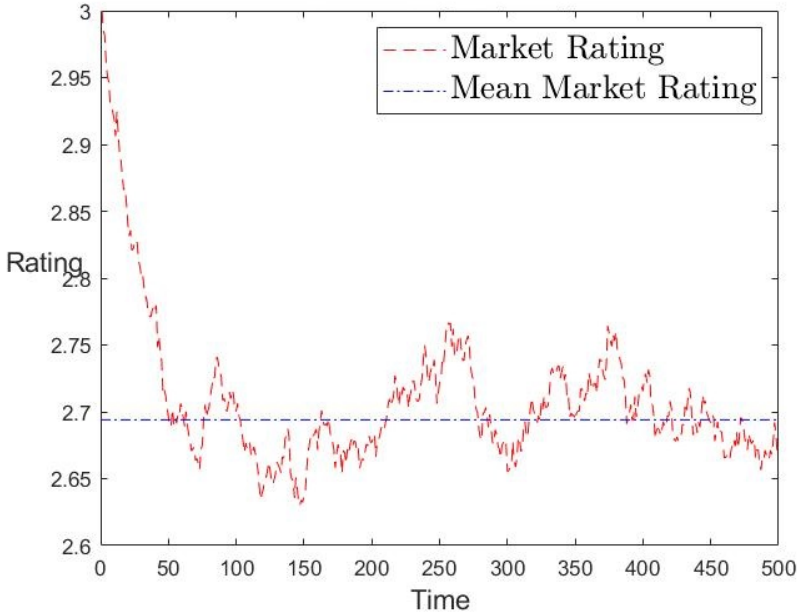

Figure 3.4. Trajectory of Market Rating Over Time

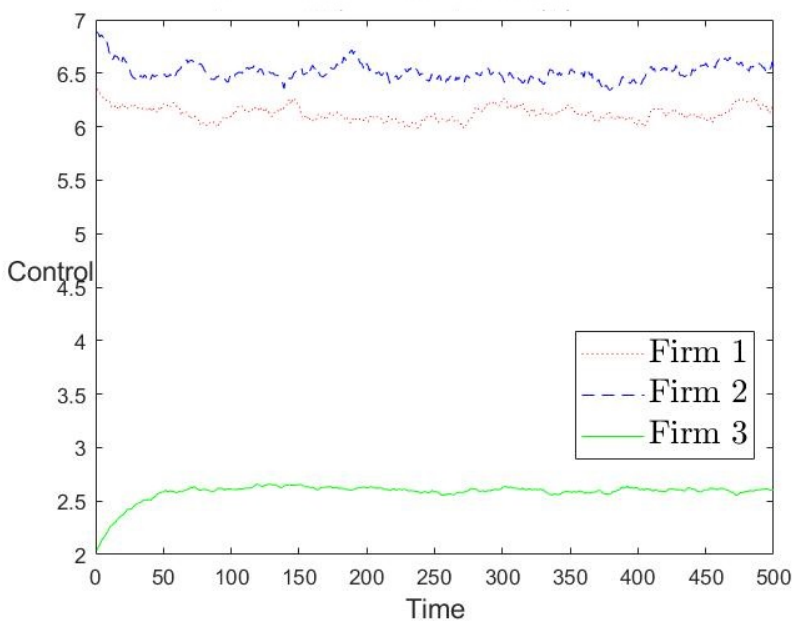

Figure 3.5. Trajectory of Control Effort of each firm in the Market

Proposition 1. *The mean market rating in equilibrium changes as the parameters of a firm (cost of control effort, sales margin, and the customer sensitivity parameter) change. The parameters for the other firms in the market are held constant.*

(i) For any firm i, as its cost of control effort c_i increases, the mean market rating in equilibrium μ decreases ($\frac{d\mu}{dc_i} < 0$).

(ii) For any firm i, as its sales margin η_i increases, the mean market rating in equilibrium μ increases ($\frac{d\mu}{d\eta_i} > 0$).

(iii) For any firm i, as the customers are more sensitive to its ratings (higher β_i), the mean market rating in equilibrium μ increases ($\frac{d\mu}{d\beta_i} > 0$).

In Proposition 1, we note that the impact of a parameter (such as the cost of control effort, sales margin, and customer sensitivity) on the equilibrium mean market rating is intuitive. For example, if the cost of control effort (c_i) for a particular firm increases, it discourages this firm to invest in effort to improve its ratings. As a result, the ratings of this firm fall, leading to a decrease in the equilibrium mean market rating. On the other hand, an increase in the sales rate margin (η_i) or the ratings sensitivity parameter (β_i) provides a greater incentive for the firm to improve its ratings. It does so by increasing the control effort and hence, its ratings increase. This leads to an increase in the mean market rating.

Corollary 3.1. *The magnitude of the self-impact of an event affecting a firm is greater than the market-impact of the same event.*

This result follows from the definition of the market mean, $N\mu = \sum_{k=1}^{N} \nu_k$. Differentiating both sides with respect to ϕ_i, where ϕ_i can be η_i, c_i, or β_i, we get

$$\frac{d\nu_i}{d\phi_i} = N\frac{d\mu}{d\phi_i} - \sum_{k\neq i}^{N} \frac{d\nu_k}{d\phi_i}.$$

In Appendix B, we show that the sign of $\frac{d\nu_k}{d\phi_i}$ is opposite to that of $\frac{d\mu}{d\phi_i}$, while the sign of $\frac{d\nu_i}{d\phi_i}$ is the same as the sign of $\frac{d\mu}{d\phi_i}$. Hence, it is easy to see that the magnitude of $\frac{d\nu_i}{d\phi_i}$ (self-impact) is greater than the magnitude of $\frac{d\mu}{d\phi_i}$ (market-impact). The implication of the above corollary is that the size of the market (N) dilutes the market impact of an event affecting a firm. This supports the equilibrium concept we use in this study, namely, that each firm *behaves* (or optimizes) its own profit assuming that its actions have no impact on the market mean. However, at the end, the market mean is affected by the aggregate behavior of the firms.

Proposition 2. *The focal firm's profit in equilibrium changes as the parameters (cost of control effort, sales margin, and the customer sensitivity parameter) change. The parameters for the other firms in the market are held constant.*

(i) For any firm i, as its sales margin η_i increases, its equilibrium profit V_i increases if and only if $N - f_{\eta_i}(N) > 0$.

(ii) For any firm i, as the customers are more sensitive to ratings (higher β_i), its equilibrium profit V_i increases if and only if $N - f_{\beta_i}(N) > 0$.

(iii) For any firm i, as its cost of control effort c_i increases, its equilibrium profit V_i increases if and only if $N - f_{c_i}(N) < 0$.

Where the expressions of $f_{\eta_i}(N)$, $f_{\beta_i}(N)$, and $f_{c_i}(N)$ can be found in Appendix B.

The impact of a change in a parameter on the equilibrium profit is more subtle. For example, take the case of an increase in the sales margin of a particular firm. Proposition 2 (i) asserts that this increase could have a positive or a negative impact on the profit of this firm. In Proposition 1 (i) we observe that an increase in the sales margin increases the equilibrium mean market rating. Therefore, with an increase in the sales margin of a firm, there are two opposing forces: (1) A *direct* effect that must be positive (because, everything

else held constant, a higher sales margin should lead to higher profit), and (2) An *indirect* (negative) effect that acts through the increase in the mean market rating. However, the *net* impact on profit can be positive or negative. This is because the (positive) direct effect can dominate the (negative) indirect effect when there are a sufficiently large number of competing firms in the platform. The reverse is true when the number of competing firms is relatively small. The impact of an increase in the customer sensitivity parameter for a firm has an effect similar to the sales margin.

The impact of a cost increase can be explained in a similar way: As the cost of a firm increases, it has less incentive to invest in effort, implying that its ratings decrease. However, because the market ratings also decrease, the net impact on the profit of the firm could be positive, despite an increase in the cost.

In Proposition 2 above, we examined the profit impact of a situation when only the focal firm was affected by an event. We next present a corollary concerning a situation where an event affects more than one firm. The corollary is demonstrated using a numerical example.

Corollary 3.2. *An event acting alone could hurt a firm's profit. However, the same event could be beneficial if it affects another firm in a specific way.*

Take a market with two firms (denoted by 1 and 2). Let us assume that the control cost of these firms changes from c_1 and c_2 to $c_1^{'} > c_1$ and $c_2^{'} >> c_2$, where the cost of control for Firm 2 increases more (see Table 3.2). To see why the result could occur, we note that if Firm 1 alone experienced an increase in its cost of control, it would exert less control effort and its ratings would fall. As a result, the mean market rating would also fall. If the cost of control for Firm 2 did not change, its profit would increase because it would benefit from the fall in the mean market rating. However, if its cost of control also increases, some of the profit benefit would be diluted. The net impact on the profit of a firm would occur from a combination of a direct (negative) effect of its increased cost of control and an indirect

(positive) effect of a reduced mean market rating. Mathematically, the above argument could be reduced to the inequalities below.

$$\frac{dV_1}{dc_1} + \frac{dV_1}{dc_2} > 0$$

$$\frac{dV_2}{dc_1} + \frac{dV_2}{dc_2} < 0$$

If the above inequalities were simultaneously true, Firm 1 would experience an increase in its profit despite an increase in its cost of control. We know that $\frac{\partial V_1}{\partial c_1} < 0$ and $\frac{\partial V_2}{\partial c_2} < 0$. That is, if everything else is held constant, the profit of a firm must decrease if its cost of control increases. Also, the total impact of an increase in control cost on the profit ($\frac{dV_1}{dc_1}$) can be negative. On the other hand, $\frac{dV_1}{dc_2} > 0$, because the impact of an increase in the control effort of another firm decreases the market equilibrium, which in turn, increases the profit for the focal firm. Or mathematically, we have

$$\underbrace{\frac{dV_1}{dc_2}}_{>0} = \underbrace{\frac{dV_1}{d\mu}}_{<0} \times \underbrace{\frac{d\mu}{dc_2}}_{<0}.$$

Thus, the joint effect on V_1 of an increase in the costs c_1 and c_2 can be positive. We illustrate this possibility in Table 2.

In Table 3.2, we observe that the profit of Firm 1 decreases when *only* its cost increases. However, its profit increases despite the same increase in its own cost when the cost associated with Firm 2 also increases. The implication of this finding is that when an adverse event (such as an increase in the wage rate in a local area, leading to an increase in the control cost in that area) strikes two firms, the firm that manages this adverse event better (e.g., by changing its operations to better respond to the cost increase), can turn the adversity to an advantage.

Proposition 3. *The entry of a new firm has consequences for the existing (incumbent) firms, the platform and the consumers.*

Table 3.2. Impact of Cost on Profit When One Firm and Two Firms are Affected

Firm	c_i	V_i	c_i'	V_i'
Firm 1	1.0	14,094	1.1	13,764
Firm 2	1.0	10,143	1.0	10,411
Firm 1	1.0	14,094	1.1	14,948
Firm 2	1.0	10,143	2	6,813

Note. $\beta_1 = 1$, $\beta_2 = 2$, $\eta_1 = 100$, $\eta_2 = 80$, $\rho = 0.1$ $S_1(0) = S_2(0) = 20$, $x_1(0) = x_2(0) = 4$, $b = 5$, $\gamma_1 = \gamma_2 = 0$.

(i) The entry of a low rating firm increases the total profit of the market and also the profits of incumbent firms. However, it lowers the mean market rating.

(ii) The entry of a high rating firm lowers the profits of the incumbent firms. It increases the mean market rating.

(iii) An average firm's entry increases the total profit of the market and leaves the incumbent firms and the market mean rating unaffected.

In equilibrium, there are two aggregate outcomes of interest to consumers and the platform. The mean market rating can be considered to be a measure related to consumer welfare, whereas the total profit of all the firms in the market can be used as a measure of the platform's goal.

The entry of a low rating firm can have two reinforcing effects. If the ratings of other firms do not change, the market mean should clearly be lower. As a consequence, the other firms in the market could decrease their efforts when faced with a lowered market mean. Hence, the market mean should reduce even more in equilibrium. The analysis of the impact of a low rating firm becomes more complex when the reaction of the existing firms in the market is considered. It can be shown the even if the reaction of the existing firms is taken into account, the equilibrium mean market rating decreases. This increases the profits of the incumbent firms in the market though. The total profit (including the new entrant) is

clearly higher than the total profit before the entry of the low rating firm. Platforms often charge a brokerage fee from firms. Thus, the higher total profits of firms, the better it is for the platform.[5]

On the other hand, a high rating firm (with a higher than average rating) has two opposing effects. If the existing firms do not react, the entry of a high rating firm should result in an increase in the mean market rating. This reduces the profits of the incumbent firms in the market. Hence, the total profit of the incumbent firms decreases from the level before the entry of the high rating firm. It can be shown that the above conclusion holds even when we account for the fact the existing firms react to the entry of the high rating firm. However, if we include the profit of the new entrant, the total profit of the market can be either higher or lower.

Table 3.3 numerically illustrates the impact of a new entry firm on the incumbent firms and the total profit of the market. Assume there are three firms (Firm 1, Firm 2, Firm 3) originally in the platform (before the entry), and one firm (with high rating, Firm 4) joins the platform later. The mean market rating in equilibrium changes from 2.63 to 2.79 after the entry. In Scenario a, the total profit of the market decreases from 209.16 to 201.92 after the entry. In Scenario b, the total profit of the market increases from 209.16 to 241.92 after the entry. For the incumbent firms (Firm 1 to 3), each of their profits decreases after the entry.

Thus, the entry of a high rating firm can be good or not so good for the platform (total profit increases or decreases), good for consumers (mean market rating is higher), but not so good for the incumbent firms (each incumbent firm's profit reduces).

When a firm with average rating (equal to the market mean) enters the market, there is no impact on the market mean. Hence, there is no impact on the incumbent firms in the market. However, the total profit (including the new entrant) must be higher than what it

[5]For example, commission rates at Expedia range from 20-25% (Page, 2018).

Table 3.3. Illustrating the Impact of Entry

	Before Entry	After Entry Scenario a	After Entry Scenario b
Firm 1's Profit	113.42	91.73	91.73
Firm 2's Profit	53.83	50.70	50.70
Firm 3's Profit	41.91	35.23	35.23
Firm 4's (Entry) Profit	–	24.27	64.27
Total Profit	209.16	201.92	241.92

Note. $\rho = 1$, $b = 5$, $S_1(0) = S_2(0) = S_3(0) = 5$, $x_1(0) = x_2(0) = x_3(0) = x_4(0) = 4$, $\beta_1 = 5.17$, $\beta_2 = 1$, $\beta_3 = 2.47$, $\eta_1 = 17.2$, $\eta_2 = 13.02$, $\eta_3 = 10.92$, $c_1 = c2 = c3 = 1$, $\beta_4 = 10$, $\eta_4 = 10$, $c_4 = 1$, $S_4(0) = 5$ (Scenario a), $S_4(0) = 1$ (Scenario b), $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$.

was before the average rating firm entered. The entry of an average rating firm is good for the platform (total profit is higher), but the incumbent firms and consumers as not affected.

Table 3.4 summarizes how a new entry influences the market rating in equilibrium (related to consumer welfare), the profits of the incumbent firms in the platform, and the total profit of the firms in the platform (related to the platform's goal).

Table 3.4. Impact of a New Entry

New Entry	Market Rating	Profits of Incumbent Firms	Total Profit of Market
Low Rating Firm	Decrease	Increase	Increase
Average Rating Firm	Unaffected	Unaffected	Increase
High Rating Firm	Increase	Decrease	Increase or Decrease

Corollary 3.3. *The exit of a firm has consequences for the existing (incumbent) firms, the platform and the consumers.*

(i) The exit of a high rating firm lowers the mean market rating, and increases the profits of the incumbent firms.

(ii) The exit of a low rating firm increases the mean market rating, lowers the profits of incumbent firms, and lowers the total profit of the market.

(iii) The exit of an average firm leaves the incumbent firms and the market mean rating unaffected, and decreases the total profit of the market.

The exit of a high rating firm should decrease the mean market rating, everything else held constant. It is possible to prove that the direct impact of the exiting firm (i.e., one of lowering the market mean), is more influential than the indirect impact (i.e., one of lowering the number of firms in the market). Because the new mean market rating is lower, the remaining firms enjoy higher profits ($\frac{dV_i}{d\mu} < 0$). However, this exit also decreases the number of firms in the market (from N to $N-1$), and hence it is not immediately clear if the total profit of the market must decrease as a result of the exiting firm. The exit of a low rating firm has the opposite effect: it increases the mean market rating and lowers the profits of incumbent firms. The total profit of the market decreases. Finally, the exit of the average firm will leave the mean market rating and profits of the remaining firms unchanged, but will lower the total profit of the firms, since the average firm no longer contributes to the total profit.

Proposition 4. *A more heterogeneous market (one where the parameters of the firms are very different) leads to a lower mean market rating in equilibrium.*

We study heterogeneity among firms using the following structure. Define a composite parameter for each firm as,

$$\delta_i = \frac{\eta_i \beta_i}{\rho c_i}.$$

Let $\bar{\delta}$ be the mean value of the parameter for the firms in the market. Next, let each firm's parameter differ by a factor α from the firm immediately above (or below) it. That is,

$$\delta_i = \bar{\delta} + \alpha \left(i - \frac{N+1}{2} \right),$$

where $i = 1, 2, ..., N$, and N is assumed to be odd. If N is even, we can add a virtual firm to represent the middle firm and ignore it for the purposes of calculating for the market mean and total profit.

In the above, the firm in the middle has the mean parameter ($\bar{\delta}$), and the firms above (below) arithmetically increase (decrease) from the mean value by the value α. In this structure, the parameter α represents the level of heterogeneity in the market. As α increases, the market is more heterogeneous. The result in the above proposition states that a higher value of α leads to a reduction in the mean market rating in equilibrium ($\frac{d\mu}{d\alpha} < 0$). The main implication of this result is for platforms. As it selects firms with more diverse parameters (as opposed to niche firms with similar parameters), it reduces the competition among the firms the platform. This lowers the market mean, implying that consumers receive lower quality or service. In next proposition, we consider the impact of market heterogeneity on the profits of the firms.

Proposition 5. *The total profit of the firms in the market increases with an increase in the extent of heterogeneity in the market.*

To understand the above result, we first let the cause of the heterogeneity arise from difference in the sales margin (η_i). It is possible to show that for $i > \frac{N+1}{2}$, $\frac{dV_i}{d\alpha} > 0$ (See Appendix B). That is, the firms with higher than average sales margin will earn more profit as the firms become more differentiated (higher α). However, a firm with lower than the average sales margin could earn more or less profit. Importantly, the total profit of the market (sum of the profits of all the firms) is higher with more heterogeneity (See Appendix B). A similar argument can be made when the source of heterogeneity comes from customer sensitivity (β_i) or the cost of control effort (c_i).

The above results can be interpreted in terms of the intensity of competition in the market. As the firms become more heterogeneous, the intensity of competition should reduce. This should result in a higher total profit in equilibrium.

3.5.2 Managerial Implications

Our study has three important take-aways for platform executives and participating firms: (1) Aim for Strategic Growth, (2) Diversity is Key, and (3) Outrun your competitor. We discuss these below.

Aim for Strategic Growth

Growth is often a core value in the business plan of any organization. Towards this end, many platforms have advocated a "get big fast" strategy to grow the platform (Cennamo and Santaló, 2013), e.g. by continually soliciting new firms to join in the platform at all costs. However, the platform needs to be strategic about growth. Cennamo and Santaló (2013) show that such a "growth at all costs" strategy can be perilous without carefully considering the potential conflicts and disincentives of such a growth strategy creates, particularly among the service providers on the platform. For example, we saw that the entry of a relatively low rating firm may increase the total profit of the market (i.e., sum of profits including the new entrant). However, the platform runs into a dilemma here. When a low rating entrant joins, the incumbent firms benefit but the consumers experience a market with lower mean market rating than before. On the other hand, when a new, high rating firm enters, the market mean increases (beneficial for consumers), but the incumbent firms earn lower profits and the total profit (including the new entrant) may be lower or higher.

Thus, the platform needs to balance its recruiting activities between low rating and high rating firms, referred to as the platform governance problem in the literature (Tiwana et al., 2010; Wareham et al., 2014; Huotari, 2017; Rietveld et al., 2019). Even when the platform benefits (higher total profit), one of its stakeholders (incumbent firms or consumers) could be hurt as a result of the new entry. Encouraging average firms to join could turn out to be most beneficial because it has no impact on the incumbent firms or consumers, but there is a limited extent to which the platform can find such firms and encourage them to join. In such cases, as much as possible, the platform should attempt to grow at the middle rather

than at the ends. Growing at the low end of the market is beneficial to incumbents but the loss to consumers (resulting from a lower market mean) must be kept in mind. In the view of such a tradeoff, Apple, for example, stipulated strict quality standards for its App developers (Huotari, 2017). Growing at the top end is good for consumers and sometimes for the platform, but the loss to incumbent firms must be weighed in while attempting to grow at the high end of the market. This is echoed by Tadelis (2016) where it pointed out that promoting seller quality may come at the expense of crowding out incumbents. Rietveld et al. (2019) also show that platform owners of games do not simply promote best in class games, rather they strategically invest in under-appreciated games where there is greater marginal value to be unlocked.

Diversity is Key

Platforms can benefit from our results concerning the heterogeneity of firms that participate in the platform. If a platform only targets similar firms (e.g., firms with similar costs and margins), it could result in intense competition, leading to lower firm profits and the total profit. Our findings show that diversity among firms is good for the platform and the firms in the platform. The caveat, however, is that the mean market rating decreases with more diversity. Thus the platform needs to carefully balance the characteristics of firms with relatively high margins with those with relatively low margins. This finding is consistent with Cennamo and Santaló (2018) and Wareham et al. (2014) where they show the need to have heterogeneous firms to serve diverse consumers and to meet evolving market demand. Nintendo in the 1990s, for instance, carefully curbed the number of competing video game titles by forbidding game developers to launch more than five titles in a category (Williams, 2002).

Don't Try to Outrun the Bear, Outrun Your Competitor

Concerning firms, our results allow a firm to anticipate the impact of changes in various parameters on the profit. For example, contrary to what one could expect, a firm could

welcome an increase in the cost of control effort (for example, if the costs of certain inputs such as labor or materials increase) if it believes that it can control the cost increase better than its competitors. As Hass and Rigby (2004) succinctly put it "The key to survival? Play the bears game [6] while others become (the prey of) it." Winning competitors are often those who are able to scrutinize their supply chains, store labor deployment, marketing programs, and overhead costs to eliminate wasted dollars to compete.

While consumers do not play a strategic role in this study, they are clearly an important stakeholder. To this end, we have provided several results with regard to the impact of various parameters on the mean market rating in equilibrium. As the mean market rating improves, it benefits consumers because they receive a higher quality product or service.

3.6 Conclusion

The ubiquity of online customer reviews is reshaping consumer perception on products or services that are being evaluated for purchase. According to a report from TripAdvisor, more than 90% of business representatives rated online reviews as one of the most important factors for their business (GuestRevu, 2016). Given the importance of online reputation, we address the question how competition impacts a firm's control effort to manage its online reputation and hence its profit. The answer to this question has important implications for the governance of platforms – that constitutes the main focus of this study.

We consider a competitive setting where all firms in a market attempt to manage their online reputations (measured by review ratings). Because finding a Nash equilibrium is intractable, we invoke the notion similar to a mean field equilibrium, where consumers interpret a firm's ratings in a relative sense, anchored around a belief about the mean market

[6] The bear game (a popular joke) is about two campers attempting to run away from bear. One of them puts on his running shoes. Upon this, the other remarks that the shoes won't help outrun the bear. The other responds: "I don't need to outrun the bear, I just need to outrun you."

rating. Then, each firm's sales are driven by its own ratings and this mean-field belief. We develop a controlled diffusion model where a firm's sales are driven by the common belief and its own ratings that are maneuvered by the control effort, chosen to maximize its profit. Thus, we derive a controlled diffusion model for the evolution of the ratings and sales of a firm. We prove the existence and uniqueness of the equilibrium mean market rating, which depends on each firm's sales margin, cost of control effort, as well as customer sensitivity to ratings, the total number of firms in the market, the discount factor, and the upper bound of ratings.

We find that when the customer sensitivity to ratings is high or when the sales margin is high, the firm exerts more effort (but at a diminishing rate) to boost its mean rating. However, it results in an increase in the mean market rating in equilibrium. This finding reflects a key structural result in this study that applies in several situations: A particular firm's actions directly impact its outcomes (such as ratings and profit) but indirectly impact these outcomes by acting through the equilibrium mean market rating. Interestingly, the direct and indirect impacts pull in opposite directions. When the direct impact is positive (negative), the indirect impact is negative (positive). Another interesting finding in this study is that, for a focal firm, the joint impact on profit of a change in the parameters of multiple firms can be different from the impact if the parameters of only the focal firm were to change. For example, if only the sales margin of one firm were to decrease, it could decrease its profit. However, if the sales margin of a competing firm were to decrease more, then the focal firm's profit could increase despite a decrease in its sales margin. This somewhat unexpected result arises from the endogenous nature of the mean market rating.

We provide several insights concerning the properties of the equilibrium mean market rating and the equilibrium profit. For example, a more heterogeneous market (one where the parameters of the firms are very different) leads to a lower mean market rating and higher total profit of the firms in the market. In addition, our results can inform platforms on how

to target certain firms to join, depending on their ratings. Adding firms with average ratings (i.e., the middle of the market) is the safest option considering the goals of the platform (increase total profit) and the other stakeholders, namely, incumbent firms and consumers. On the other hand, growing the market at the bottom hurts consumers (i.e., adding firms with lower than average ratings), but benefits both the platform and the incumbent firms. Finally, growing at the top is good for the platform and consumers, but hurts the profit of incumbent firms.

This study is not without limitations. We considered a homogeneous market where all firms compete with one another. Instead, we could consider a network-like structure where the weight between two nodes (firms) represents the intensity of competition between the nodes. Also, in addition to *intra* market competition, there could be *inter* market competition. That is, beyond competing firms in one segment, there could be another layer of (possibly weaker) competition across segments. Finally, we did not consider temporal effects, but instead, restricted ourselves to a steady-state analysis. Given the stochastic nature of the setting, future work could consider how firms jostle one another for competitive positions in the market, and how the equilibrium outcomes (mean market rating, mean individual ratings and control efforts, etc.) unfold over time.

CHAPTER 4

ARE YOU PAYING TOO MUCH FOR FINANCIAL ADVICE? THE TRANSPARENCY-REVENUE CONUNDRUM IN SOCIAL TRADING

4.1 Introduction

The emergence of new financial technologies (Fintech) has helped bridge the digital divide of financial services, especially in terms of access to financial advising and wealth management services. Among these technologies is social trading, through which a retail investor can manage her wealth (no matter how small it is) by directly following the financial advice of other traders (Eldridge, 2017). These traders often share their opinions (financial advice) on financial markets through specialized social media platforms (e.g., StockTwits and Seeking Alpha), where traders can make friends, post their opinions, and directly communicate with other investors (Doering et al., 2015). Investors can then use such financial advice to make trading decisions.

Recently, a more disruptive type of social trading business model has emerged, represented by platforms such eToro, Zulutrade, and Collective2. These firms go beyond merely providing a platform for traders to share financial advice; rather, they allow investors to observe and follow the actual trading activities of peer traders on stocks, currencies, and cryptocurrencies (Pelster and Hofmann, 2017). Our focus in this study is on this new type of social trading, referred to as *copy trading*.

Social trading platforms are gaining popularity as evidenced by multiple rounds of venture capital funding (Reuters, 2018) and the growing pool of traders on these platforms with over 13.9 million online retail traders as of 2018 (BrokerNotes, 2018). eToro alone has attracted over 9 million active users by 2018, mostly small retail investors, who are allowed to open an account with as little as 200 USD.[1] The demand for social trading services is

[1]https://bitcoinist.com/9-million-traders-flocked-etoros-social-trading-platform/

projected to explode in the future (Empire, 2017), covering 37% of the investor population by 2021 (eToro, 2017). Most of these social trading platforms are regulated. For example, the Financial Conduct Authority (FCA) in 2015 stipulated a rule that all traders in social trading need to comply with the MiFID II ruling to qualify as investment managers.[2] A handful of these platforms are allowed to operate in the U.S. market to serve U.S. residents, including Collective2, Peertrade and Zulutrade's Forex and FXCM markets.[3]

Of central interest to us is the social trading feature called copy trading. This feature enables investors to replicate – in real-time and in their own accounts – the actual trades of other investors. An investor is allowed to follow other traders for a fee. There are several key advantages of such a copy trading mechanism, compared to other wealth management services: (1) more transparency since investors will observe every single trade of a trader,[4] (2) higher-level of control since the execution is done in the investor's own account,[5] (3) more reasonable fee (compared to funds management), typically zero management fee and a small percent when there is a gain,[6] and (4) reducing potential manipulation that may

[2]The key regulations on social trading include: (1) No hedging of trades - a trader cannot have a trade on the same instrument open in two opposite directions (Buy and Sell simultaneously); (2) First In First Out (FIFO) trades rule - when a trader has multiple trades open of the same pair in the same direction, they must however be closed in the order they are opened (https://financefeeds.com/mifid-ii-entering-age-completely-self-directed-traders-final-nail-goes-copy-trading-coffin/).

[3]https://socialtradingguru.com/social-trading-for-us-residents

[4]For example, although institutional wealth managers with $100 million or more in qualifying assets are required to file quarterly an SEC Form 13F detailing their investment holdings, investors do not know exactly when the fund manager bought and sold a specific security in the portfolio.

[5]In institutional markets, investors have less flexibility over their portfolios. For example, less control over timing the realization of gains and no way to customize the specific set of securities in the portfolio.

[6]The fund industry is overdue for a change (Kapadia, 2018). It is reported that investors earned an average of 4.67% on mutual funds over the last 20 years, which is 3.52% less than the average S&P 500 index return; mutual funds made an average return of 6.92% over the last 5 years, around 3% less than the S&P 500 index over the same period (Kim, 2017). As Warren Buffett described about fund managers: "Professionals in other fields, like dentists, bring a lot to the layman, but people get nothing for their money from professional money managers." The fee war is just part of the problem; price-conscious investors also want more transparency around how they pay for advice (https://www.zulutrade.com/trader-guide).

occur in the financial advising industry, where financial advisors may strategically distort their recommendations by "speaking in two tongues", for example, issuing overly positive recommendations but less optimistic forecasts (Malmendier and Shanthikumar, 2014). Under copy trading, it is hard, if not impossible, for traders to "speak in two tongues" since their real investment actions are observed and copied. Top social trading networks (platforms) that enable copy trading include eToro, ZuluTrade, Ayondo, Tradeo, etc. For example, eToro launched this feature back in 2007 for followers to replicate traders' trades (Kortekaas, 2013). According to eToro, among its 300 billion U.S. dollars worth of trades, two-thirds were executed through copy trading (Brand, 2017).

Among the differences between social trading and traditional investment services discussed above, improved information transparency has been one of the key drivers underlying the current wave of financial technology innovations (Lee and Shin, 2018). Compared with traditional fund management, social trading platforms provide much higher information transparency by sharing not only aggregated metrics but detailed trade-level information of traders. Many social trading platforms have gone to the extreme of publicizing all the trades of each trader (Glaser and Risius, 2018). This complete information transparency policy is one of the main drivers behind the explosive growth of the user base in social trading (Röder and Walter, 2017).

Information transparency in social trading involves two aspects: (1) *what* information should be disclosed (e.g., detailed versus aggregate trading performance), and (2) *when* information should be disclosed (e.g., immediate versus delayed). Timing the release of information is particularly relevant in social trading since trading information quickly loses value in a fast-paced financial market. A trend in the social trading industry is to release as much information to investors as possible (e.g. eToro and ZuluTrade have started releasing

each trade transacted by each trader including the securities bought or sold, time and price of the transaction etc.).[7]

We seek to devise an information revelation policy to optimize the right level of information transparency in social trading. Because, most social trading platforms have adopted the practice of releasing trade-level information as of 2019, our study focuses on the second dimension of information transparency, namely, the decision of *when* to release information. In copy trading, timing the release of information is crucial. Potential investors have to rely on the information the platforms provide to evaluate traders in order to choose a subset of traders to follow. Information becomes less transparent if it is released with longer delay. On the other hand, releasing trade information with very little or no delay can also backfire. Copy trading is not free for followers: the platform functions as an online broker and, in addition to a brokerage commission, it also charges followers a following fee to follow the trades of traders in real-time. Traders receive a commission calculated as a fraction of the follower fees (e.g., 25% in ZuluTrade). Followers will not pay to follow if detailed trade information is already released for free in real-time. Thus, traders will lose paid followers and consequently, will lose the incentive to participate in the platform. Therefore, a crucial decision for a copy trading platform is to choose the right level of information transparency with regard to the timing of releasing trade information.[8]

At the high level, we answer the simple but fundamental question: is copy trading worthwhile for followers? In other words, *how valuable is it for followers to know (in real-time)*

[7]Concerning the decision of what information to release, most social trading platforms nowadays release full trading information. By providing the detailed trade-level information, the platform protects itself by delegating the responsibility of evaluating traders to followers, thereby reducing the probability of followers making biased decisions when acting on aggregated performance indicators (e.g., monthly returns). However, because releasing detailed information could increase the information processing burden of investors, detailed information is usually made available on-demand (e.g., by clicking on a link), while summarized information is directly visible to users.

[8]A concomitant dual problem that can be considered is to optimally price the following fee for a given level of information transparency. In this chapter, we will focus on the problem where the following fee is fixed, but the information release policy can be controlled at the level of each trader.

the trades of traders, given that they must pay a fee to receive this information? This is a fundamental question all copy trading platforms need to answer before designing an optimal information revelation policy.

The answer to this question is vital for several reasons. First, the copy trading business model will collapse if followers do not see the real-time value in their service. Thus, the answer will help followers decide when it is valuable to follow in real-time and when it is not. Second, answering the above question is key for the platform to best monetize the dissemination of its financial advice. Third, it helps platforms determine the right magnitude of delay to add before releasing trade information to the public for free, while preventing arbitrage opportunities and at the same time guaranteeing a certain level of information transparency. For example, ZuluTrade releases traders' trades after 30 minutes to allow potential followers to evaluate different traders before following them. Lastly, the answer offers cues for the platform to personalize the delay (and, equivalently use a customized following fee) for each trader, based on the characteristics of the trader and the market conditions associated with the investment product.

To summarize, we specifically investigate the following three research questions.

- How does the level of information transparency (as measured by delay) affect the profit of a trade?
- How is the amount following a trader affected by information transparency?
- How should the platform design its information revelation policy?

Empirically, we examine the copy trading phenomenon implemented by ZuluTrade, a leading social trading platform that mainly operates in the Foreign Exchange market. It mainly deals with day trading, and thus the time to open and close a trade within a market is important. ZuluTrade provides two types of accounts: trader and follower. A follower can

"follow" one or more traders. Trades from trader accounts will be copied in real-time and automatically executed in a follower's own accounts. The platform functions as an online broker and charges followers a fee (per trade with some rate) to follow the trades of a trader. While traders themselves also pay a commission to the platform to execute their trades, they receive a kickback bonus (usually settled on a monthly basis) that is calculated based upon the amount of money that follows their trades. The platform operates as a two-sided market, where more followers will motivate more traders to join in and vice versa. Followers pay to copy real-time trades; however, these trades are available for free after some time of delay (for example, 30 minutes during our study period). Thus, investors can free ride if they are willing to tolerate the time delay introduced by the platform.

The data we obtained consists of individual traders' trades executed on ZuluTrade, for the Foreign Exchange (Forex, or currency trading) market. Our interest is to quantify the information value of knowing these trades in real-time. We define *profit-gap*, as the difference between the real-time profit of a trade and the simulated profit (calculated by comparing with historical Forex spot price data) of the trade, but executed with some open delay and/or some close delay. We use the data to examine how delay affects the profit-gap, after controlling for various factors related to the trader and other market conditions. Next, from the data, we present empirical evidence to demonstrate the platform's dilemma: lowering the delay increases transparency (that could potentially increases future revenue) but risks losing the amount of money following (and hence could hurt the platform's current revenue).

We then formulate and solve several optimization problems that address the platform's transparency-revenue conundrum. First, we examine the current approach adopted by ZuluTrade and study how it can be improved. Currently ZuluTrade seeks to maximize transparency while holding the Money-at-Risk at or below an acceptable level, where Money-at-Risk measures the vulnerability of the platform's current revenue as a function of the delay and the fees the platform charges followers to receive information on a trader's trades in

real-time. We study an improved information release policy that is customized at the trader level and demonstrate that it can substantially increase transparency (reduce delay) while maintaining the Money-at-Risk at a level equal to ZuluTrade's current release policy.

We further study – using a Stochastic Optimal Control formulation of the problem – how the information release policy can be designed to optimize revenue. That is, rather than indirectly addressing the problem by maximizing transparency subject to a given (exogenously chosen) Money-at-Risk, we set up a model to directly maximize platform revenue. The Stochastic Control problem is solved to yield a feedback control (i.e., delay) that is based on the current amount following a trader and the current attention being received by the trader from potential followers (measured by the trader's profile page views). The calculated revenue can be incorporated into the ranking algorithm to provide a systematic way to infuse the platform's goals into the ranking of the traders. In so doing, this study also helps followers to determine whether it is worthwhile to follow a specific trader in real-time and what is an appropriate following fee to pay a trader.

4.2 Literature Review

In this section, we briefly summarize the literature on social trading. The first stream of literature studies investment decisions made by individuals as a result of their social interactions. Ammann and Schaub (2016) investigate the role of social interaction in investment decisions by mining trader posts and other communications. They find that traders with superior performance are more likely to discuss their investment strategies. Heimer (2016) focuses on a phenomenon called the *disposition effect*. This anomaly, discovered in behavioral finance, uncovers the tendency of investors selling too early in the up market, while holding too long in a down market. Disposition effect is found to be magnified when investors receive advice from their friends. Some studies focus on the investment decisions made by institutional investors as a result of the social interaction (Pool et al., 2015; Jiang and Verardo, 2018).

Pool et al. (2015) find that socially connected fund managers have more similar holdings and trades; the overlap of funds whose managers reside in the same neighborhood is considerably higher than that of funds whose managers live in the same city but in different neighborhoods. Jiang and Verardo (2018) investigate the herding behavior and trading skill in the mutual fund and find that herding funds underperform their anti-herding peers by over 2% per year.

The second related stream of literature studies the phenomenon of copy trading. Doering et al. (2015) describe how copy trading platforms are organized and discuss the basic mechanics behind the relationship between signal provider (portfolio manager) and signal followers (investors). They find that signal providers typically engage in active trading rather than buy-and-hold strategies, which result in non-normal return distribution. Lee and Ma (2015) propose a system identifying traders with good and consistent performance to answer the question "whom to follow". Oehler et al. (2016) show that, on average, traders on wikifolios (a copy trading platform) do not outperform the market on average but the best performing traders earn significant short-term excess returns.

Within this literature on social trading, several studies analyzed the trading behavior and performance of traders in copy trading. Pan et al. (2012) examine the role of social mechanisms in a financial system and find that social trades outperform individual trades. Röder and Walter (2017) discover that traders who communicate actively with investors attract significantly more attention, and visibility of their trading portfolios boosts the volume of investments. Breitmayer et al. (2017) investigate the trading patterns of traders who received social recognition for their investment advice. They show that confirmatory social recognition leads to increased trading activity. Pelster and Hofmann (2017) study the relationship between providing financial advice and the disposition effect. They find that leading traders are more susceptible to the disposition effect than traders without followers.

The third stream of relevant literature is on information transparency with regard to what information and how much information to release in financial services. For example, in

crowdfunding, platforms need to decide what level of borrower information to release to help these lenders evaluate a loan. Crowdfunding studies find that what information (explicit or implicit) to release strongly influences the overall market efficiency and the lenders' decisions to participate (Herzenstein et al., 2011; Mäschle, 2012). Zhou et al. (2018) build a structural model to uncover how lenders' behaviors are affected by an exogenous information-disclosure policy change. They show that displaying extra information leads to a higher browsing propensity, which helps lenders to make sound investment decisions.

The fourth stream of related literature is on the value of information for devising trading strategies. Mutual funds are mandated to disclose their portfolio holdings to investors periodically, e.g. quarterly. Some investors might mimic the trading strategy from the released portfolio, called copycat funds in finance. Verbeek and Wang (2013) indicate that free-riding on disclosed fund holdings is an attractive strategy and suggest that mutual funds may suffer from such information disclosure regulations. The timeliness of portfolio holdings disclosure has been of interest among regulators, academics and practitioners since the Investment Company Act of 1940. The Securities Exchange Commission (SEC) has been trying to strike a balance (a uniform delay across all mutual funds) between investors' interest in timely disclosure and the potential costs associated with revealing the strategies of investment managers (Choi and Chhabria, 2012), where the information required to be released in mutual funds is the portfolio holdings, not the exact time when the mutual fund manager bought and sold the portfolio holdings. In this study, our data comes from a social trading platform focusing on day trading and the platform does releases the time traders execute a trade.

Different from the prior research that studies the social trading phenomenon mainly from a social or behavioral perspective, our study is more normative. We address a fundamental design problem in copy trading to determine the right level of transparency when releasing information to the public. Specifically, the main goal of this study is to measure the economic

value of paying for, and obtaining, real-time trade information in social trading, and to choose the optimal level of information transparency to maximize the economic goals of the platform. Prior research has not considered this important perspective.

4.3 Data and Platform Operations

In this section, we first introduce the data and then describe the operations of the social trading platform we consider.

4.3.1 Data

Our data comes from ZuluTrade, one of the largest copy trading platforms in the world. The platform allows followers to auto-copy Forex trades made by financial experts (traders). Each trader owns a public profile page, which reveals information on her past trading performance tracing back to the first day the trader joined the platform. ZuluTrade releases various performance metrics including the total profit of all the trades a trader has executed via the platform, the best and the worst realized profit of the trade among all her trades, the percentage of winning trades, the number of followers following the trader, the total amount of money following the trader, the number of weeks the trader has been trading on the platform, the number of views the trader's profile has received, etc. Importantly, ZuluTrade also releases on-going trades (trades have been opened but not closed yet) for free after adding some time delay to the public. Figure 4.1 provides a screen shot of a trader's profile page.

We obtained individual trading information of 15,352 traders during a 17 month period from August 2016 to December 2017. The data is at the most granular trade level possible for each trader. The detailed trade information includes the currency (e.g. EUR/USD), type (buy or sell), standard lot size, date open, date close, open price (the spot price at the time when the trade was opened), close price (the spot price at the time when the trade was

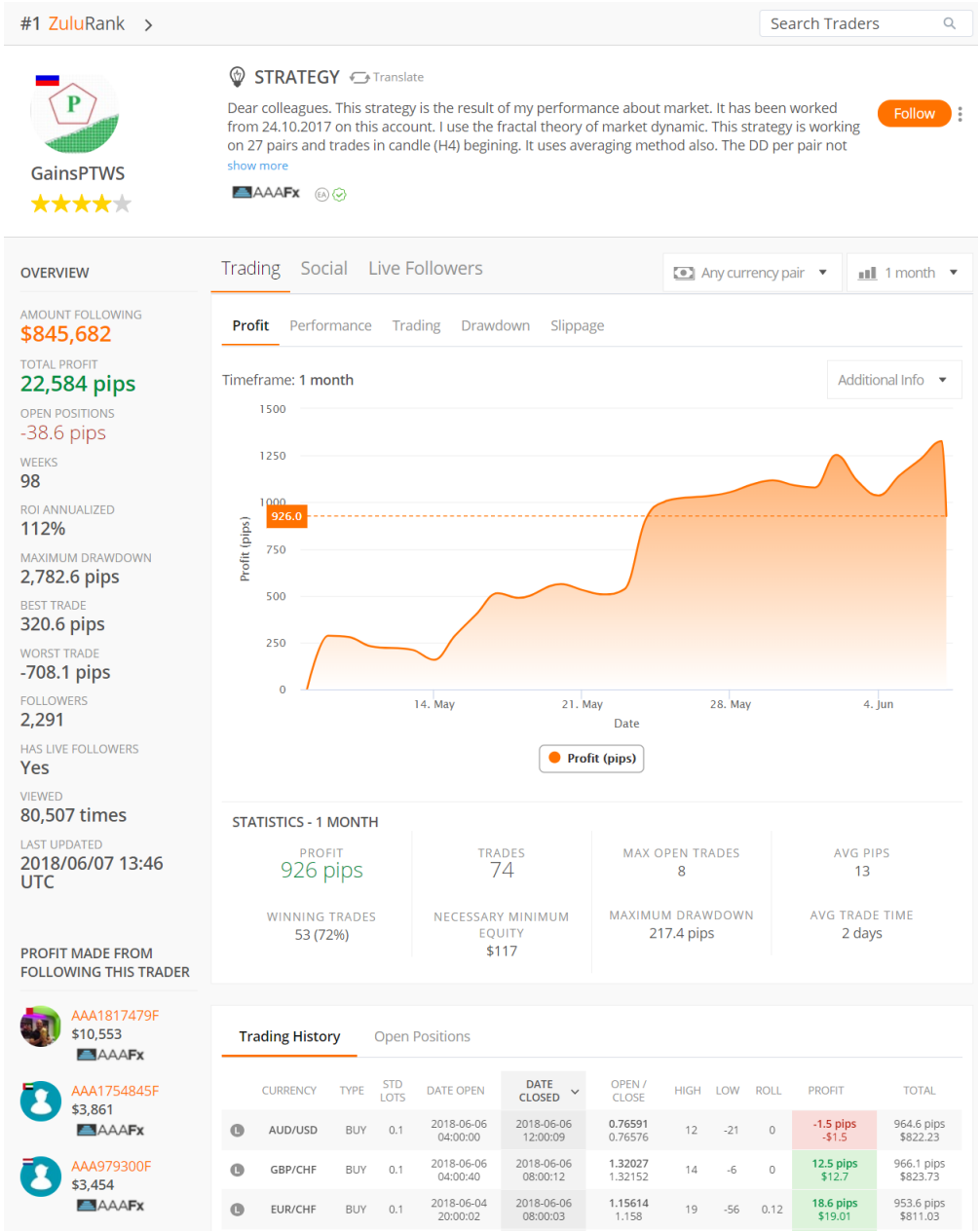

Figure 4.1. Snapshot of a Trader's Profile Page

Notes: The profile page describes the trader's investment strategy (top) and provides an overview of the trader's aggregate details, such as, trader performance (Annualized ROI, Maximum Drawdown,etc.) as well as social details such as Amount Following, Number of Views of the Profile Page, etc. (left column). Importantly, this page also provides the complete trading history (bottom). Here, trades are released for all to see with a 30-minute delay.

closed), profit, the highest potential profit during the holding period of the trade, the worst potential loss (maximum drawdown) during the holding period of the trade, etc. We focus on the intraday trades since the Forex market is volatile and more than half trades are open and closed within the same day. The Forex market provides an ideal market (environment) to study the time value of trade information since the profit of a trade is very sensitive to the magnitude of delay. In this study, we focus on the five largest currency markets – EUR/USD, GBP/USD, GBP/JPY, USD/JPY, and USD/CAD – that together account for 64% of all the trades on this platform. Table 4.1 presents the proportions of the five currencies among all the intraday trades on ZuluTrade.

Table 4.1. Proportions of Five Currencies on ZuluTrade

Currency	Frequency	Percent
EUR/USD	585,939	24.21%
GBP/USD	385,702	15.94%
GBP/JPY	226,230	9.35%
USD/JPY	224,461	9.28%
USD/CAD	126,421	5.22%
Total	1,548,753	64.00%

Notes: the total number of trades on ZuluTrade is 2,419,926 among the 15,352 traders during a 17 month period from August 2016 to December 2017.

The second data set we obtain is the historical spot prices for these five currency markets from the Dukascopy Historical Data Feed. We use this data set to calculate the hypothetical profit (loss) of a trade, upon which we will elaborate in the next section.

4.3.2 ZuluTrade Operations

On ZuluTrade, once a follower clicks to "follow" (copy) a trader, the trades of this trader are automatically copied and executed in this follower's account. The platform charges followers commission (following fee): 2 pips for a complete open-close trade per standard lot size.[9]

[9]In the Forex market, the standard lot size is $\$100,000$. A pip is the smallest price move that a given exchange allows. Major currency pairs are priced to four decimal places; the smallest change is that of the

This commission is already factored in the buy and sell price. Followers can choose how to *copy* traders' trades. For example, followers can choose a fraction $\gamma \in [0, 1]$ for each trader; only a γ fraction of the trader's trade amount will be executed in the copied trade. The copying trades will automatically be executed in real-time in the follower's account. Followers who do not follow a trader cannot receive this trader's trade information in real-time. However, the platform releases this trade information with some delay on the trader's profile page for the public to see, and for free.

The platform shares revenue with traders: 0.5 pip for each standard lot executed (open and then closed) in a follower account (that is \$5 for a trade of \$100,000). During our study period, the traders compensation is calculated on a monthly calendar basis, but a trader is compensated only if the profit earned in the trader's account for that month is positive. Traders either get fully compensated or zero for their trades in the month.

4.4 Empirical Investigations

In this section, we first describe how we measure the hypothetical profit (loss) of executing a trade after some time delay is added to the original, real-time trade. We then explore what factors influence the profit-gap, especially how magnitude of delay impacts the profit-gap. Finally, we study what drives the amount of money following a trader and empirically demonstrate the platform's dilemma concerning delay.

4.4.1 Simulating Delayed Trades

Let us assume that a trader opens a particular trade at time t_1 and closes this trade at time t_2; the difference $(t_2 - t_1)$, is the *holding time*. Such trade level data is directly observed in this study. Also observed is the profit or loss associated with the trade. Now, consider

last decimal point. For example, for the currency EUR/USD, 1 pip is equivalent to \$0.0001. Therefore, for a trade of size \$100,000, the platform charges followers \$20 as commission (following fee).

a hypothetical trade that is the same as the above trade (the same currency, type, and standard lot size) but the open time and close time are, respectively, $t_1 + \delta_1$, and $t_2 + \delta_2$. Here, δ_1 and δ_2 denote the open delay and close delay respectively. We can recover the profit (loss) of this delayed trade using the historical spot prices in the currency market traded at time $t_1 + \delta_1$, and $t_2 + \delta_2$. We vary the values for open and close delay at different levels: 0, 5, 15, 30, 60, 90, and 120 minutes.[10] When $\delta_1 = \delta_2 = 0$, it represents a real-time trade. The open and close delay are chosen such that the simulated close time is later than the simulated open time. Thus, we simulate a maximum of 48 distinct hypothetical trades corresponding to each original (or real-time) trade in the data sample. Figure 4.2 depicts how we generate the hypothetical trade after considering time delay.

For each simulated trade, we calculate the profit-gap as the real-time profit minus the simulated profit. The profit-gap directly measures the economic value of acting upon real-time information, relative to acting upon delayed information. The higher the profit-gap value, the more worthwhile for followers to follow the real-time trade (with follower fees). We would like to mention that the profit-gap value can be negative for some trades, meaning the delayed trades may even earn higher profit than the real-time trades. However, we believe that a good trader is able to time the market (when to enter and exit the market). Delaying the trade would then lower the value of her original trade.

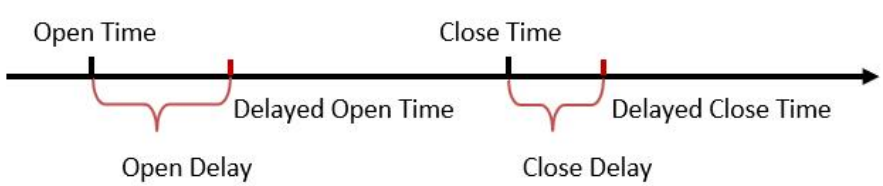

Figure 4.2. Simulating a Hypothetical Trade

[10] A delayed trade of more than 2 hours is hardly worth considering in the volatile Forex market.

4.4.2 What drives profit-gap?

We next investigate how different factors affect the value of following the trade in real-time. To this end, we specifically explore how the magnitude of delay influences profit-gap.

We use i to index a trader, j to index a trade, m to index a currency market, and t to index each 15-minute time period across the 17 months.[11] The dependent variable is $ProfitGap_{i,j,t}$. The explanatory variables of interest include $OpenDelay_{i,j,t}$, $CloseDelay_{i,j,t}$, and $Followers_{i,t}$. The variable $OpenDelay_{i,j,t}$ ($CloseDelay_{i,j,t}$) is the open (close) delay added for trade j, executed by trader i at time period t. We also include the quadratic term to capture nonlinearity. The variable $Followers_{i,t}$ represents the total number of followers following trader i at time t.[12] The econometric model we consider is specified as follows:

$$ProfitGap_{i,j,t} = \beta_1 OpenDelay_{i,j,t} + \beta_2 (OpenDelay_{i,j,t})^2 + \beta_3 CloseDelay_{i,j,t}$$
$$+\beta_4 (CloseDelay_{i,j,t})^2 \quad + \beta_5 Followers_{i,t} + X_{i,j,t} + Trader_i + Currency_m + Time_t + \varepsilon_{i,j,t} \tag{4.1}$$

In equation (4.1), $Trader_i$ is the trader-level fixed effects, $Currency_m$ is the currency (market) specific fixed effects [13], and $Time_t$ is the time-specific fixed effects capturing the market status corresponding to a specific time period. The term $X_{i,j,t}$ represents a vector of control variables, including the standard lot size of trade j ($StandardLots_{i,j,t}$), the highest potential profit (measured in pips) during the holding time period of trade j by trader i at time t ($HighestProfit_{i,j,t}$), the lowest potential profit (worst drawdown) during the holding time of trade j by trader i at time t ($WorstDrawdown_{i,j,t}$), the holding time of trade j by trader i at time t ($HoldingTime_{i,j,t}$), the total profit (with unit dollars) of trader i by time t ($Profit_{i,t}$), the total profit (with unit pips) of trader i by time t ($Profitpips_{i,t}$), the highest

[11] We use a relatively short time window of 15 minutes because of the high volatility of the Forex market.

[12] In the regression model, we take the logarithm of the followers to account for skewness in this variable.

[13] For notation simplification, we suppressed m from the subscription of all the other variables.

profit of a trade among all the trades of trader i until time t ($BestTrade_{i,t}$), the lowest profit of a trade among all the trades of trader i until time t ($WorstTrade_{i,t}$), the number of trades of trader i by time t ($Trades_{i,t}$), the annual return of investment of trader i by time t ($ROI_{i,t}$), the percentage of trades with positive profits among all the trades of trader i by time t ($WinRatio_{i,t}$), the number of weeks trader i has been traded on the platform by time t ($Age_{i,t}$), the rank of trader i at time t ($Rank_{i,t}$), the average profit of trader i by time t ($AvgTrade_{i,t}$), the maximum number of open trades trader i has been held by time t ($MaxOpenTrade_{i,t}$), the minimum capital (with unit dollars) required to trade all the trades from trader i in a follower's account at time t ($MinEquity_{i,t}$), the profit of open (unrealized) positions of trader i at time t ($OpenPosition_{i,t}$), and the number of views trader i's profile page has received by time t ($View_{i,t}$). Table 4.2 tabulates the summary statistics of profit-gap, open delay, close delay, etc.

To test potential existence of multicollinearity, we calculate the variance inflation factors (VIF) and find that they are well below the acceptable threshold (10), indicating the absence of multicollinearity. We use the White test to check for heteroscedasticity, and the result (the chi-square value is 1,467,927 with p-value less than 0.001) shows significant heteroscedasticity in the error term. We therefore control for it using robust standard errors.

We simulate hypothetical trades with different magnitudes of delay. Because the volumes of the hypothetical trades are minuscule relative to the gigantic Forex market, it is unlikely that any trade in our data would have any discernible market impact. For example, the daily trading volume of the entire Forex market is on average \$1.8 trillion in January 2018, while the annual trading volume in ZuluTrade is about \$800 billion in 2018. The maximum amount of money following a trader in our data is about \$3 million. Therefore, we do not expect any endogeneity in the variables $OpenDelay$ and $CloseDelay$. On the other hand, potential endogeneity might arise in variable $Followers_{i,t}$. For example, it can be argued that higher a profit-gap may likely cause followers to follow a trader in the first place, i.e. simultaneity may occur.

Table 4.2. Summary Statistics

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
$ProfitGap$	48,125,024	0.69	19.79	0.30	-1,002.20	1,031.90
$OpenDelay$	48,125,024	31.30	37.15	15	0	120
$CloseDelay$	48,125,024	42.22	42.89	30	0	120
$Followers$	48,125,024	9.22	59.14	0	0	2,589
$StandardLots$	48,125,024	3.75	86.63	0.06	0.01	100,000
$HighestProfit$	48,125,024	18.79	24.35	11	0	881
$WorstDrawdown$	48,125,024	-19.66	22.56	-12	-1,351	0
$HoldingTime$	48,125,024	190.74	250.64	95.35	1	1,440
$Profit$	48,125,024	144.95	21,221.78	1.35	-1,770,000	12,800,000
$Profitpips$	48,125,024	5.38	31.32	6.10	-655.60	772.40
$AvgTrade$	48,125,024	5.30	72.04	1.00	-10,611.30	4,898
$MaxOpenTrade$	48,125,024	45.94	49.99	30	0	428
$Rank$	48,125,024	9,826.72	12,034.6	2,598	1	32,001
$Trades$	48,125,024	1,888.94	2,353.35	911	0	14,676
Age	48,125,024	39.45	49.71	23	0	478
$Amount$	48,125,024	10,308.21	63,718.35	0	0	3,267,296
$BestTrade$	48,125,024	508.49	1,360.10	200	-499.60	69,938.40
$WorstTrade$	48,125,024	-785.72	8,291.01	-371	-1,807,046	0
$WinRatio$	48,125,024	65.45	22.57	68	0	100
ROI	48,125,024	392.28	3,261.81	32	-11,963	84,266.6
$MinEquity$	48,125,024	1,866.37	2,050.23	1,064	0	9,955.44
$OpenPosition$	48,125,024	-1,529.13	11,028.39	0.00	-1,486,899	622,768.30
$View$	48,125,024	8,912.46	30,339.39	1,834	0	2,258,260

Notes: The unit of $ProfitGap$ is pip; the units of $OpenDelay$, $CloseDelay$, and $HoldingTime$ are minute; $Followers$ is the number of followers; one unit of $StandardLots$ is $100,000 in the Forex market; the units of $HighestProfit$, $WorstDrawdown$, $BestTrade$, $WorstTrade$, $OpenPosition$, $AvgTrade$, and $Profitpips$ are pip; the units of $Profit$ and $MinEquity$ are U.S. dollar; the unit of Age is week; $View$ is the cumulative number of views by time t.

To account for such potential endogeneity, we use the two-stage least squares (2SLS) regression with instrument variables (IV). We construct two IVs for the variable $Followers_{i,t}$. The first IV (IV1) uses the amount following other traders from the same set of followers following the focal trader. Specifically, IV1 is a Hausman type of instrument, constructed as the total number of followers (from the set of followers following the focal trader) following other traders (but excluding the focal trader) at time t. The variable IV1 can be expected to be correlated with $Followers_{i,t}$ because it is constructed from the same set of followers; but it should not directly influence the focal trader's performance (the dependent variable $ProfitGap$) because the focal trader is excluded in its construction. Another IV (IV2) is the average rank of other traders who are followed by the focal trader's set of followers at time t. The rank of a trader is displayed by ZuluTrade. Similar to IV1, the average rank of other traders should be correlated with the number of followers because this is a metric followers should care about when deciding whether to follow a trader. However, IV2 should not directly influence the focal trader's performance because it is constructed using other traders, and not including the focal trader (Rossi, 2014).

To validate our IVs, we perform a Hausman test, where under the null hypothesis the specified endogenous regressor can actually be treated as exogenous. We have a chi-square value 1190.6 with p-value < 0.001, indicating the preference of IV based estimation. The IVs should further satisfy two prerequisites: the relevance assumption and the exogeneity assumption (Green, 2007). The former requires that the IVs should be correlated with the endogenous variables and that this correlation should not be weak. The F-statistics for the endogenous variable (followers) is 46,145 (p-value $= 0.000$). The latter condition, exogeneity, requires that instruments excluded from the structural equation must be uncorrelated with the structural errors, which is typically done using a test of over-identifying restrictions via Hansen's J statistic (Hansen, 1982). The Hansen J statistic is 2.642 with p-value $= 0.1041$, indicating that we should not reject the null hypothesis that the instruments are exogenous and excludable.

Table 4.3 presents the estimation results corresponding to the econometric model in equation (4.1). Larger magnitudes of open and close delay increase the profit-gap, indicating that it is more valuable for investors to follow in real-time if the platform releases a trader's trade with longer delay. From Table 4.3, we see that the coefficients of the squared delay terms are significant and negative. Thus, the profit-gap increases with delay in a concave manner. Holding everything else equal, the marginal effect of delay on the profit-gap decreases with a larger magnitude of delay. Equivalently, if the platform were to add more delay, the impact on profit-gap would decrease. Besides, the profit-gap increases with larger number of followers following a trader. The profit-gap is heterogeneous across different currency markets, and it varies as the level of market volatility varies in different markets.

Table 4.3. Impact of Delay on Profit-Gap

Dependent Variable: $ProfitGap$	
VARIABLES	Coefficients
$OpenDelay$	0.016***(0.000)
$(OpenDelay)^2$	-0.0009***(0.000)
$CloseDelay$	0.017***(0.000)
$(CloseDelay)^2$	-0.0007***(0.000)
$Followers$	0.154***(0.001)
Control Variables	Yes
Currency-level FE	Yes
Trader-level FE	Yes
Time-level FE	Yes
Observations	48,125,024
R-squared	0.128

Notes: The robust standard error is reported in parenthesis. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

4.4.3 Transparency-revenue Conundrum

To investigate what drives the amount following a trader, we estimate the regression model as specified in equation (4.2). The dependent variable is the amount following trade j from trader i at time t. The variables of interests are the number of (cumulative) views of trader i

received by time t ($View_{i,j,t}$), the profit-gap (with 30-minute open delay and zero close delay as implemented by ZuluTrade) of trade j from trader i at time t ($ProfitGap_{i,j,t}$), and the interaction term of view and profit-gap. The number of views captures the potential number of interested investors that the trader might be able to convert into followers. From the results with respect to equation (4.1), we know that larger magnitude of delay is associated with higher profit-gap. The profit-gap measures the information transparency level. $X_{i,j,t}$ represents the set of control variables as specified in equation (4.1). We take logarithms of amount and view. We also control for the trader-level and time-level fixed effects.

$$\begin{aligned} Amount_{i,j,t} = \theta_1 View_{i,j,t} + \theta_2 ProfitGap_{i,j,t} + \theta_3 View_{i,j,t} * ProfitGap_{i,j,t} + X_{i,j,t} \\ + Trader_i + Time_t + \varepsilon_{i,j,t} \end{aligned} \tag{4.2}$$

The variance inflation factors are all well below 10, indicating absence of multicollinearity. We control for heteroscedasticity using robust standard errors. Both variables $View$ and $ProfitGap$ might be subject to endogeneity. For example, it can be argued that a trader's profile attracts more views because the trader has higher amount of money following her; followers choose to follow a trader because it is worthwhile to follow her in real time (higher profit-gap), i.e. simultaneity may occur. Hence, we construct instrument variables for both variables. The IV we construct for $ProfitGap_{i,j,t}$ is the holding time of trade j. The profit-gap is sensitive to the holding time of the trade, but a follower's decision regarding whether to follow a trader or not does not directly rely on the holding time of the trade.

We construct two IVs for variable $View_{i,j,t}$. The first IV (IV3) is a Hausman type of instrument, constructed as the total number of views (from the set of followers following the focal trader) following other traders (but excluding the focal trader) at time t. Variable IV3 is expected to be correlated with $View_{i,j,t}$ because it is constructed from the same set of followers; but it should not be systematically co-determined with the focal trader's performance (the dependent variable $Amount_{i,j,t}$) because of the exclusion the focal trader. The other IV (IV4) is the average rank of other traders who are followed by the same set of

followers at time t. The rank of a trader is displayed by ZuluTrade; traders with higher ranks are listed at more prominent positions, attracting more views. Likewise, the average rank of the other traders should be correlated with the views because this is a metric followers care about when deciding whether to follow a trader; but IV4 should not directly influence the focal trader's views because it is constructed using the rank of the other traders (Rossi, 2014).

The Hausman test statistics is 864.6 with p-value < 0.001, indicating the preference of IV based estimation. The F-statistics for the three variables ($ProfitGap$, $View$, and the interaction term) are 8,144.60 (p-value < 0.01), 2,246.30 (p-value < 0.01), and 17,621.87 (p-value < 0.01) respectively, pointing to strong IVs. The Hansen's J statistic is $\chi^2_{(1)} = 1.647$ with p-value $= 0.199$, satisfying the exogeneity condition.

The results in Table 4.4 show that both coefficients of view and profit-gap are significant and positive. Interestingly, the interaction term of the two is negative, implying that view and profit-gap weaken each other. The marginal effect of profit-gap on the amount is $(0.289 - 0.031 \times View)$. When the number of views of a trader attracted is low, higher profit-gap is associated with higher amount following the trader. However, when the number of views is large, higher profit-gap results in lower amount following the trader. In other words, when a large number of potential followers are interested in a trader, low transparency (or high delay) makes it more difficult for non-following investors to evaluate the trader's ability and thus reduces the chance of converting them into followers. The marginal effect of view on the amount $(3.281 - 0.031 \times ProfitGap)$ can be negative and can be interpreted in a similar manner.

Taken together, the above negative interaction presents evidence that the platform faces a conundrum: the dilemma between increasing transparency (to facilitate the conversion of interested investors to followers) and reducing transparency (to prevent free-riding and protect existing revenue). Having demonstrated the transparency-revenue conundrum, we

Table 4.4. Impact of Profit-Gap and View on Amount Following a Trader

Dependent Variable: *Amount*	
VARIABLES	Coefficients
View	**3.281*** (0.055)**
ProfitGap	**0.289*** (0.095)**
View*ProfitGap	**-0.031*** (0.01)**
Rank	-0.003*** (0.000)
AgeLog	-4.055*** (0.081)
BestTrade	0.001** (0.000)
WorstTrade	-0.001*** (0.000)
WinRatio	0.052*** (0.002)
ROI	0.001*** (0.000)
Trades	-0.002*** (0.000)
AvgTrade	0.001*** (0.000)
MinEquity	-0.001*** (0.000)
StandardLots	0.002 (0.000)
HighestProfit	-0.005*** (0.001)
WorstDrawdown	-0.005*** (0.002)
Profit	0.001 (0.001)
MaxOpenTrade	-0.008*** (0.001)
Profitpips	0.003*** (0.001)
OpenPosition	0.001*** (0.000)
Trader-level FE	Yes
Time-level FE	Yes
Observations	730,812
R-squared	0.912

Notes: The standard error is reported in parenthesis. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. The variable $View$ is the number of cumulative views (in logarithm) by time t; the results are qualitatively the same when using the number of views during time t (noncumulative).

next devise various information release policies to optimize the level of transparency and platform revenue.

4.5 Optimizing Information Release

In this section, we develop and compare several information release policies that address the transparency-revenue conundrum in different ways. Central to the manner in which the tradeoff is handled, is the notion of Money-at-Risk, a concept we operationalize later in

this section. This concept was introduced to us during our interactions with the ZuluTrade management team. It is important to mention that the policies we study in this section (including the current release policy employed by ZuluTrade), do not explicitly optimize the revenue of the platform. Rather, these policies can be expected to indirectly impact the revenue in a desirable manner. Direct revenue optimization policies are introduced in the next section.

Conceptually, the platform would like to maximize the transparency of each trade in order for potential followers to evaluate traders, without endangering its revenue collected from commissions and follower fees. Thus, the platform needs to strike a fine balance between transparency and its revenue objectives. Transparency is maximized when a trader's trades are released to the public without any delay. However, such a (complete) transparency policy runs the risk that followers might manually copy the trades of a trader and execute them elsewhere without paying their due follower fees. This suggests the notion of *Money-at-Risk*, defined as the expected loss in the amount following a trader for a given delay and follower fee. To put it simply, Money-at-Risk (MaR) measures the risk of losing the amount following a trader. Mathematically, the Money-at-Risk for a trader is the amount following the trader (a_i) multiplied by the probability that the profit-gap of the trades associated with this trader is less than the follower fee (c) or,

$$a_i P_i\{x_i(\delta) \leq c\},$$

where $x_i(\delta)$ is the random variable representing the profit-gap of the trades associated with a trader. From the platform's perspective, an important goal is to limit the total Money-at-Risk across all the traders, namely, $\sum_{i=1}^{A} a_i P_i\{x_i(\delta) \leq c\}$, where A is the total number of traders.[14]

[14]More precisely, we should interpret A as the total number of trader *accounts*, because a trader is allowed to open multiple accounts (up to 10 accounts in ZuluTrade). For convenience, however, we will use the term trader to mean trader account. These two terms differ in practice, but inconsequential to our model.

We first introduce a theoretical model to capture how the profit-gap of a trade evolves as a function of the delay. This model will be proven useful in the analysis of the information release policies that are studied in this section.

4.5.1 Profit-Gap Model

The profit-gap evolves over time with some randomness, which comes from the fluctuation of market price. In modelling how the profit-gap evolves with time, the randomness needs to be considered. Therefore, we regard the profit-gap as a random process and use a Stochastic Differential Equation (SDE) to model it. Based on the empirical findings in Table 4.3, the mean profit-gap was observed to be a concave increasing function of the delay. This finding suggests the following stochastic process to model the profit-gap of a trader ($x_i(t)$) after introducing a delay of t from the point in time when the real-time trade was executed by trader i.

$$dx_i(t) = \frac{\alpha_i}{2\sqrt{t}}dt + \sigma_i dz(t),$$

where $dx_i(t)$ is the change of profit-gap in a small time interval from t to $t + dt$, $dz(t)$ is the increment of the Wiener process (following $\mathcal{N}(0, dt)$) capturing the randomness, the parameters α_i and σ_i are trader specific.

The above model (VABM) is a *variant* of an Arithmetic Brownian Motion (ABM), where the drift term is inversely proportional to the square root of time (t). The above SDE is consistent with the empirical finding in equation (4.1), namely, that the profit-gap is an increasing concave function of the delay. If we take the expectation of $dx_i(t)$, we get $\mathbb{E}(dx_i(t)) = \frac{\alpha_i}{2\sqrt{t}}dt$. Integrating both sides with respect to time yields $\mathbb{E}(x(t)) = \alpha_i\sqrt{t}$, implying that the expected profit gap, $\mathbb{E}(x(t))$, is a concave and increasing function of time. The above SDE model can be analytically solved to show that the density $f(x_i(t))$ is Normal with a mean of $\alpha_i\sqrt{t}$ and a variance of $\sigma_i^2 t$. The mean $\alpha_i\sqrt{t}$, reflects the concave, increasing relationship with delay. The parameter α_i can be interpreted as the information value

associated with the trades of trader i: the higher this parameter, the higher the expected profit gap for the same level of delay. As we will see later, the value of α_i plays an important role in the information release policy: a trader with a higher value of this parameter will allow the platform to provide greater information transparency.

We apply Maximum Likelihood Estimation (MLE) to estimate the parameters for the profit-gap of a trader ($\hat{\alpha}_i$ and $\hat{\sigma}_i$ for trader i). The VABM model fits the data better than two common alternative models: (1) a model with no drift term, or Brownian Motion (BM), and (2) a model with a linear drift term, or Arithmetic Brownian Motion (ABM). The average Bayesian Information Criterion (BIC) values for a random sample of of 1,000 traders are: VABM (85,200), ABM (86,600) and BM (87,600), showing that VABM has the best fit among the three models. This finding shows that, generally speaking, the profit-gap of a trade evolves with delay with a positive and concave drift, not just as white noise or with linear drift. However, there are indeed some traders with negative drift and some others with zero drift.

For any given trader and delay δ, we can use the above VABM process to determine the probability that the profit-gap associated with the trades of the trader is less than the following fee c. This probability is calculated as the cumulative distribution function of the normal distribution with mean $\hat{\alpha}_i\sqrt{\delta_i}$ and variance $\hat{\sigma}_i^2\delta_i$. $F_i(c|\delta_i)$ represents the probability that the profit-gap is less than c given the delay δ_i for trader i. Figure 4.3 illustrates how the probability changes with delay for a representative trader with id 102885. We see that the fitted probability $\hat{F}_i(c|\delta_i)$ decreases with δ_i in a convex manner. This convex property is seen to hold for most traders.

4.5.2 Comparison of Different Release Policies

We first propose an indifference information release policy that maximizes trade transparency while ensuring that the release of trade information is such that the expected profit-gap for

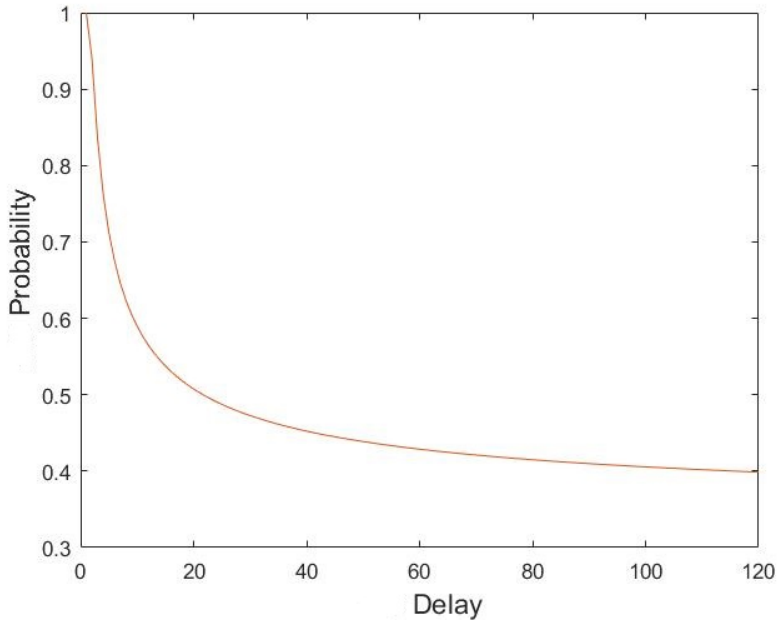

Figure 4.3. Fitted Probability versus Delay for Trader ID 102885

a trader is exactly equal to the following fee. This policy is based on the premise that, on average, there will be no loss of commission revenue because the release of trade information is such that a *risk neutral* follower will be indifferent between following in real time (with a fee) and executing the same trade with delay (for free). In this sense, it is an arbitrage-free policy. Next, we propose two policies that consider the profit-gap distribution to decide on the delay, rather than basing this decision on the mean of this distribution. The first is a *uniform* release policy where all trades are publicly posted for everyone to see with a delay of δ. The second policy is a *customized money-at-risk* release policy where the information release policy (i.e., delay, δ_i) depends on the characteristics of trader i. The uniform policy is, of course, a special case of the customized money-at-risk release policy.

Customized Indifference Policy

In this policy, a trader's trades are released with a delay such that the expected profit-gap equals the difference in commission fees between real-time and delayed trades.

Let $F_i(x|\delta_i)$ ($f_i(x|\delta_i)$) denote the cumulative probability function (probability density function) for the profit-gap (x) associated with trader i when the trades of this trader are released with a delay of δ_i. Because the profit-gap could be negative, we allow the support of the profit-gap distribution to be $x \in$ [-∞, +∞]. Then, the customized indifference delay would be given by the value of δ_i such that

$$\int_{-\infty}^{+\infty} x f_i(x) dx = c,$$

where c is the fee paid to follow the trades of a trader.

Customized Money-at-Risk Policy

In this policy, each trader (or more generally, each trader's *account*) can be customized with a different release delay, δ_i. The optimization problem can be stated as below.

$$\begin{aligned} \underset{\delta_i, i \in \{1,2,\ldots,A\}}{\text{Minimize}} \quad & \sum_{i=1}^{A} v_i \delta_i \\ \text{subject to} \quad & \sum_{i=1}^{A} a_i F_i(c|\delta_i) \leq \eta \end{aligned}$$

In the above specification, δ_i is the decision variable, A is the total number of traders, a_i is the amount of money following trader i, v_i is the number of views received by trader i in a period, $F_i(c|\delta_i)$ is the probability that the profit-gap is less than the following fee c given delay δ_i, and η is maximum money-at-risk (the platform can tolerate). Intuitively, the platform should release a trader's trades with less delay if the trader receives more views, to help the trader convert more potential followers.

In the above problem formulation, the objective function is a measure of trade *opacity*; thus we wish to minimize this measure. Our data tracks the number of potential followers viewing a trader. We denote the total number of page views per period as (v_i). These page views are a direct measure of potential followers evaluating the trader. The constraint in the

above problem represents the Money-at-Risk (MaR). When the probability that the profit-gap for a trader is below the follower fee ($F_i(c|\delta_i)$), investors who follow the trader could switch from following in real-time to free-riding with some delay (δ_i). This probability is a measure of the vulnerability of losing the follower fees the platform receives from the followers of a particular trader. The Money-at-Risk captures this vulnerability, and is calculated as the probability $F_i(c|\delta_i)$ multiplied by the amount following.

The probability $F_i(c|\delta_i)$ decreases with δ_i. At one extreme, if $\delta_i = 0$, the real-time profit and the delayed profit are the same; hence, the profit-gap is zero. Thus, the probability that the profit-gap is less than c is 1. As the delay increases, the delayed profit reduces (while the real time profit stays the same); hence the profit-gap increases. Thus, $F_i^{'}(c|\delta_i) < 0$. Based on our regression results (see Table 4.3), we would expect the profit-gap to exhibit diminishing returns with respect to delay, implying that the profit-gap can be expected to be a concave function of the delay. Thus, we can expect that the cumulative density F to be convex with respect to the delay, i.e., $F_i^{''}(c|\delta_i) > 0$, a property that is also supported in data.

The function $F_i(c|\delta_i)$ is trader specific and embeds in its knowledge of how proficient the trader is at spotting short-lived opportunities in the market. $F_i(c|\delta_i)$ can also be considered as a measure of the risk of losing followers for a given follower fee (c) and delay(δ_i). For the same delay (say δ_1), if $F_b(c|\delta_1) < F_a(c|\delta_1)$, it implies that trader b's trades carry more information value, or equivalently, for the same delay, these trades are more worthwhile to follow in real time, as illustrated in Figure 4.4. Thus, they can be released earlier to serve the goal of increasing transparency. However, relative to trader b, trader a's trades need more protection (from the perspective of information value), and need to be released later. In Figure 4.4, to limit the same risk of losing followers (p), the trades for trader b can be released with greater transparency ($\delta_1 < \delta_2$).

To solve the above problem, it is sufficient to note that it is a convex optimization problem. This is because the objective function is linear in the decision variables while the

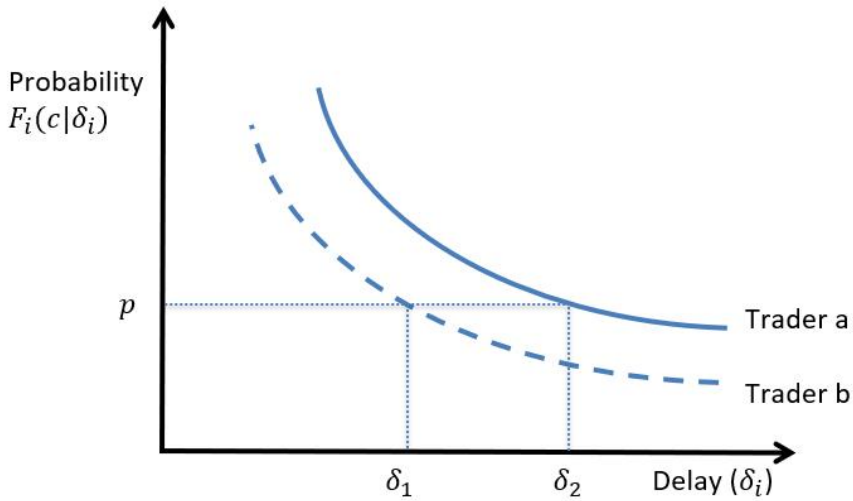

Figure 4.4. Risk of Losing Followers as a Function of Delay: Comparison of Two Traders

constraint is jointly convex in the decision variables. In the constraint, it is easy to see that the decision variables do not interact with one another and the data supports the property that the cumulative density associated with trader i, $F_i(c|\delta_i)$, is convex in δ_i.

Uniform Policy

In this policy, a uniform delay is decided for all traders. The optimization problem is stated as below.

$$\begin{aligned} \underset{\delta}{\text{Minimize}} \quad & \sum_{i=1}^{A} v_i \delta \\ \text{subject to} \quad & \sum_{i=1}^{A} a_i F_i(c|\delta) \leq \eta \end{aligned}$$

As before, the objective function is a measure of trade *opacity*; thus we wish to minimize this measure. The constraint is also conceptually similar. The uniform policy is a special case of the customized money-at-risk policy where all trader's trades are released with the same time delay (δ).

Numerical Illustration

We randomly choose ten traders to illustrate how the different release policies work. The amount of money following each trader (a_i) is $[2800, 400, 126, 773, 150, 58, 324, 650, 58, 200]$. The average views per period (day) for each trader (v_i) is $[1, 1, 1, 2, 3, 3, 3, 3, 5, 8]$. We take the following fee as $c = 1$ (with unit pips). For each trader, we first estimate the parameters α_i and σ_i using the data and obtain the cumulative density function $F_i(c|\delta_i)$.

Given different values of the maximum money-at-risk (η), we can calculate the optimal opacity and draw the Pareto curve associated with the uniform policy as shown in Figure 4.5 (a). The uniform 30-minute delay policy (marked as point U_1) adopted by the Zulu-Trade platform corresponds to a money-at-risk of 1,694.8 and an opacity measure of 900. Given the same money-at-risk value, the optimal opacity is only 460.9 under the customized money-at-risk release policy (marked as point C_1). The optimal delay for each trader is $[91.2, 31.7, 24.5, 32.9, 3.3, 4.6, 28.7, 19.1, 3.3, 8]$. As can be seen from the data, the first three traders have the same number of views ($v_1 = v_2 = v_3 = 1$), but the amounts of money following these traders are different. It is better to use a longer delay for a trader with a higher amount of money following. This is done to protect the revenue earned from the current followers of such traders. We can see that $a_1 > a_2 > a_3$ results in $\delta_1^* > \delta_2^* > \delta_3^*$. Conversely, everything else held constant, for a trader with a higher number of views, it is better to use a shorter delay.

In Figure 4.5 (a), point I_1 represents the customized indifference policy with a money-at-risk value of 2,769.5 and an opacity value of 380. The delay added for each trader's trades is $[0.4, 26.6, 1.9, 47.4, 2.1, 4.4, 11.3, 0.4, 1.6, 24.2]$. Given the same level of money-at-risk, the optimal opacity is 28.8 under the uniform policy with 1 minute delay (point U_2), and 16.3 under the customized money-at-risk policy (point C_2). The optimal delay for each trader is $[4.5, 1.8, 0.5, 1.3, 0.2, 0, 1.1, 1, 0, 0]$.

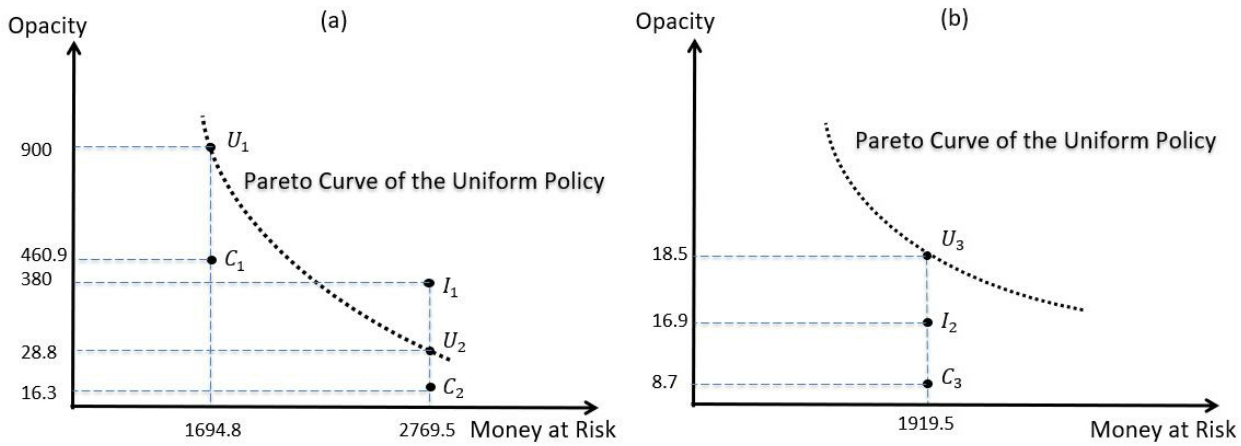

Figure 4.5. Comparison of Different Release Policies
Notes: U (Uniform Release Policy); I (Indifference Release Policy); C (Customized Release Policy). In Figure 5 (a) we show that the Customized Release Policy outperforms both other policies. Observe that for the same money at risk, C_1 and C_2 yield lower opacity than the other two policies. However, the Indifference policy and the Uniform Policy cannot be strictly ordered: In 5(a) the Uniform Policy has lower opacity, whereas in 5(b) the Indifference Policy does better.

In Figure 4.5 (a), we can see that, specific to these sample traders, the uniform release policy performs better than the customized indifference policy in terms of opacity given the same level of maximum money-at-risk. However, these two policies do not have a strict ordering in terms of their opacity values for the same money-at-risk. For example, as illustrated in Figure 4.5 (b), for a different set of traders, the customized indifference policy (point I_2) can perform better than the uniform policy (point U_3). Point C_3 represents the customized money-at-risk policy.

It is not surprising that the customized money-at-risk release policy always dominates the uniform release policy, since the uniform policy is a special case of customized money-at-risk policy. However, the customized money-at-risk policy can significantly outperform the uniform policy given the same money-at-risk.

4.6 Revenue Optimization

So far, we have considered information release policies that attempted to strike a balance between information transparency (represented as views-weighted delay) and revenue (represented as Money-at-Risk). Our optimization formulations for these policies, however, do not directly consider platform revenue as the objective to optimize. We next present a stochastic control model to directly optimize platform revenue. This model is developed at the trader level, i.e., for every trader, depending on the trader's parameters, we implement the feedback control $\delta(a(t), v(t))$, representing the delay introduced for a particular trader when the current amount following is $a(t)$ and the number of cumulative views is $v(t)$.

To validate the objective function, we first conduct a model-free analysis to check the correlation between the platform's revenue collected from following fees of a trader[15] and the trader's rank provided by the platform. We find that the revenue and rank are significantly correlated (higher revenue is correlated with higher rank) with an average magnitude -0.08 (smaller value representing higher rank), validating the assertion that the revenue collected from following fees is a reasonable objective.

The objective of the platform would be to maximize its revenue – from following fees – that is generated from the amount following a trader, $a(t)$. The revenue per unit time (or the revenue *rate*) earned by the platform can be expected to increase with the amount. Thus, we model the revenue rate as k times $a(t)$, where k is a trader specific constant that could be chosen based upon the frequency of trades executed by the trader. As mentioned above, there are two state variables that drive the feedback control, $a(t)$ and $v(t)$. We accordingly propose two stochastic differential models for the evolution over time of both state variables.

[15]The revenue collected from a trader is calculated by $a \times r \times 75\% \times n$, where a is the total amount of money following the trader, r is the following fee, which is 5 dollars per 100,000 dollars of trade, of the following fees the platform shares 25% to the trader and thus the platform keeps 75% share, and n is the number of trades of the trader during the time period.

The amount following the trader is modeled using a stochastic differential equation ($da(t)$, the change of the amount in a small time interval from time t to $t + dt$) whose drift term is a function of the control $\delta(a(t), v(t))$, the amount $a(t)$, and the views $v(t)$, as $g(\delta(a(t), v(t)), a(t), v(t))$. This is consistent with our understanding of the forces that drive the amount following a trader. For any given trader, we expect that the number of views generates an inflow into the set of current followers. However, this inflow is throttled if the opacity increases, i.e., the recent investment activities of the trader are not visible to potential followers. On the other hand, increasing transparency runs the risk of losing current followers. The noise term is modeled as $\sigma_a(a)dZ(t)$, where $\sigma_a(a)$ is the volatility coefficient and $dZ(t)$ is the increment of the Wiener process following $\mathcal{N}(0,\, dt)$. Following convention in control theory, we suppress the time argument and write the state equation for the amount as below.

$$da(t) = g(\delta, a, v)dt + \sigma_a(a)dZ(t)$$

Since ZuluTade uses a one-size-fits-all uniform release policy ($\delta = 30$ minutes), the impact of the control on the change in the amount following a trader ($da(t)$) cannot be empirically observed. However, concerning the impact of delay, we expect that the inflow of new followers to be a function of the delay whereas the outflow of current followers to depend on the square root of delay. The influence of the delay on the inflow is based on the assumption that new followers expect more transparency. On the other hand, the outflow should depend on the *square root* of delay, rather than directly on the delay. This is because the square root of delay is proportional to the profit-gap (as we show in Section 4.1), a quantity that ultimately controls a follower's decision to pay the follower fee and execute the real-time trade, or execute the delayed trade released by the platform for free.

We therefore propose that the drift term (or the mean change in the amount following a trader) to be given by

$$g(\delta, a, v) = \frac{-\alpha a}{\sqrt{\delta_0 + \delta}} + \frac{\gamma v}{\delta_0 + \delta}.$$

We expect $\alpha, \gamma > 0$. The parameter δ_0 represents the minimum possible delay. The first term in the expression for $g(\delta, a, v)$ represents the *outflow* in the amount; increasing δ reduces the outflow since it protects the current amount of money following the trader. We take the square root of delay because profit gap, rather than delay, is expected to influence the outflow. The outflow should increase with the amount. The second term in the drift is the inflow in the amount following a trader. This term captures the effect that a higher delay makes the trades less transparent; this throttling effect of delay is aggravated when the number of views is high.

The number of views received by the trader is modeled as another stochastic differential equation ($dv(t)$, the change of the number of views in a small time interval from time t to $t + dt$) whose drift term is a function of the amount following and the current number of views, $h(a, v)$. The noise term is modeled as $\sigma_v(v)dW(t)$, where $\sigma_v(v)$ is the volatility coefficient and $dW(t)$ is the increment of the Wiener process. Thus we have,

$$dv(t) = h(a, v)dt + \sigma_v(v)dW(t).$$

Conspicuously absent in the above relationship is the impact of the control variable (delay). However, this is expected. While the ability of potential followers to evaluate a trader depends on the delay, we do not expect the control to affect the evaluation *interest* (or views) from potential followers. Because we do not expect the control (delay) to affect the change in views in a direct manner, we look for empirical support for the function $h(a, v)$. Turning to the data, it is natural to expect that a higher amount following generates more interest. In Table 4.5, we find empirical support for a relatively simple form for the *change* in the number of views: $\Delta v(t) = v(t + 1) - v(t)$, by running the regression in equation (4.3) considering the control variables $X_{i,t}$ as well as trader-level and time-level fixed effects. Therefore, empirical evidence supports the proposed form $h(a, v) = pa + qv$.

$$\Delta View_{i,t} = \theta_1 Amount_{i,t} + \theta_2 View_{i,t} + X_{i,t} + Trader_i + Time_t + \epsilon_{i,t} \tag{4.3}$$

Table 4.5. Impact of Amount and Views on the Change in Views

Dependent Variable: $\Delta View$	
VARIABLES	Coefficients
Amount	0.003***(0.0003)
View	0.005***(0.0006)
Control Variables	Yes
Trader-level FE	Yes
Time-level FE	Yes
Observations	181,666
R-squared	0.020

Notes: The standard error is reported in parenthesis. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

4.6.1 Stochastic Control Model

Based on the above discussion, we present the revenue optimization problem for the platform as the following stochastic control model, after suppressing all implicit time arguments.

$$\begin{aligned} \max_{\delta} \quad & \mathbb{E}\left[\int_0^T kae^{-\rho t}dt\right] \\ \text{subject to} \quad & da = \left(\frac{-\alpha a}{\sqrt{\delta_0+\delta}} + \frac{\gamma v}{\delta_0+\delta}\right)dt + \sigma_a(a)dZ \\ & dv = (pa+qv)dt + \sigma_v(v)dW \end{aligned} \tag{4.4}$$

where δ is the control parameter, a is the amount of money following a trader at time t, v is the number of views at time t, and ρ is the discount factor.

Solution

To solve the stochastic optimal control problem, the optimal control δ should satisfy the following Hamilton-Jacobi-Bellman (HJB) equation for the value function V.

$$\begin{aligned} \rho V(a,v,t) = \max_{\delta}\{ka + V_a\left(\frac{-\alpha a}{\sqrt{\delta_0+\delta}} + \frac{\gamma v}{\delta_0+\delta}\right) + V_v(pa+qv) + V_t + \frac{1}{2}V_{aa}\sigma_a^2(a) \\ + \frac{1}{2}V_{vv}\sigma_v^2(v)\} \end{aligned} \tag{4.5}$$

Using the above equation, a point-wise optimization problem can be solved as below.

The optimal solution is *bang-bang*, implying that the trades of a trader should either be released with minimum possible delay (δ_0) or the highest delay that is reasonable. In this context, we can set δ_0 to be a small value (say a few seconds) to represent the minimum possible delay that can be implemented given information technology constraints. Concerning the maximum delay, we can set this value to a relatively large value, e.g., 120 minutes. A transaction released after too long delay (e.g., beyond a few hours) would likely be independent of the original real-time transaction in the volatile day-trading Forex market (on average each trader conducts four trades per day) and increasing the delay further would render the trade information close to zero value. The above values of minimum possible delay and maximum delay are illustrative, and could change with the platform's technology and the market. Note that, the optimal delay changes dynamically in the life of a trader, depending on the current values of the state variables associated with the trader (amount and number of views).

To prove the optimal policy structure, let $g(\delta)$ denote the terms involving the control variable in the HJB as shown below.

$$g(\delta) = \frac{-\alpha a}{\sqrt{\delta_0 + \delta}} + \frac{\gamma v}{\delta_0 + \delta}$$

The first derivative yields

$$g'(\delta) = \frac{1}{2(\delta_0 + \delta)^{\frac{3}{2}}} \left(\alpha a - \frac{2\gamma v}{\sqrt{\delta_0 + \delta}} \right)$$

The second derivative yields

$$g''(\delta) = \frac{1}{(\delta_0 + \delta)^{\frac{5}{2}}} \left(-\frac{3}{4}\alpha a + \frac{2\gamma v}{\sqrt{\delta_0 + \delta}} \right)$$

The optimal solution is a corner solution as shown below.

- If $g'(0) \geq 0$, then $g(\delta)$ increases in δ. Hence, the maximum delay should be chosen. Thus, $\delta^\star = \delta_m$, where $(\delta_0 + \delta_m)$ represents the maximum delay.

- When $g'(0) < 0$, then the optimal solution is either $\delta^\star = 0$ or $\delta^\star = \delta_m$. This can be seen by setting $g'(\delta) = 0$. Solving for $\delta = \delta_1$ yields a minimum because $g''(\delta_1) > 0$. The value of δ_1 is unique because once $g'(\delta) > 0$, it stays positive. Hence, the optimal solution can be found by evaluating the value of the function $g(\delta)$ at the two extreme points $(0, \delta_m)$ and choosing the higher of the two values.

Theorem 4. *The optimal information release policy can be completely characterized using the amount to views ratio ($\frac{a}{v}$).*

The delay should be set to the maximum value ($\delta^\star = \delta_m$) if

$$\frac{a}{v} \geq \left(\frac{\gamma}{\alpha}\right) \left(\frac{\frac{1}{\delta_0} - \frac{1}{\delta_0+\delta_m}}{\frac{1}{\sqrt{\delta_0}} - \frac{1}{\sqrt{\delta_0+\delta_m}}} \right),$$

$$\delta^\star = 0 \quad \textit{otherwise}.$$

The proof is provided in Appendix C.

A bang-bang policy is quite reasonable in the copy trading context. It is unlikely that followers will react to a small change in the information release policy. However, they will likely be sensitive to changes from maximum transparency to complete opacity. The optimal information release policy is sensitive to the ratio $\frac{a}{v}$ (or the amount to views ratio). When a trader's amount to views ratio is low, it means that there is potential to increase the follower base (views are relatively high relative to the amount). Hence, the optimal policy is to use minimum delay in order to convert more viewers into followers. However, the optimal policy switches to maximum delay once the amount to views ratio increases beyond a certain level. That is, in the life of a trader, the information release policy is not fixed and can switch from one extreme to the other.

Another point important to mention is that the value of the maximum delay (δ_m) is trader specific. That is, assuming the following fee is the same across traders, for a trader with more information value, we can afford to release the trades of this trader with less delay. The

value of the maximum delay can be chosen such that it is still a reasonable period of time (e.g., a few hours, to signal that the platform would like to be as transparent as possible) but it protects the amount following the trader. For example, we can use a global maximum for the maximum delay (say Δ_m) and set a trader's maximum delay to be the lower of Δ_m and a value such that $\mathbb{E}(da)$ is zero.[16] Mathematically, we can set

$$\delta_m^i = \min\left[\Delta_m, \left(\frac{\gamma v}{\alpha a}\right)^2 - \delta_0\right]$$

While the above bang-bang policy does not easily lend itself to an analytical solution for the value function, a numerical value of the optimal discounted revenue for a trader can be obtained for given values of a and v and all the other trader specific parameters. Currently, ZuluTrade provides a ranking of traders, whose ranking algorithm is not made public. However, this ranking can be assumed to be based on trader characteristics such as profit, risk, and so on, that signal the quality of the trader to potential followers. It could also be based on factors that are in the interest of the platform, such as the amount following. The numerical value of the optimal revenue can be incorporated into the current ranking algorithm to provide a systematic way to infuse the economic goals of the platform into ranking traders.

To illustrate the economic impact of the optimal policy, we compare, for a hypothetical average trader (all parameters are held at their mean values), the current policy (30 minute delay) and the optimal (bang-bang) policy. Figure 4.6 shows how the amount following evolves over time under the 30-minute policy and the optimal policy. We see that the optimal policy outperforms the 30-minute policy in terms of the amount of money following. In Figure 4.6 (a) the initial value of amount is high, the optimal delay is $\delta_m = 120$ minutes at $t = 0$ to protect the amount at risk.

[16] In our data, 75% of the trades are held less than 5 hours; the average holding time is 190 minutes. Thus, a reasonable value of Δ_m would be about 5 hours.

However, in Figure 4.6 (b) where the initial value of amount is low, the optimal policy is to set the delay to 0 at the beginning of the time to encourage the amount to grow. We also compare the value function (objective function in equation (4.4) using numerical integration. The value function is 1.23 under the 30-minute policy and 7.86 under the optimal policy in Figure 4.6 (a). The value function is 281.8 under the 30-minute policy and 330.0 under the optimal policy in Figure 4.6 (b). As mentioned earlier, evaluating the value function under the optimal policy (e.g. 330.0) is a useful factor to consider in the ranking of traders.

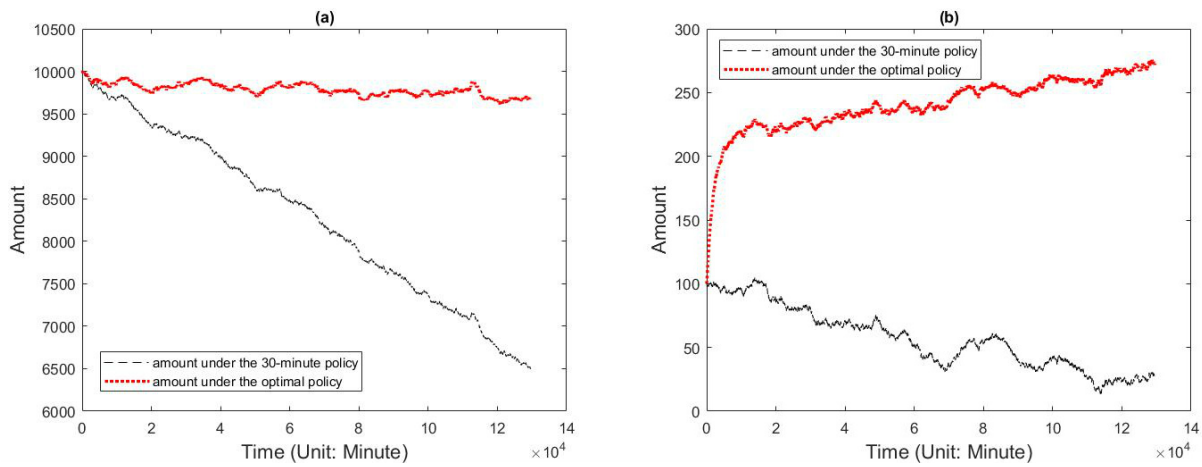

Figure 4.6. Comparison of the 30-minute Policy and the Optimal Policy
Notes: $\rho = 0.000000093$, $k = 2.6e{-}7$, $T = 129600$, $dt = 1$, $\delta_0 = 0.1$. Figure (a): $a(0) = 10000$, $v(0) = 50$, $\alpha = 1.6268e - 5$, $\gamma = 0.001$, $p = 0.0000001$, $q = 0.0000001$, $\sigma_a = 1$, $\sigma_v = 0.01$. Figure (b): $a(0) = 100$, $v(0) = 20$, $\alpha = 0.00015$, $\gamma = 0.0005$, $p = 0.000015$, $q = 0.00001$, $\sigma_a = 0.1$, $\sigma_v = 0.01$. The trajectories are the average of 100 iterations.

4.7 Conclusion

We studied information release strategies in social trading using data from ZuluTrade that primarily engages in the Foreign Exchange (currency trading) market. The data tracks 15,352 (primarily) day traders, copied by 15,492 followers across five major currencies. Our main interest was to determine the optimal information release policy for the platform. Towards this goal, we estimate the hypothetical profit earned by a trade if it was executed

with some delay (i.e., a virtual, *simulated* trade constructed by adding some time delays to the open and close time of the original trade). To estimate the profit earned by a simulated trade, we used historical spot price data of the currency being traded. Having estimated the profit earned from the simulated trade, we developed a measure called *profit-gap*, defined as the real-time profit minus the simulated profit. We estimated various econometric models to unravel the drivers of the profit-gap, the amount of money following a trader, and the number of views received by the trader. Using the empirical evidence gathered from the ZuluTrade platform, the study devised different ways for the platform to address the transparency-revenue dilemma, both implicitly (by maximizing transparency while respecting a Money-at-Risk constraint) and explicitly (by directly optimizing the revenue generated from commissions and follower fees).

The key contribution of our proposed approach was in its ability to address the conundrum, customized at the level of an individual trader. This is in sharp contrast to the current approach of the platform that adopts a simple, one-size-fits-all policy that chooses the same level of delay for all trades occurring on the platform. Like the platform, our focus was on the use of delay as a control variable, while holding the follower fee at a fixed level. A natural extension would be to consider both the delay and the follower fee as control variables that are trader tailored. The inclusion of customized follower fees would provide a finer degree of control for the platform: rather than running the risk of losing existing followers, the platform could offer a lower follower fee to maintain the existing set of followers associated with a trader. The platform could also consider offering a menu of options for followers, each with a given level of delay and an associated follower fee. The current options offered by the platform lie at the two extremes: zero delay and a follower fee versus a 30-minute delay and zero follower fee.

While the focus of our study was on the goals of the platform, our study also helps followers answer the question whether they are paying too much for financial advice. To

illustrate, consider a follower that faces an uncertain profit-gap (x), whose density $f(x)$ is Normal with a mean of $\alpha\sqrt{\delta}$ and variance of $\sigma^2\delta$. A natural question is: is paying a follower fee of c (per trade) justified? As a simple answer, a comparison could be made between the expected profit-gap ($\alpha\sqrt{\delta}$) and the following fee c: If $\alpha\sqrt{\delta} > c$, then real-time following is worthwhile. On the other hand, if risk aversion is a concern, then for an investor with a utility function $U(x)$, where x is the profit gap, we can replace the left side of the above inequality by the expected utility associated with a random profit gap, $\mathbb{E}(U(x)) = \int_{-\infty}^{+\infty} U(x)f(x)dx$.

At a broader level, our study is one of the first of its kind to explore the value of information in a real-world setting. The context in this study, namely the Forex market, provided us with an ideal testbed to study information value measured by the profit impact of delaying a trader's trades. This is because the Forex market is very speculative and volatile, and the impact of a small delay on the profit of a trade can be significant. In other markets, e.g., the traditional stock market where the fundamentals do not change rapidly, the impact of a small delay on profit may be too small to measure. Thus, to study the impact of delay in the stock market, it will be necessary to simulate much larger delays. The problem with introducing large delays, of course, is that exogenous events that occur during the delay period could affect the profit-gap. Even so, the problem of following in real-time versus doing so with delay is a prevalent phenomenon, not unique to the Forex market.

In this study, we discussed the question of *how* to follow (i.e., real-time versus delay) and the platform should manage delay in a way to maximize its revenue from commissions and following fees. Future work could explore the additional question of *whom* to follow. That is, it is possible that a trader-currency may not be worthwhile to follow at all, whether in real-time or with delay. Together, the joint decision of whom to follow, and if so, how to follow, is what must eventually be addressed for a more comprehensive study of the problem.

APPENDIX A

APPENDIX FOR CHAPTER 2

Text Analysis

Sentiment Analysis

This part describes the procedure of our sentiment analysis for labeling reviews. First, we conduct word segmentation and part-of-speech tagging to convert the unstructured textual data into individual words or phrases.[1] Next, the Naive Bayes method is used to train a classifier on the polarity of these words or phrases by manually labeling each review text from training set as positive or negative. The training set is composed of a large pool of labeled online reviews. Table A.1 lists some examples of positive and negative words from the review text. Finally, using this trained classifier, we calculate the sentiment score (the probability of a positive sentiment) for each review text to classify it as negative or positive. We validate our approach by comparing the polarity of reviews derived from our sentiment analysis with those derived directly from the review rating. When using the rating as the basis to derive polarity, we consider a rating as negative if the rating is below the mean, and positive otherwise. Table A.2 compares the agreement (the number of counts) between the labels derived from these two methods. We see that the review polarity obtained from sentiment analysis is generally consistent with that derived from the rating, with relatively low false positive and false negative rates, compared to the state-of-the-art sentiment mining tools that typically achieve about 70% of accuracy (Liu, 2015). We further conduct a Chi-squared Goodness-Of-Fit Test that results in a chi-square statistic of 1.79 with a p-value

[1]These steps were implemented using modules from the Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS), a popular Chinese lexical analyzer with a word segmentation accuracy 98.45%.

0.18. Therefore, we cannot reject the null hypothesis and hence the polarities defined from review ratings and text sentiments are statistically the same.

Table A.1. Example of Positive and Negative Words

Sentiment	Words
Positive	timely, good, clean, happy, friendly, best, tasty, reliable, affordable satisfied, lucky, beautiful, welcomed, enjoy, responsible, patiently
Negative	wasting, chaotic, yelling, noisy, rubbish, snobbish, inefficient, force disappointed, unfortunately, awkward, busy, failed, inconvenient, but

Table A.2. Comparison of Review Polarity defined by Rating versus Text Sentiment

		Rating		
		Positive	Negative	Total
Text Sentiment	Positive	33,204	9,890	43,094
	Negative	10,043	8,721	18,764
	Total	43,247	18,611	61,858

Machine versus Human Generated Response

We identify machine-generated responses as those that follow a standard template (regardless of the nature of the complaint). For example, to respond to negative customer reviews, a machine-generated response reads something like "Dear Customer, thanks for your review. We apologize for any inconvenience in your trip. We will try our best to improve our service. Hope you continue supporting us. Thanks." To respond to positive customer reviews, it reads something like "Dear Customer, thanks for your review. We are happy to know you have a good experience with our product or service. Hope you continue supporting us. Thanks."

To label machine-generated responses, we pre-defined a standard template-type of response for positive and negative reviews respectively, and then calculated the similarity score between each response text with the positive and negative template response texts using the term frequency-inverse document frequency (TF-IDF) value. Then we choose the maximum TF-IDF value as the similarity score. We label responses with high similarity score

(with similarity score greater than 0.6) as machine-generated responses. After labelling, we manually went over each response to make sure customized (or human-generated) responses were not mis-labelled as machine-generated.

Based on the above scheme, we find that when the current review is negative, most of the responses are customized (76%). This makes sense because when customers report a serious problem with a tour, a standard, template-like reply normally does not help address the complaint. Prior research shows that tailored (customized) management responses are effective in this regard (Wang and Chaudhry 2017). Thus, we have focused on customized responses since these responses are costly to firm, but they are necessary and effective.

Content Similarity between Review Texts

In this part, we first describe how the content similarity between negative reviews (in Section 4.2) is calculated. Suppose a review does not receive a response, we first extract its next and prior n (window size) reviews. Then we calculate a cosine similarity score between those negative review texts before and after the (focal) review (tuple) within the window size. The cosine similarity score is calculated based on the TF-IDF value. We repeat the above calculation for each review without a response (control group), and then compute the average similarity score. Following the same calculation, we then compute the average similarity score between negative review texts before and after reviews with a response (treatment group). We have used window sizes of 10 and 20 in the analysis. We then conduct t-test to empirically verify whether the mean values (similarity scores) of the treatment and control group are statistically different or not for each window size.

The similarity between positive reviews (in Section 8.1) is calculated following the same scheme.

Details of Quasi-Experiment

We use Propensity Score Matching (PSM) to first identify similar tours based on observed tour characteristics. The tour characteristics extracted from data include length of travel (with unit of days, Length), flexible length of travel or not (dummy, Flexibility), departure city (Departure), destination (Destination), travel agent (i.e. the travel company, Agent), transportation (including a flight or not, Transportation), travel type (e.g. with a tour guide or not, Type), hotel star level (Hotel), the average price of the tour (Price), and tour age (Age, the number of days since the first customer review arrived on Ctrip.com to December 31th, 2014, and it is log transformed). All these characteristics identified above are exogenous (because they predate the response and are very unlikely that they are systematically co-determined with the firm's response decisions), which is a critical condition for PSM to work well.

The treatment in our setting is firm responses. However, when matching tours, one concern is that tours never respond to any reviews are systematically different from those that have done so. Therefore, we attempt to match among the tours which have at least responded to reviews once. We focus on the three-month period after the first response in each tour. Tours without (with) responses in this period are defined as control (treatment) group. This design guarantees a high degree of similarity between the treatment and control groups since both provided responses. We then run a Probit model (as the first stage) to match the treated and control tours based on their predicted propensity scores. The Probit model results are shown in Table A.3. We use the nearest neighbor (NN) matching with replacement. We further check whether the covariates of the matched treated and control tours are balanced. Table A.4 shows after matching the tour characteristics are comparable. The PSM yields 68 tours in the treatment group and 49 in the control group (a few control tours are used as the matching firm multiple times).

Table A.3. Probit Regression of Responding to Reviews

Treatment	Coef.	Std. Err.	z value	p value
Length	0.189	0.215	0.880	0.379
Flexibility	-0.265	0.716	-0.370	0.711
Departure	-0.732	0.682	-1.070	0.283
Destination	-0.051	0.310	-0.170	0.869
Agent	-1.716	6.764	-0.250	0.800
Transportation	1.058	0.677	1.560	0.118
Type	-0.424	0.381	-1.110	0.265
Hotel	-0.214	0.269	-0.800	0.426
Price	-0.001	0.000	-1.770	0.077
Age	-2.367	0.996	-2.380	0.017
Intercept	-65.661	18.757	-3.500	0.000
Number of obs	117			
LR chi2(11)	89.760			
Prob >chi2	0.000			
Log likelihood	-34.670			

Table A.4. Balance Check Before and After Matching

	Mean Difference		p-value from t-test	
	Before Matching	After Matching	Before Matching	After Matching
Length	-0.124	1.346	0.398	0.115
Flexibility	0.179	-0.206	0.027	0.205
Departure	-0.137	0.059	0.061	0.398
Destination	-0.016	0.397	0.456	0.188
Agent	0.026	-0.015	0.191	0.160
Transportation	0.042	0.191	0.327	0.231
Type	0.127	-0.926	0.295	0.106
Hotel	-0.550	1.072	0.040	0.156
Price	-377.056	553.162	0.099	0.194
Age	-0.146	0.108	0.000	0.208

For each matched pair, let i index a tour, j index a matched pair, and t index the period. $Rating_{i,j,t}$ is the average review ratings for tour i in pair j at period t. $Treatment_{i,j}$ is a dummy variable with 1 representing a tour in the treatment group and 0 in the control group. $Period_{i,j,t}$ is a dummy variable representing the period before (with value 0) or after (with value 1) the first response in a tour. δ_j captures the pair-level fixed effect, and $\epsilon_{i,j,t}$ is the error term. Equation (A.1) specifies the DID econometric estimation model to study the

relationship between responses and future review ratings. Using equation (A.1), we examine the treatment effect of management response on future review ratings.

$$Rating_{i,j,t} = \beta_0 + \beta_1 Treatment_{i,j} Period_{i,j,t} + \beta_2 Treatment_{i,j} + \beta_3 Period_{i,j,t} + \epsilon_{i,j,t} \quad \text{(A.1)}$$

Table A.5 provides the estimation results. The estimated coefficient of interest corresponding to the treatment effect (β_1) is positive and statistically significant, suggesting a positive effect of firm response on review ratings. However, there is no significant difference on ratings between the treatment and control group in the before-response period. The control group exhibits a negative trend on rating from the before-response to the after-response period.

Table A.5. Results of DID in combination with PSM

$Rating_{i,j,t}$	Coef.	Std. Err.	t value	p value
$Intercept$	4.636	0.104	44.720	0.000
$Treatment_{i,j} * Period_{i,j,t}$	0.139	0.049	2.810	0.005
$Treatment_{i,j}$	0.016	0.035	0.450	0.654
$Period_{i,j,t}$	-0.065	0.035	-1.850	0.065
Pair-level Fixed Effect	Yes			
Number of obs	272			
F (70, 201)	1.870			
Prob >F	0.000			
R-squared	0.395			

As an alternative, we also conducted another quasi-experiment. Here tours with low responding rate (the total number of responses divided by the total number of reviews) less than 0.02 are defined as the control group and otherwise treatment group. The Probit model results are shown in Table A.6. We use the NN matching with replacement. We further check whether the covariates of the matched treated and control tours are balanced. Table A.7 shows that after matching the tour characteristics are comparable. The PSM yields 80 tours in the treatment group and 37 in the control group.

We focus on review ratings following each focal review with versus without a response controlling the state where the focal review locates. That is, for one focal review (r_0), we

calculate the average of its next n (e.g. 20) review ratings ($\sum_{i=1}^{n} \frac{r_i}{n}$) and record its current state value ($\sum_{i=0}^{-(n-1)} \frac{r_i}{n}$). We only consider those reviews where the pre- and post- reviews do not contain a response to expunge possible confounds. For each pair, we examine a review with response in the treatment tour and record the state value. We then identify a review of the control tour in the no-response period that has the same value of state as the treated tour. Then we investigate the impact of management response on the subsequent review ratings controlling different state values ($x_{i,j}$) and pair-level fixed effect (δ_j) by running regression defined in equation (A.2).

$$Rating_{i,j} = \beta_0 + \beta_1 Treatment_{i,j} + \beta_2 x_{i,j} + \delta_j + \epsilon_{i,j} \tag{A.2}$$

where i indexes a tour, j represents a matched pair, and $\epsilon_{i,j}$ is the error term. Table A.8 provides the estimation results. The estimated coefficient of interest corresponding to the treatment effect (β_1) is positive and statistically significant, suggesting a positive effect of firm response on review ratings.

Table A.6. Probit Regression of Responding to Reviews

Treatment	Coef.	Std. Err.	z value	p value
Length	0.077	0.145	0.530	0.598
Flexibility	-0.201	0.605	-0.330	0.739
Departure	-0.879	0.499	-1.760	0.078
Destination	-0.513	0.242	-2.120	0.034
Agent	0.849	0.781	1.090	0.277
Transportation	-0.025	0.595	-0.040	0.966
Type	-0.539	0.301	-1.790	0.073
Hotel	-0.233	0.234	-0.990	0.320
Price	0.361	0.270	1.340	0.180
Age	-0.194	0.923	-0.210	0.834
Intercept	-8.585	33.701	-0.250	0.799
Number of obs	117			
LR chi2(11)	28.140			
Prob >chi2	0.021			
Log likelihood	-58.940			

Table A.7. Balance Check Before and After Matching

	Mean Difference		p-value from t-test	
	Before Matching	After Matching	Before Matching	After Matching
Length	0.741	0.475	0.073	0.275
Flexibility	-0.007	0.025	0.472	0.437
Departure	-0.009	-0.050	0.464	0.367
Destination	-0.064	0.088	0.338	0.349
Agent	0.004	-0.038	0.464	0.282
Transportation	0.124	0.025	0.107	0.437
Type	-0.260	-0.125	0.151	0.379
Hotel	0.227	0.175	0.256	0.382
Price	549.830	429.663	0.043	0.107
Age	0.026	0.029	0.275	0.341

Table A.8. Results of Regression

$Rating_{i,j}$	Coef.	Std. Err.	t value	p value
$Intercept$	3.875	0.115	33.810	0.000
$Treatment_{i,j}$	0.066	0.011	6.290	0.000
$x_{i,j}$	0.128	0.023	5.570	0.000
Pair-level Fixed Effect	Yes			
Number of obs	1378			
F (81, 1296)	7.450			
Prob >F	0.000			
R-squared	0.318			

Initial Value Estimation

To run the Maximum Likelihood Estimation (MLE), we need to provide good initial values for quick convergence. It is common to use OLS to obtain initial parameter estimates, but for our problem simply applying OLS is not able to determine the initial values of α, ρ, β, and σ simultaneously, which we will elaborate later. Therefore we determine the initial values of the estimators by utilizing some of the specific characteristics of our data.

We first infer the review arrival rate λ and the negative review probability p directly from the data. A closer examination of the probability density function of y_t reveals that its log-likelihood function is maximized over the compounds k_1 and k_3. To exactly identify the optimal value of $\hat{\theta}$, it is necessary to further decompose the optimal values of $\hat{k}_1$ and $\hat{k}_3$.

Notice that $k_1 k_3 = b\beta p\lambda = 5\beta p\lambda$, and the term $\beta\lambda p$, always appears as a whole. To separate β out, we can first estimate λ and p from the data directly, and then solve for β from the equation $k_1 k_3 = 5\beta p\lambda$.

After estimating β, p, and λ, we compute the estimated value $\alpha + \rho(1-p)\lambda$ as a whole by solving the equation $k_1 = \alpha + \rho(1-p)\lambda + \beta p\lambda$. However, knowing the value of the expression $\alpha + \rho(1-p)\lambda$ still does not inform us whether an increased consumers' perception of quality (review rating) is due to the firm's control activity or due to the arrival of positive reviews. We propose the method to separate these two effects as follows.

We observe that specific to our data, many firms initially do not respond to reviews ($\alpha = 0$, in this period), and then begin to respond ($\alpha > 0$, after this period). Thus each tour can be split into two periods: pre-period (no-response period) where response is not allowed by the platform or adopted by the firm and post-period (after-no-response period) that is after the first response.[2] For these tours, we run MLE on the pre-period and are able to estimate ρ (boost parameter) using the data in pre-period where there were no responses.[3] Then using the data from post-period, we are able to recover the control parameter α by plugging in the estimated ρ and β in the previous step. This is how the initial values α_0, ρ_0, β_0, and σ_0 are generated. To summarize, the estimation procedure for the initial values (separately carried out for each tour) is described as:

1. Estimate the review arrival rate λ and the probability of negative reviews p from the pre-period.

2. Plugging $\hat{\lambda}$ and $\hat{p}$, estimate the boost parameter ρ_0 (by setting $\alpha = 0$) and damage parameter β_0 using the data in the pre-period.

[2]The pre-period (at the very beginning of the data generation process) expunges any possible impact from firm response.

[3]To run MLE on the pre-period (post-period) data, the initial value of ρ and β (α and σ) are obtained using OLS.

3. Plugging ρ_0, β_0, $\hat{\lambda}$, and $\hat{p}$, estimate the control parameter α_0, and the magnitude of the random component σ_0 using the data in the post-period.

Using these initial values $(\alpha_0, \rho_0, \beta_0, \sigma_0)$, we maximize the log-likelihood over α, ρ, β, and σ, and get the maximum likelihood estimators of the parameters.

Lastly, we would like to clarify that regarding to the generation of negative reviews, our SDE model allows the probability (proportion) of negative reviews (p) to change over time and vary under different response strategies. As described above, p is estimated from an initial period where no responses occur. This value of p is assumed to be constant, which is reasonable since there were no responses made by management during this period. We then use this value of p to estimate the management response strategy (α) in the portion of the data where the responses are seen to occur. During this period, the actual value of the probability of negative review (say p') can be different from the value during the no-response period. If the goal is to estimate this value of p', then we would need to absorb the effect of the management response into p'. This value of p' might indeed be different (typically lower) from p.

Besides, response to a positive (negative) review affects how that review is perceived by other customers or reviewers. In other words, responses have an indirect effect through reviews. Our model has the control term $\alpha(b - x_t)$ in an equation where the variables λ and p are independent of α, i.e., these variables are the arrival rate and probability of a negative review in the absence of responses. We inferred λ and p in the initial part of the data where there were no responses. Once responding begins, we expect λ and p to change according to:

$$\mathbb{E}(dx_t) = (\lambda(1-p)\rho(b - x_t) - \lambda p\beta x_t + \alpha(b - x_t))dt = (\lambda'(1-p')\rho(b - x_t) - \lambda' p'\beta x_t)dt.$$

Thus, the arrival rate and the probability of negative reviews appearing will change as a result of the indirect (moderating) effect of firm response.

Detailed Results by Tour

In this appendix, we present the detailed breakdown results tour by tour. Table A.9 shows the estimation results for the rest 107 tours. The first half of Table A.10 reports the steady state means estimated from SDE, ARMA, GARCH, MA, and ES, as well as the observed means inferred from the data. The second half of Table A.10 presents the in-sample RMSE for all the models considered for each tour. Table A.11 shows the in-sample MAE and SMAPE of all the methods.

With regard to the out-of-sample predictive performance, Table A.12, A.13, and A.14 present the RMSE, MAE, and SMAPE of all the 117 tours in terms of long-term and short-term out-of-sample prediction.

RMSE, MAE, and SMAPE are common metrics to compare model performance when the target to predict is continuous. We further examined several additional model comparison metrics by discretizing the state variable into a discrete variable, which we call "trending". This transformation is done by comparing the predicted value with the in-sample mean: if the predicted value is above the mean, we label it as 1 (up) and otherwise 0 (down). We then are able to calculate precision, recall, and F1 values (which is the harmonic mean of precision and recall) for the out-of-sample prediction. Table A.15 presents the F1 Score of long-term and short-term out-of-sample predictions. We conduct paired t-test to examine whether the two sets of means are statistically different and the results are shown in Table A.16. We can see that in terms of F1 score, SDE achieves comparable performance with respect to ARMA, but performs better than GARCH, MA, and ES, lending further support to the robustness of our analyses. Note that we dropped the naive method from consideration in calculating the F1 score as it becomes trivial after concretization.

Lastly we examined the metric of accuracy using the discretized state variable. Table A.17 shows the accuracy percentage for long-term and short-term out-of-sample predictions

with paired t-test results shown in Table A.18. We see that in terms of accuracy, SDE is again at least comparable with the benchmark methods.

To summarize, the SDE method is superior or comparable to all the benchmark methods, regardless of predictive performance metrics considered. However, it is the only method that can be applied in a prescriptive fashion.

Additional Robustness Checks

Different window size of the state variable

Instead of using the first two pages of reviews, we examine an alternative window size by constructing the state variable x_t using only the first page (10 reviews). Table A.19 reports the estimation results of the Ctrip data using a window size n equal to 10. We observe qualitatively the same results, e.g., there is a large variance in γ_1 and small variance in γ_2 across different tours. We find that the percentage of γ_1 values greater than 1 is larger than the case with window size of 20, meaning that the relative boosting impact from firms' control (as opposed to positive reviews) is bigger when considering one page than two pages.

Predictive Performance of SDE without response

To better understand the role of incorporating response in our SDE model, we conducted additional analyses by setting $\alpha = 0$, to emulate the case with no responses, which is denoted by SDE without response. We compare its predictive performance with our SDE model, ARMA and GARCH using the Ctrip data. The new model (SDE without response) performs worse than SDE, ARMA and GARCH in general. Table A.20 presents the mean and standard deviation (in parenthesis) of the out-of-sample (both long-term and short-term) RMSE for the 117 tours. We further conduct paired t-test to examine whether the two sets of results are statistically different as shown in Table A.21. The results clearly show that the model

without considering responses is inferior. This additional analysis further strengthens the importance of factoring in responses.

Sample Snapshot from Ctrip

Figure A.1 presents a translated snapshot of the page for writing reviews at Ctrip.com. As we can see in Ctrip's review page, it prominently lists recent reviews on the right side, and thus future reviewers have a higher chance to read them.

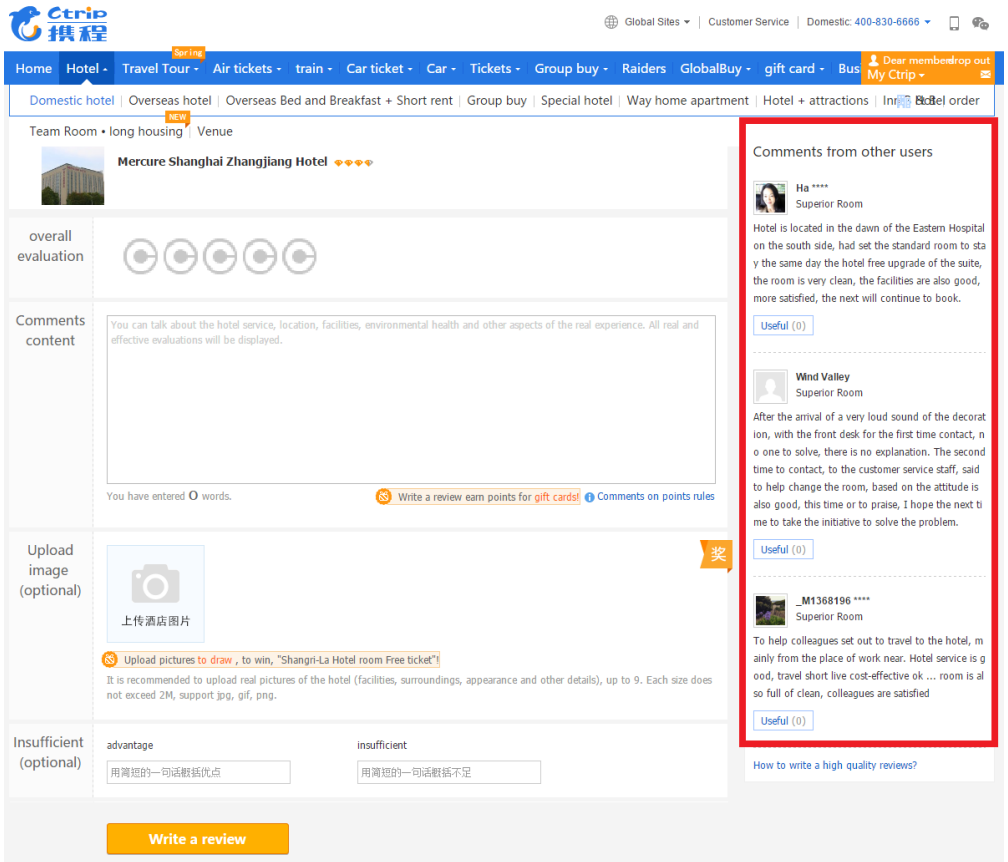

Figure A.1. Translated Snapshot of Reviewer's Writing Review Page from Ctrip.com

Table A.22 and A.23 present ARMA and GARCH estimates (with t-statistic in parenthesis) for 117 tours in Ctrip data.

Table A.9. The Parameter Estimation Results for the rest 107 tours in Ctrip Data

Tour ID	$\hat{\lambda}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\sigma}$	γ_1	γ_2
56737	7.437***(0.804)	0.199***(0.020)	0.053* (0.033)	0.011** (0.006)	0.005***(0.000)	0.261***(0.004)	0.78	0.14
71478	0.624***(0.093)	0.365***(0.030)	0.035***(0.005)	0.051***(0.014)	0.022***(0.002)	0.056***(0.003)	1.72	0.21
49049	1.565***(0.227)	0.333***(0.030)	0.035***(0.007)	0.023***(0.007)	0.010***(0.001)	0.097***(0.003)	1.49	0.20
72526	0.932***(0.214)	0.258***(0.028)	0.030***(0.010)	0.062***(0.015)	0.024***(0.003)	0.071***(0.004)	0.71	0.20
32391	1.735***(0.372)	0.309***(0.026)	0.012* (0.009)	0.046***(0.007)	0.011***(0.001)	0.107***(0.003)	0.21	0.37
72527	0.819***(0.187)	0.302***(0.051)	0.031***(0.008)	0.016 (0.013)	0.010***(0.002)	0.060***(0.003)	3.50	0.10
69278	1.199***(0.273)	0.278***(0.036)	0.025***(0.009)	0.032***(0.010)	0.011***(0.002)	0.087***(0.003)	0.93	0.20
28245	1.052***(0.351)	0.357***(0.077)	0.011** (0.006)	0.012* (0.009)	0.005***(0.001)	0.075***(0.003)	1.39	0.23
30814	0.802***(0.398)	0.288***(0.113)	0.022** (0.013)	0.010 (0.022)	0.008***(0.003)	0.090***(0.004)	3.83	0.08
16311	2.268***(0.293)	0.372***(0.020)	0.049***(0.009)	0.047***(0.006)	0.014***(0.001)	0.114***(0.003)	0.74	0.34
2753	0.339***(0.042)	0.187***(0.02)	0.021***(0.004)	0.017*(0.013)	0.033***(0.004)	0.05***(0.003)	4.56	0.04
3150	0.191***(0.054)	0.288***(0.013)	0.005*(0.003)	0.32***(0.02)	0.09***(0.005)	0.036***(0.003)	0.1	0.37
4171	0.231***(0.024)	0.313***(0.027)	0.017***(0.002)	0.022***(0.006)	0.027***(0.001)	0.043***(0.002)	4.88	0.08
5107	1.265*(0.928)	0.172**(0.088)	0.004 (0.004)	0.002 (0.005)	0.002 (0.002)	0.088***(0.001)	1.67	0.08
5452	0.506***(0.067)	0.292***(0.015)	0.003*(0.002)	0.072***(0.006)	0.015***(0.001)	0.061***(0.001)	0.12	0.37
5603	0.461**(0.227)	0.184***(0.043)	0.004*(0.003)	0.017**(0.007)	0.009***(0.002)	0.051***(0.002)	0.65	0.14
6712	0.137**(0.078)	0.264***(0.03)	0.001 (0.003)	0.195***(0.022)	0.068***(0.007)	0.03***(0.004)	0.07	0.33
7297	0.637**(0.32)	0.244**(0.119)	0.007**(0.004)	0 (0.01)	0.005**(0.003)	0.056***(0.002)	33.89	0.01
9474	0.389***(0.049)	0.282***(0.009)	0.005**(0.002)	0.16***(0.008)	0.04***(0.002)	0.046***(0.002)	0.11	0.35
11626	0.245***(0.102)	0.297***(0.03)	0.007*(0.005)	0.164***(0.029)	0.042***(0.006)	0.045***(0.004)	0.26	0.34
12043	0.184**(0.096)	0.121***(0.027)	0.003*(0.002)	0.031***(0.008)	0.03***(0.004)	0.029***(0.002)	0.65	0.08
17184	0.24***(0.062)	0.293***(0.051)	0.003*(0.002)	0.034***(0.01)	0.007***(0.001)	0.045***(0.002)	0.55	0.27
18201	0.193***(0.032)	0.2***(0.024)	0.01***(0.002)	0.031**(0.015)	0.047***(0.007)	0.04***(0.003)	2.14	0.08
20131	0.304 (0.604)	0.296**(0.144)	0.004 (0.013)	0.07**(0.04)	0.012**(0.005)	0.066***(0.008)	0.26	0.33
21148	0.18***(0.053)	0.321***(0.024)	0.003**(0.002)	0.121***(0.016)	0.027***(0.003)	0.028***(0.002)	0.21	0.39
21509	0.475***(0.054)	0.36***(0.028)	0.037***(0.006)	0.063***(0.007)	0.037***(0.001)	0.064***(0.004)	1.93	0.19
22510	0.435***(0.048)	0.422***(0.014)	0.019***(0.007)	0.448***(0.015)	0.064***(0.002)	0.083***(0.005)	0.17	0.63
22809	0.484***(0.042)	0.293***(0.006)	0.02**(0.01)	1.023***(0.014)	0.302***(0.004)	0.137***(0.007)	0.06	0.39
26561	0.356**(0.185)	0.327***(0.06)	0.003*(0.002)	0.021**(0.01)	0.008***(0.002)	0.041***(0.002)	0.53	0.32
28977	0.462***(0.151)	0.361***(0.04)	0.001 (0.002)	0.038***(0.007)	0.006***(0.001)	0.053***(0.001)	0.05	0.54
44981	0.188**(0.082)	0.375***(0.04)	0.002**(0.001)	0.044***(0.012)	0.012***(0.002)	0.026***(0.001)	0.41	0.43
46254	0.313***(0.051)	0.418***(0.012)	0.004*(0.003)	0.352***(0.016)	0.067***(0.003)	0.047***(0.003)	0.06	0.67
49063	0.214***(0.033)	0.366***(0.019)	0.011***(0.004)	0.356***(0.02)	0.062***(0.003)	0.054***(0.004)	0.23	0.47
49232	0.333***(0.065)	0.303***(0.041)	0.01***(0.002)	0.014*(0.01)	0.014***(0.002)	0.037***(0.002)	2.99	0.11
49435	0.421***(0.135)	0.262***(0.059)	0.01***(0.004)	0.016*(0.011)	0.011***(0.002)	0.048***(0.003)	2.11	0.11
50327	0.583***(0.034)	0.35***(0.009)	0.008*(0.006)	0.363***(0.009)	0.062***(0.001)	0.103***(0.003)	0.06	0.51
51324	0.599***(0.049)	0.219***(0.007)	0.005**(0.003)	0.136***(0.007)	0.055***(0.002)	0.069***(0.002)	0.07	0.26
52165	2.07***(0.19)	0.236***(0.008)	0.009*(0.006)	0.087***(0.003)	0.028***(0.001)	0.114***(0.002)	0.07	0.29
52950	0.284***(0.048)	0.206***(0.017)	0.011***(0.002)	0.05***(0.015)	0.056***(0.008)	0.034***(0.002)	1.02	0.13
53795	0.186***(0.027)	0.236***(0.015)	0.003*(0.002)	0.151***(0.015)	0.027***(0.002)	0.034***(0.002)	0.14	0.27
55106	0.174 (0.361)	0.642***(0.045)	0 (0.002)	0.2***(0.03)	0.018***(0.003)	0.026***(0.003)	0.01	1.76
55830	0.34***(0.088)	0.455***(0.03)	0.004*(0.002)	0.094***(0.017)	0.019***(0.003)	0.039***(0.002)	0.21	0.69
60195	0.485***(0.183)	0.152***(0.055)	0.014***(0.006)	0.001 (0.012)	0.019***(0.006)	0.058***(0.004)	24.2	0.01
60639	0.129*(0.081)	0.189***(0.05)	0.006***(0.002)	0.045**(0.024)	0.039***(0.012)	0.029***(0.004)	1.37	0.1
63798	0.3***(0.065)	0.532***(0.015)	0.003*(0.002)	0.277***(0.017)	0.034***(0.002)	0.036***(0.002)	0.09	1.05
66207	0.318 (0.312)	0.311*(0.238)	0.008**(0.004)	0.004 (0.026)	0.007 (0.008)	0.031***(0.004)	9.38	0.04
66215	0.272***(0.073)	0.267***(0.025)	0.005*(0.003)	0.103***(0.022)	0.037***(0.006)	0.04***(0.003)	0.23	0.3
66287	0.339***(0.021)	0.361***(0.006)	0.003**(0.002)	0.291***(0.007)	0.072***(0.002)	0.059***(0.002)	0.05	0.54
66865	0.838*(0.616)	0.103***(0.041)	0.003 (0.01)	0.02 (0.018)	0.012**(0.007)	0.07***(0.004)	0.22	0.09
69925	0.352***(0.057)	0.554***(0.023)	0.002 (0.005)	0.292***(0.018)	0.05***(0.003)	0.06***(0.005)	0.04	1.19
71821	0.35***(0.136)	0.333***(0.045)	0.009*(0.007)	0.108***(0.018)	0.019***(0.002)	0.049***(0.004)	0.34	0.37
72202	0.559***(0.231)	0.191***(0.044)	0.014*(0.009)	0.039***(0.016)	0.013***(0.003)	0.055***(0.004)	0.81	0.13
72532	0.707***(0.117)	0.225***(0.025)	0.027***(0.005)	0.024***(0.006)	0.023***(0.002)	0.061***(0.003)	2.04	0.1
72603	0.128*(0.095)	0.525***(0.148)	0.001*(0.001)	0.016 (0.017)	0.005**(0.002)	0.018***(0.002)	1.24	0.49
72723	2.243***(0.194)	0.141***(0.012)	0.026**(0.014)	0.001 (0.01)	0.005***(0)	0.144***(0.003)	9.29	0.02
72869	0.421***(0.048)	0.514***(0.007)	0.011***(0.004)	0.668***(0.018)	0.08***(0.002)	0.055***(0.003)	0.08	0.98

Notes. The estimation is based on the in-sample data for each tour.

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Tour ID	$\hat{\lambda}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\sigma}$	γ_1	γ_2
73163	0.272***(0.034)	0.364***(0.014)	0 (0.002)	0.182***(0.01)	0.034***(0.002)	0.042***(0.002)	0.01	0.57
73186	0.319***(0.074)	0.37***(0.076)	0.01***(0.002)	0.004 (0.01)	0.009***(0.002)	0.039***(0.002)	12.36	0.04
74983	0.401***(0.133)	0.333***(0.064)	0.011**(0.005)	0.029***(0.012)	0.011***(0.002)	0.042***(0.003)	1.38	0.21
78546	0.377***(0.136)	0.564***(0.037)	0.005***(0.002)	0.07***(0.011)	0.016***(0.002)	0.035***(0.002)	0.43	0.9
79499	1.629***(0.535)	0.38***(0.036)	0.06*(0.041)	0.226***(0.036)	0.046***(0.006)	0.122***(0.015)	0.26	0.49
79769	0.732***(0.182)	0.366***(0.056)	0.035*(0.027)	0.149***(0.025)	0.024***(0.002)	0.145***(0.012)	0.5	0.38
83878	0.304***(0.033)	0.515***(0.012)	0.004*(0.003)	0.459***(0.012)	0.063***(0.001)	0.059***(0.003)	0.06	1
84052	0.748***(0.053)	0.5***(0.015)	0.054***(0.009)	0.28***(0.007)	0.041***(0.001)	0.18***(0.005)	0.52	0.66
93877	0.777***(0.187)	0.216***(0.03)	0.024**(0.012)	0.067***(0.021)	0.024***(0.004)	0.076***(0.005)	0.6	0.17
1605378	3.638***(0.911)	0.421***(0.019)	0.031*(0.02)	0.116***(0.01)	0.021***(0.001)	0.124***(0.005)	0.13	0.64
1605628	0.52**(0.248)	0.485***(0.053)	0.016*(0.012)	0.17***(0.04)	0.016***(0.003)	0.047***(0.006)	0.34	0.7
1607268	1.714***(0.418)	0.522***(0.019)	0.015*(0.011)	0.161***(0.011)	0.024***(0.001)	0.094***(0.005)	0.11	0.98
1608443	0.473***(0.08)	0.339***(0.048)	0.042*(0.033)	0.188***(0.026)	0.041***(0.002)	0.291***(0.016)	0.71	0.3
1609587	2.147***(0.896)	0.436***(0.039)	0.04*(0.031)	0.141***(0.019)	0.024***(0.003)	0.131***(0.01)	0.24	0.63
1612648	0.606***(0.1)	0.158***(0.022)	0.033*(0.022)	0.105***(0.018)	0.065***(0.004)	0.224***(0.011)	0.61	0.12
1614660	0.569***(0.122)	0.196***(0.037)	0.071**(0.032)	0.064*(0.04)	0.082***(0.009)	0.164***(0.019)	2.42	0.07
1617321	0.769***(0.293)	0.222***(0.036)	0.013*(0.009)	0.048***(0.012)	0.022***(0.003)	0.073***(0.004)	0.47	0.19
1620855	4.505***(0.748)	0.414***(0.017)	0.037*(0.028)	0.143***(0.012)	0.03***(0.002)	0.154***(0.007)	0.1	0.64
1621632	2.546***(0.198)	0.399***(0.01)	0.035***(0.014)	0.207***(0.004)	0.039***(0.001)	0.148***(0.004)	0.11	0.6
1621699	0.828***(0.243)	0.196***(0.013)	0.029*(0.02)	0.306***(0.043)	0.123***(0.014)	0.089***(0.01)	0.14	0.21
1623624	1.59**(0.755)	0.227***(0.022)	0.065 (0.052)	0.289***(0.036)	0.091***(0.009)	0.133***(0.017)	0.18	0.25
1624666	1.18**(0.595)	0.309***(0.052)	0.06**(0.035)	0.121***(0.041)	0.024***(0.005)	0.085***(0.011)	0.61	0.28
1633110	0.577***(0.228)	0.258***(0.099)	0.015**(0.007)	0.001 (0.016)	0.017***(0.006)	0.075***(0.005)	30.31	0.01
85207	0.871***(0.171)	0.176***(0.015)	0.012*(0.009)	0.096***(0.014)	0.034***(0.004)	0.079***(0.004)	0.17	0.18
85282	0.652***(0.077)	0.281***(0.021)	0.013**(0.008)	0.108***(0.007)	0.024***(0.001)	0.117***(0.004)	0.26	0.31
93804	0.541***(0.06)	0.3***(0.01)	0.015**(0.007)	0.368***(0.009)	0.08***(0.002)	0.08***(0.004)	0.11	0.39
60051	0.13***(0.02)	0.4***(0.012)	0.017**(0.007)	1.825***(0.017)	0.437***(0.004)	0.127***(0.011)	0.12	0.6
44248	0.178***(0.036)	0.32***(0.039)	0.007***(0.001)	0.027**(0.012)	0.032***(0.004)	0.025***(0.002)	2.13	0.15
69982	0.326*(0.198)	0.24***(0.055)	0.005*(0.004)	0.036**(0.016)	0.025***(0.007)	0.044***(0.004)	0.57	0.2
76106	0.17***(0.016)	0.2***(0.01)	0.008***(0.002)	0.206***(0.016)	0.066***(0.003)	0.047***(0.002)	0.3	0.19
1601410	0.672***(0.268)	0.231***(0.042)	0.017**(0.01)	0.052***(0.017)	0.024***(0.005)	0.062***(0.005)	0.61	0.19
11869	0.314***(0.048)	0.25***(0.018)	0.003*(0.003)	0.09***(0.01)	0.025***(0.002)	0.052***(0.002)	0.16	0.29
78511	0.195***(0.025)	0.56***(0.01)	0.004***(0.001)	0.406***(0.015)	0.04***(0.001)	0.037***(0.002)	0.1	1.15
82421	0.251***(0.065)	0.4***(0.028)	0.014***(0.004)	0.189***(0.031)	0.048***(0.005)	0.043***(0.004)	0.49	0.45
15618	0.124***(0.023)	0.32***(0.006)	0.005*(0.003)	1.322***(0.021)	0.305***(0.005)	0.041***(0.004)	0.04	0.45
36953	0.143***(0.034)	0.333***(0.039)	0.002 (0.002)	0.103***(0.018)	0.017***(0.002)	0.037***(0.002)	0.25	0.4
66666	0.146***(0.011)	0.5***(0.007)	0.000 (0.001)	0.595***(0.007)	0.091***(0.001)	0.044***(0.002)	0	1
94870	0.535***(0.098)	0.28***(0.019)	0.048***(0.012)	0.223***(0.02)	0.065***(0.004)	0.063***(0.007)	0.56	0.25
91836	0.36***(0.04)	0.4***(0.01)	0.009**(0.004)	0.426***(0.017)	0.092***(0.003)	0.063***(0.004)	0.1	0.61
22349	0.182***(0.073)	0.28***(0.073)	0.006**(0.003)	0.022*(0.015)	0.018***(0.004)	0.031***(0.003)	2.21	0.12
83648	1.07***(0.178)	0.379***(0.041)	0.045***(0.011)	0.042***(0.009)	0.016***(0.001)	0.099***(0.005)	1.61	0.23
44983	0.298***(0.071)	0.174***(0.019)	0.006***(0.002)	0.039***(0.008)	0.037***(0.004)	0.036***(0.002)	0.66	0.13
71801	0.25***(0.019)	0.321***(0.008)	0.052***(0.008)	0.999***(0.012)	0.168***(0.002)	0.115***(0.005)	0.31	0.36
58776	0.128***(0.023)	0.24***(0.018)	0.008***(0.002)	0.155***(0.02)	0.059***(0.005)	0.028***(0.003)	0.53	0.21
74359	0.376**(0.167)	0.429***(0.021)	0.009**(0.005)	0.232***(0.02)	0.032***(0.002)	0.036***(0.004)	0.18	0.64
1603059	0.447***(0.132)	0.214***(0.021)	0.014**(0.008)	0.141***(0.015)	0.045***(0.004)	0.058***(0.005)	0.29	0.21
76500	7.223***(0.731)	0.373***(0.017)	0.708***(0.121)	0.245***(0.019)	0.034***(0.001)	0.269***(0.016)	0.64	0.36
66306	0.451***(0.037)	0.586***(0.013)	0.048***(0.004)	0.326***(0.025)	0.059***(0.003)	0.054***(0.004)	0.78	0.79
1626293	0.62***(0.068)	0.348***(0.011)	0.019**(0.01)	0.438***(0.015)	0.11***(0.003)	0.101***(0.006)	0.11	0.48
1629646	1.662***(0.307)	0.72***(0.023)	0.051**(0.028)	0.494***(0.036)	0.027***(0.001)	0.146***(0.011)	0.22	2.11
90928	0.555***(0.057)	0.68***(0.014)	0.033***(0.005)	0.41***(0.02)	0.036***(0.001)	0.065***(0.004)	0.45	1.46

Notes. The estimation is based on the in-sample data for each tour.

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.10. In-Sample Performance Results (Steady State Mean and RMSE)

	Steady State Mean						In-Sample RMSE				
Tour ID	$\hat{\mu}_{SDE}$	$\hat{\mu}_{ARMA}$	$\hat{\mu}_{GARCH}$	$\hat{\mu}_{MA}$	$\hat{\mu}_{ES}$	μ	SDE	ARMA	GARCH	MA	ES
5106	4.6352	4.6421	4.6385	4.8916	4.8943	4.6547	0.1364	0.1350	0.1376	0.1780	0.1837
72528	4.5189	4.5088	4.5247	4.5916	4.5943	4.5089	0.1504	0.1504	0.1576	0.1502	0.1502
73154	4.5507	4.5440	4.5664	4.7082	4.7055	4.5444	0.1491	0.1522	0.1529	0.2159	0.2145
80961	4.7626	4.7616	4.8101	4.8166	4.8110	4.7610	0.1552	0.1563	0.1614	0.1625	0.1622
29336	4.6961	4.6883	4.7240	4.7582	4.7555	4.6862	0.1363	0.1385	0.1466	0.1808	0.1754
23222	4.7165	4.7124	4.7098	4.7499	4.7499	4.7065	0.1251	0.1246	0.1331	0.1257	0.1244
30938	4.6682	4.6480	4.7442	4.6082	4.6055	4.6582	0.1776	0.1778	0.2139	0.1835	0.1837
71480	4.7005	4.6928	4.7576	4.7082	4.7055	4.6906	0.1567	0.1572	0.1704	0.1603	0.1598
88292	4.7203	4.7100	4.7907	4.6249	4.6332	4.7103	0.1681	0.1613	0.1786	0.2209	0.2167
1618693	4.7964	4.8063	4.8370	4.8249	4.8332	4.8057	0.1129	0.1130	0.1167	0.1156	0.1162
56737	4.7338	4.7406	4.7969	4.7999	4.7999	4.7416	0.1342	0.1371	0.1392	0.1348	0.1350
71478	4.5837	4.5682	4.6153	4.5499	4.5499	4.5700	0.1449	0.1464	0.2189	0.2390	0.2394
49049	4.5757	4.5899	4.6636	4.9749	4.9832	4.5987	0.1990	0.1571	0.2096	0.3640	0.3670
72526	4.6391	4.6631	4.6089	4.6665	4.6777	4.6585	0.1213	0.1209	0.1620	0.1693	0.1670
32391	4.5847	4.5960	4.5707	4.8499	4.8499	4.5989	0.1563	0.1596	0.1644	0.2515	0.2493
72527	4.7113	4.6964	4.6314	4.6999	4.6999	4.6940	0.1342	0.1401	0.2973	0.1405	0.1412
69278	4.6799	4.6822	4.6980	4.8499	4.8499	4.6887	0.1319	0.1322	0.1414	0.2161	0.2161
28245	4.5205	4.5106	4.6387	4.7999	4.7999	4.5535	0.1994	0.2018	0.2231	0.2889	0.2886
30814	4.6986	4.7082	4.7306	4.4832	4.4888	4.6886	0.1315	0.1305	0.1457	0.2360	0.2332
16311	4.5463	4.5510	4.5674	4.7166	4.7277	4.5476	0.1529	0.1533	0.1557	0.1959	0.2010
2753	4.6230	4.6581	4.8523	4.1331	4.1355	4.6131	0.2564	0.2517	0.4700	0.2416	0.2414
3150	4.5309	4.5996	4.5828	4.6499	4.6499	4.5408	0.0741	0.0651	0.0990	0.2070	0.2029
4171	4.5754	4.6067	4.6412	4.5699	4.5742	4.5854	0.1434	0.1427	0.1465	0.1764	0.1753
5107	4.6252	4.6397	4.7013	4.7999	4.7999	4.6263	0.1632	0.1669	0.1751	0.1993	0.1991
5452	4.6348	4.6522	4.6641	4.5932	4.5870	4.6354	0.1391	0.1380	0.1396	0.2078	0.2089
5603	4.6369	4.6939	4.7474	4.6499	4.6499	4.6467	0.1078	0.1088	0.1790	0.1560	0.1560
6712	4.4796	4.6918	4.8978	4.5834	4.5859	4.4895	0.1390	0.1489	0.4250	0.1781	0.1789
7297	4.5287	4.5459	4.5185	4.4565	4.4627	4.5190	0.1790	0.1777	0.2531	0.2664	0.2654
9474	4.5919	4.6448	4.5968	4.5032	4.5028	4.5886	0.1186	0.1279	0.1253	0.2630	0.2656
11626	4.6009	4.6168	4.6467	4.7067	4.7057	4.6108	0.1155	0.1151	0.1309	0.2272	0.2269
12043	4.6287	4.6548	4.6340	4.5768	4.5729	4.6289	0.0790	0.0785	0.1032	0.2375	0.2350
17184	4.7572	4.7284	4.7629	4.4399	4.4413	4.7473	0.1583	0.1502	0.1513	0.2310	0.2263
18201	4.4643	4.5317	4.2819	4.2266	4.2227	4.4060	0.2375	0.2268	0.3958	0.4029	0.4070
20131	4.7258	4.6338	4.6456	4.7434	4.7443	4.7358	0.1094	0.1062	0.2996	0.2427	0.2407
21148	4.6064	4.5909	4.6241	4.4463	4.4541	4.5984	0.1100	0.1096	0.1188	0.1139	0.1135
21509	4.4951	4.6069	4.5439	4.3999	4.3999	4.5050	0.1463	0.1461	0.1546	0.1936	0.1930
22510	4.5913	4.5770	4.5270	4.7097	4.7155	4.5918	0.1257	0.1251	0.2036	0.1520	0.1520
22809	4.4810	4.5099	4.4718	4.4767	4.4728	4.4828	0.1445	0.1598	0.1523	0.2892	0.2840
26561	4.4538	4.4654	4.5345	4.4699	4.4742	4.4479	0.1833	0.1990	0.2042	0.2764	0.2754
28977	4.6040	4.5864	4.6708	4.7699	4.7671	4.6102	0.1601	0.1613	0.1740	0.1653	0.1657
44981	4.4678	4.4695	4.4451	4.4299	4.4328	4.4486	0.1486	0.1412	0.1432	0.1562	0.1535
46254	4.4352	4.4545	4.4517	4.1466	4.1471	4.4268	0.1233	0.1264	0.1440	0.3042	0.3038
49063	4.6195	4.5692	4.6151	4.6094	4.6225	4.5809	0.1457	0.1443	0.1440	0.1541	0.1539
49232	4.5243	4.5361	4.5256	4.4133	4.4113	4.5241	0.0956	0.0956	0.1008	0.1922	0.1918
49435	4.6193	4.6656	4.6828	4.4065	4.4055	4.6081	0.1389	0.1471	0.1479	0.2410	0.2410
50327	4.6032	4.6014	4.5796	4.5833	4.5785	4.6029	0.1866	0.1853	0.1854	0.2376	0.2416
51324	4.5243	4.5294	4.5783	4.6465	4.6542	4.5240	0.1799	0.1804	0.1910	0.3589	0.3597
52165	4.5681	4.6475	4.6417	4.5832	4.5928	4.5706	0.1567	0.1600	0.1641	0.2822	0.2853

	Steady State Mean						In-Sample RMSE				
Tour ID	$\hat{\mu}_{SDE}$	$\hat{\mu}_{ARMA}$	$\hat{\mu}_{GARCH}$	$\hat{\mu}_{MA}$	$\hat{\mu}_{ES}$	μ	SDE	ARMA	GARCH	MA	ES
52950	4.3679	4.4099	4.2787	4.2898	4.2913	4.3637	0.1248	0.1090	0.4716	0.4583	0.4596
53795	4.7648	4.7849	4.8000	4.8965	4.8970	4.7707	0.0976	0.0983	0.1013	0.1298	0.1340
55106	4.3268	4.3092	4.3252	4.3635	4.3615	4.3338	0.0869	0.0871	0.0884	0.0896	0.0878
55830	4.4012	4.4603	4.4556	4.1432	4.1442	4.4062	0.1509	0.1643	0.1499	0.1900	0.1930
60195	4.5605	4.6133	4.5924	4.4231	4.4269	4.5559	0.1234	0.1141	0.1501	0.2628	0.2589
60639	4.6059	4.5662	4.7691	4.4106	4.4016	4.4770	0.1190	0.1214	0.3695	0.2146	0.2138
63798	4.4312	4.4141	4.4511	4.5399	4.5413	4.4335	0.1176	0.1174	0.1198	0.1273	0.1289
66207	4.6460	4.4759	4.4851	4.3904	4.3844	4.4108	0.1763	0.1765	0.2612	0.1885	0.1885
66215	4.5198	4.5737	4.7546	4.6463	4.6540	4.5296	0.1757	0.1764	0.3604	0.2275	0.2306
66287	4.4088	4.4169	4.3960	4.3999	4.4070	4.4088	0.1968	0.1965	0.1948	0.2510	0.2492
66865	4.7346	4.7997	4.8720	4.8632	4.8613	4.7443	0.1336	0.1338	0.1905	0.1551	0.1551
69925	4.1616	4.0761	4.3093	4.4892	4.5053	4.1707	0.3684	0.3038	0.4363	0.3883	0.3898
71821	4.7002	4.6814	4.6499	4.6399	4.6413	4.6999	0.0946	0.0945	0.1780	0.1000	0.0995
72202	4.7965	4.7833	4.8020	4.8495	4.8568	4.7874	0.0865	0.0861	0.0883	0.1257	0.1257
72532	4.5873	4.6608	4.6569	4.6732	4.6699	4.5971	0.1145	0.1151	0.1231	0.2282	0.2282
72603	4.2802	4.3201	4.2927	4.2499	4.2499	4.2815	0.1056	0.0971	0.1294	0.1984	0.2010
72723	4.7587	4.7428	4.7659	4.8366	4.8313	4.7687	0.1533	0.1438	0.1376	0.2317	0.2356
72869	4.4774	4.4618	4.4670	4.5965	4.5970	4.4759	0.0960	0.0959	0.0955	0.0958	0.0958
73163	4.5147	4.5228	4.5823	4.5633	4.5613	4.5118	0.2162	0.2162	0.2284	0.2161	0.2158
73186	4.5360	4.4722	4.5129	3.9966	3.9971	4.4659	0.2207	0.2237	0.2254	0.3376	0.3369
74983	4.6482	4.6000	4.5646	4.4800	4.4756	4.6510	0.1146	0.1278	0.1487	0.1779	0.1801
78546	4.1531	4.1715	4.0845	3.9366	3.9385	4.1463	0.1369	0.1094	0.3270	0.4560	0.4502
79499	4.5474	4.4577	4.4261	4.3936	4.3945	4.5374	0.1059	0.1084	0.1111	0.1392	0.1359
79769	4.7022	4.5644	4.5886	4.4994	4.5068	4.6923	0.0934	0.0940	0.0956	0.1313	0.1253
83878	4.3949	4.4235	4.3995	4.3966	4.3971	4.3947	0.2151	0.2114	0.2654	0.2088	0.2078
84052	4.5626	4.5462	4.5535	4.4232	4.4270	4.5625	0.1698	0.1701	0.1719	0.1760	0.1745
93877	4.7032	4.6971	4.7321	4.5129	4.5183	4.6932	0.1443	0.1433	0.1532	0.2183	0.2187
1605378	4.4834	4.4790	4.4839	4.4399	4.4413	4.4832	0.1269	0.1268	0.1292	0.1508	0.1508
1605628	4.6951	4.6090	4.6165	4.7558	4.7624	4.7051	0.1447	0.1438	0.3069	0.1680	0.1682
1607268	4.3601	4.3412	4.3803	4.2800	4.2757	4.3504	0.1393	0.1401	0.1596	0.4051	0.4025
1608443	4.6949	4.6146	4.6025	4.7499	4.7499	4.6995	0.2188	0.2060	0.2004	0.2069	0.2058
1609587	4.5205	4.4588	4.5140	4.5794	4.5896	4.5138	0.1278	0.1278	0.1395	0.1716	0.1709
1612648	4.6655	4.6989	4.8509	4.5367	4.5314	4.6642	0.2328	0.2052	0.2712	0.2712	0.2712
1614660	4.5802	4.6900	4.6190	4.3223	4.3338	4.5735	0.1571	0.1501	0.4708	0.2010	0.2014
1617321	4.5973	4.6027	4.6542	4.5498	4.5570	4.5914	0.1476	0.1484	0.1601	0.1713	0.1724
1620855	4.4122	4.4054	4.4101	4.2767	4.2657	4.4027	0.1669	0.1683	0.1740	0.2372	0.2312
1621632	4.4985	4.5187	4.5171	4.5398	4.5484	4.4996	0.1758	0.1756	0.1739	0.1738	0.1741
1621699	4.6055	4.7151	4.8200	4.5697	4.5742	4.6147	0.0990	0.0993	0.3621	0.1371	0.1365
1623624	4.6379	4.7126	4.7457	4.6739	4.6630	4.6382	0.1007	0.1017	0.1081	0.1227	0.1227
1624666	4.7369	4.6872	4.6881	4.6957	4.7037	4.7338	0.1060	0.1009	0.2755	0.2075	0.2046
1633110	4.3277	4.3529	4.6673	3.9999	3.9999	4.3189	0.3075	0.2095	0.4758	0.4707	0.4729
85207	4.6976	4.7500	4.7651	4.6499	4.6499	4.6974	0.1371	0.1276	0.1301	0.1895	0.1860
85282	4.6811	4.6883	4.7328	4.4932	4.4942	4.6783	0.1640	0.1705	0.1782	0.1740	0.1740
93804	4.6127	4.6240	4.6070	4.5999	4.5999	4.6121	0.1234	0.1237	0.1784	0.1313	0.1305
60051	4.3744	4.3851	4.3571	4.2400	4.2414	4.3781	0.1533	0.1545	0.1584	0.1660	0.1650
44248	4.2595	4.3643	4.3299	4.4134	4.4115	4.2695	0.1434	0.1494	0.1867	0.2409	0.2440
69982	4.3847	4.4640	4.4762	4.3999	4.3999	4.3890	0.1325	0.1345	0.1645	0.2697	0.2708
76106	4.7090	4.7129	4.7527	4.6399	4.6413	4.7043	0.1534	0.1825	0.1607	0.1566	0.1571
1601410	4.6032	4.5860	4.5995	4.5662	4.5711	4.5932	0.0906	0.0908	0.0901	0.1025	0.1083
11869	4.6387	4.6382	4.6279	4.6833	4.6785	4.6430	0.1558	0.1556	0.2306	0.1777	0.1786

	Steady State Mean						In-Sample RMSE				
Tour ID	$\hat{\mu}_{SDE}$	$\hat{\mu}_{ARMA}$	$\hat{\mu}_{GARCH}$	$\hat{\mu}_{MA}$	$\hat{\mu}_{ES}$	μ	SDE	ARMA	GARCH	MA	ES
78511	4.4893	4.4910	4.4360	4.3999	4.3999	4.4882	0.1177	0.1175	0.1266	0.1753	0.1701
82421	4.4880	4.4446	4.3822	4.3431	4.3441	4.4805	0.1245	0.1346	0.1330	0.1849	0.1771
15618	4.5284	4.5046	4.5340	4.4999	4.4999	4.5270	0.0760	0.0801	0.4833	0.1234	0.1211
36953	4.6925	4.6804	4.7021	4.7635	4.7544	4.7025	0.1566	0.1528	0.1412	0.1480	0.1465
66666	4.3377	4.3502	4.3193	4.2933	4.2871	4.3422	0.1747	0.1665	0.1720	0.1861	0.1832
94870	4.6595	4.6565	4.6691	4.6397	4.6412	4.6655	0.1149	0.1194	0.1047	0.1178	0.1142
91836	4.4234	4.3992	4.5619	4.3034	4.2957	4.4331	0.1548	0.1614	0.3223	0.1812	0.1827
22349	4.5442	4.5192	4.6298	4.3296	4.3326	4.5438	0.1590	0.1158	0.1699	0.2170	0.2167
83648	4.5822	4.5711	4.4989	4.3898	4.3984	4.5899	0.1813	0.1778	0.1960	0.1760	0.1760
44983	4.4668	4.4982	4.5682	4.6031	4.6098	4.4638	0.1735	0.1719	0.2072	0.3714	0.3649
71801	4.7133	4.7056	4.6535	4.5900	4.5843	4.7144	0.0969	0.0984	0.1109	0.1209	0.1237
58776	4.6379	4.6144	4.5289	4.4439	4.4302	4.6314	0.1402	0.1349	0.1484	0.1524	0.1499
74359	4.5922	4.5639	4.5606	4.6096	4.6155	4.5826	0.0566	0.0565	0.0554	0.1162	0.1141
1603059	4.6827	4.6683	4.7571	4.6063	4.6126	4.6727	0.1019	0.1016	0.2846	0.1280	0.1280
76500	4.7610	4.7014	4.7085	4.7666	4.7642	4.7590	0.1056	0.1056	0.1215	0.1817	0.1778
66306	4.3690	4.2041	4.3756	4.0231	4.0269	4.3591	0.1587	0.1135	0.2430	0.3753	0.3745
1626293	4.4624	4.4702	4.4335	4.5863	4.5954	4.4691	0.1704	0.1688	0.1737	0.1665	0.1664
1629646	4.4894	4.4830	4.6025	4.4398	4.4413	4.4994	0.1294	0.1223	0.2958	0.5764	0.5786
90928	4.4284	4.4047	4.4628	4.2931	4.2941	4.4289	0.1747	0.1865	0.3413	0.2452	0.2449
average	**4.5680**	**4.5706**	**4.5937**	**4.5374**	**4.5387**	**4.5636**	**0.1457**	**0.1431**	**0.1952**	**0.2107**	**0.2102**

Table A.11. In-Sample Performance Results (MAE and SMAPE)

	MAE					SMAPE				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
5106	0.1079	0.1102	0.1094	0.1465	0.1510	0.1788	0.1833	0.1804	0.2554	0.2651
72528	0.1185	0.1185	0.1179	0.1183	0.1183	0.1343	0.1343	0.1337	0.1340	0.1340
73154	0.1194	0.1225	0.1214	0.1757	0.1743	0.1416	0.1543	0.1443	0.2223	0.2200
80961	0.1242	0.1236	0.1198	0.1207	0.1205	0.2712	0.2775	0.2706	0.2731	0.2724
29336	0.1132	0.1155	0.1170	0.1423	0.1370	0.1982	0.2013	0.2058	0.2603	0.2476
23222	0.1040	0.1034	0.1030	0.1034	0.1026	0.2003	0.1993	0.2019	0.1986	0.1977
30938	0.1489	0.1491	0.1641	0.1554	0.1556	0.2312	0.2315	0.2628	0.2365	0.2365
71480	0.1250	0.1285	0.1308	0.1278	0.1277	0.2172	0.2213	0.2284	0.2219	0.2215
88292	0.1384	0.1294	0.1349	0.1911	0.1873	0.2479	0.2397	0.2561	0.3011	0.2978
1618693	0.0902	0.0905	0.0919	0.0916	0.0918	0.2561	0.2571	0.2646	0.2631	0.2640
56737	0.1066	0.1072	0.1036	0.1018	0.1016	0.2511	0.2529	0.2524	0.2456	0.2454
71478	0.1188	0.1207	0.1428	0.2126	0.2130	0.2394	0.2425	0.2649	0.3410	0.3413
49049	0.1606	0.1234	0.1386	0.3243	0.3276	0.2298	0.2021	0.2076	0.8400	0.8603
72526	0.1019	0.1017	0.1225	0.1425	0.1403	0.2140	0.2134	0.2490	0.2622	0.2597
32391	0.1299	0.1325	0.1367	0.2038	0.2016	0.1857	0.1886	0.1946	0.3239	0.3188
72527	0.1124	0.1182	0.1579	0.1195	0.1200	0.2590	0.2670	0.3042	0.2692	0.2696
69278	0.1053	0.1057	0.1115	0.1763	0.1763	0.1770	0.1782	0.1886	0.3379	0.3379
28245	0.1676	0.1697	0.1722	0.2205	0.2201	0.2182	0.2212	0.2267	0.3119	0.3112
30814	0.1105	0.1102	0.1071	0.2088	0.2062	0.1710	0.1708	0.1664	0.2698	0.2676
16311	0.1196	0.1189	0.1188	0.1545	0.1592	0.1481	0.1475	0.1473	0.1987	0.2062
2753	0.2134	0.2110	0.2996	0.2059	0.2060	0.3005	0.2967	0.4070	0.2884	0.2883
3150	0.0603	0.0538	0.0768	0.1851	0.1811	0.0779	0.0669	0.1006	0.2823	0.2741
4171	0.1094	0.1079	0.1100	0.1330	0.1320	0.1423	0.1406	0.1435	0.1785	0.1768
5107	0.1349	0.1365	0.1354	0.1505	0.1503	0.1950	0.1984	0.1990	0.2250	0.2245
5452	0.1127	0.1100	0.1118	0.1685	0.1695	0.1605	0.1570	0.1598	0.2636	0.2657
5603	0.0874	0.0882	0.1103	0.1245	0.1245	0.1473	0.1486	0.1763	0.2229	0.2229
6712	0.1160	0.1235	0.3015	0.1381	0.1382	0.1744	0.1833	0.4293	0.2063	0.2064
7297	0.1452	0.1419	0.1717	0.2251	0.2241	0.1763	0.1716	0.1939	0.2833	0.2817
9474	0.0900	0.0961	0.0944	0.2333	0.2360	0.1150	0.1229	0.1212	0.3814	0.3880
11626	0.0998	0.0986	0.1035	0.1910	0.1906	0.1286	0.1273	0.1383	0.2972	0.2963
12043	0.0667	0.0668	0.0847	0.2174	0.2148	0.0976	0.0976	0.1305	0.4334	0.4254
17184	0.1202	0.1126	0.1096	0.2089	0.2044	0.2261	0.2174	0.2147	0.3294	0.3254
18201	0.1996	0.1938	0.2553	0.3349	0.3383	0.1928	0.1895	0.2302	0.3653	0.3707
20131	0.0897	0.0863	0.1519	0.2083	0.2060	0.1354	0.1299	0.1930	0.3982	0.3913
21148	0.0922	0.0918	0.0933	0.0912	0.0910	0.1085	0.1080	0.1098	0.1072	0.1070
21509	0.1223	0.1217	0.1287	0.1568	0.1563	0.1555	0.1547	0.1617	0.2049	0.2040
22510	0.0967	0.0988	0.1352	0.1129	0.1129	0.1228	0.1254	0.1558	0.1378	0.1378
22809	0.1158	0.1322	0.1245	0.2488	0.2433	0.1166	0.1353	0.1265	0.3008	0.2914
26561	0.1558	0.1610	0.1576	0.2250	0.2239	0.1535	0.1559	0.1549	0.2375	0.2359
28977	0.1346	0.1371	0.1411	0.1404	0.1406	0.1775	0.1801	0.1877	0.1830	0.1832
44981	0.1220	0.1161	0.1141	0.1200	0.1191	0.1237	0.1180	0.1157	0.1202	0.1195
46254	0.0943	0.0960	0.1119	0.2688	0.2684	0.0815	0.0832	0.1000	0.2931	0.2924
49063	0.1135	0.1136	0.1134	0.1184	0.1183	0.1439	0.1443	0.1449	0.1518	0.1516
49232	0.0779	0.0779	0.0828	0.1661	0.1657	0.0862	0.0863	0.0917	0.2060	0.2053

			MAE					SMAPE		
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
49435	0.1090	0.1189	0.1195	0.1938	0.1938	0.1635	0.1785	0.1823	0.3410	0.3410
50327	0.1518	0.1506	0.1522	0.1970	0.2006	0.2033	0.2017	0.2022	0.2371	0.2398
51324	0.1456	0.1458	0.1498	0.3104	0.3112	0.1661	0.1664	0.1716	0.4547	0.4567
52165	0.1190	0.1207	0.1245	0.2384	0.2416	0.1841	0.1900	0.1905	0.4480	0.4575
52950	0.0949	0.0835	0.1902	0.4303	0.4317	0.0885	0.0770	0.1279	0.5469	0.5497
53795	0.0840	0.0850	0.0850	0.1050	0.1092	0.1964	0.2000	0.2006	0.2564	0.2698
55106	0.0697	0.0699	0.0706	0.0736	0.0702	0.0504	0.0505	0.0510	0.0531	0.0507
55830	0.1224	0.1360	0.1189	0.1576	0.1603	0.1141	0.1296	0.1111	0.1405	0.1425
60195	0.1040	0.0944	0.1270	0.2222	0.2186	0.1428	0.1299	0.1765	0.3586	0.3498
60639	0.1013	0.1055	0.2752	0.1668	0.1665	0.1322	0.1359	0.3363	0.2065	0.2061
63798	0.0992	0.0991	0.0979	0.1042	0.1054	0.0875	0.0875	0.0864	0.0923	0.0934
66207	0.1287	0.1306	0.2005	0.1579	0.1579	0.1646	0.1662	0.2474	0.1937	0.1937
66215	0.1486	0.1493	0.2703	0.1788	0.1806	0.1852	0.1860	0.3413	0.2257	0.2284
66287	0.1578	0.1577	0.1552	0.2067	0.2049	0.1423	0.1422	0.1390	0.1918	0.1898
66865	0.1044	0.1051	0.1321	0.1086	0.1086	0.2504	0.2521	0.3121	0.2582	0.2582
69925	0.2727	0.2302	0.2668	0.3155	0.3170	0.1651	0.1475	0.1803	0.2353	0.2368
71821	0.0801	0.0797	0.1093	0.0739	0.0734	0.1337	0.1329	0.1690	0.1225	0.1219
72202	0.0664	0.0661	0.0674	0.1065	0.1065	0.1703	0.1697	0.1722	0.2368	0.2368
72532	0.0904	0.0901	0.0965	0.1975	0.1975	0.1416	0.1414	0.1521	0.3657	0.3657
72603	0.0850	0.0760	0.0990	0.1717	0.1739	0.0647	0.0578	0.0753	0.1364	0.1384
72723	0.1227	0.1145	0.1109	0.2005	0.2046	0.2945	0.2774	0.2777	0.3633	0.3670
72869	0.0792	0.0791	0.0784	0.0777	0.0777	0.0737	0.0737	0.0730	0.0723	0.0723
73163	0.1615	0.1609	0.1576	0.1584	0.1589	0.1604	0.1599	0.1561	0.1571	0.1577
73186	0.1632	0.1636	0.1610	0.2677	0.2669	0.1449	0.1452	0.1458	0.2796	0.2784
74983	0.1007	0.1073	0.1174	0.1396	0.1416	0.1551	0.1622	0.1716	0.1928	0.1946
78546	0.1158	0.0915	0.1912	0.4275	0.4217	0.0690	0.0546	0.1146	0.3325	0.3263
79499	0.0894	0.0922	0.0930	0.1116	0.1093	0.0826	0.0852	0.0858	0.1046	0.1022
79769	0.0694	0.0700	0.0706	0.1115	0.1043	0.0834	0.0841	0.0848	0.1262	0.1193
83878	0.1600	0.1591	0.2076	0.1560	0.1554	0.1311	0.1303	0.1741	0.1277	0.1271
84052	0.1341	0.1346	0.1362	0.1412	0.1404	0.1549	0.1554	0.1571	0.1609	0.1604
93877	0.1209	0.1197	0.1271	0.1759	0.1763	0.2049	0.2032	0.2190	0.3308	0.3318
1605378	0.0989	0.0988	0.1016	0.1241	0.1241	0.0949	0.0948	0.0974	0.1163	0.1163
1605628	0.1142	0.1131	0.1904	0.1413	0.1414	0.1212	0.1200	0.2023	0.1478	0.1479
1607268	0.1128	0.1126	0.1250	0.3744	0.3716	0.0846	0.0843	0.0969	0.3838	0.3797
1608443	0.1855	0.1751	0.1699	0.1737	0.1732	0.2645	0.2573	0.2498	0.2513	0.2512
1609587	0.1054	0.1055	0.1112	0.1441	0.1435	0.0986	0.0987	0.1041	0.1291	0.1287
1612648	0.1816	0.1720	0.1981	0.1980	0.1980	0.3146	0.2971	0.3583	0.3577	0.3577
1614660	0.1270	0.1214	0.2748	0.1661	0.1665	0.2129	0.2026	0.3079	0.2940	0.2948
1617321	0.1232	0.1230	0.1268	0.1327	0.1331	0.1597	0.1595	0.1647	0.1730	0.1736
1620855	0.1389	0.1399	0.1465	0.1940	0.1885	0.1204	0.1211	0.1268	0.1737	0.1678
1621632	0.1400	0.1399	0.1383	0.1381	0.1382	0.1520	0.1519	0.1502	0.1501	0.1500
1621699	0.0781	0.0782	0.2695	0.1079	0.1073	0.1416	0.1419	0.4370	0.2045	0.2031
1623624	0.0844	0.0853	0.0870	0.1035	0.1035	0.1525	0.1539	0.1564	0.1792	0.1792
1624666	0.0911	0.0862	0.1686	0.1824	0.1796	0.1357	0.1284	0.2342	0.3160	0.3091
1633110	0.2645	0.1862	0.3919	0.3691	0.3713	0.2165	0.1637	0.3611	0.3389	0.3421
85207	0.1115	0.1040	0.1039	0.1525	0.1493	0.2493	0.2198	0.2218	0.3665	0.3555

			MAE					SMAPE		
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
85282	0.1290	0.1376	0.1407	0.1416	0.1416	0.2233	0.2347	0.2425	0.2372	0.2372
93804	0.1013	0.1015	0.1316	0.1055	0.1049	0.1406	0.1410	0.1748	0.1472	0.1463
60051	0.1224	0.1245	0.1215	0.1253	0.1246	0.1075	0.1092	0.1063	0.1089	0.1084
44248	0.1092	0.1153	0.1498	0.2058	0.2085	0.0922	0.0968	0.1235	0.1741	0.1767
69982	0.1062	0.1116	0.1402	0.2246	0.2257	0.1062	0.1115	0.1397	0.2450	0.2466
76106	0.1279	0.1498	0.1283	0.1268	0.1269	0.2242	0.2807	0.2283	0.2252	0.2256
1601410	0.0749	0.0751	0.0711	0.0833	0.0878	0.0897	0.0899	0.0852	0.1008	0.1068
11869	0.1232	0.1231	0.1517	0.1439	0.1446	0.1910	0.1912	0.2155	0.2292	0.2303
78511	0.1001	0.0997	0.1025	0.1375	0.1335	0.0999	0.0996	0.1014	0.1283	0.1254
82421	0.1034	0.1149	0.1123	0.1536	0.1480	0.0881	0.0978	0.0961	0.1351	0.1294
15618	0.0604	0.0586	0.2802	0.1003	0.0978	0.0663	0.0644	0.2481	0.1021	0.0999
36953	0.1311	0.1327	0.1213	0.1262	0.1255	0.2307	0.2372	0.2194	0.2302	0.2287
66666	0.1476	0.1367	0.1406	0.1545	0.1528	0.1195	0.1109	0.1137	0.1261	0.1246
94870	0.1000	0.1019	0.0844	0.0908	0.0885	0.1416	0.1454	0.1211	0.1309	0.1271
91836	0.1256	0.1316	0.1966	0.1339	0.1348	0.1175	0.1222	0.1738	0.1284	0.1295
22349	0.1360	0.0953	0.1317	0.1627	0.1623	0.1441	0.1042	0.1447	0.1880	0.1874
83648	0.1486	0.1487	0.1518	0.1493	0.1493	0.1973	0.1983	0.1969	0.2006	0.2006
44983	0.1432	0.1465	0.1565	0.3207	0.3135	0.1367	0.1455	0.1538	0.4123	0.3978
71801	0.0796	0.0812	0.0828	0.0922	0.0949	0.1472	0.1499	0.1492	0.1610	0.1644
58776	0.1243	0.1189	0.1217	0.1250	0.1239	0.1626	0.1558	0.1563	0.1658	0.1642
74359	0.0462	0.0460	0.0444	0.1041	0.1018	0.0538	0.0536	0.0517	0.1311	0.1278
1603059	0.0833	0.0834	0.1697	0.0955	0.0955	0.1274	0.1276	0.2295	0.1486	0.1486
76500	0.0881	0.0878	0.1032	0.1465	0.1435	0.1643	0.1638	0.1862	0.2300	0.2272
66306	0.1239	0.0894	0.1654	0.3380	0.3372	0.0808	0.0620	0.1070	0.1991	0.1988
1626293	0.1449	0.1449	0.1440	0.1440	0.1435	0.1490	0.1491	0.1474	0.1486	0.1480
1629646	0.1061	0.1001	0.2070	0.5340	0.5363	0.1062	0.0980	0.2067	0.3500	0.3510
90928	0.1435	0.1481	0.1939	0.2113	0.2110	0.1245	0.1289	0.1574	0.1710	0.1708
average	**0.1182**	**0.1163**	**0.1414**	**0.1750**	**0.1745**	**0.1569**	**0.1561**	**0.1826**	**0.2372**	**0.2367**

Table A.12. Out-of-Sample Predictive Performance Results (RMSE)

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
5106	0.1742	0.1887	0.1767	0.1696	0.1696	0.1888	0.1705	0.1577	0.2128	0.1972	0.1951	0.1910
72528	0.1543	0.1538	0.1658	0.1510	0.1519	0.1520	0.0700	0.0677	0.1244	0.0878	0.0870	0.0849
73154	0.2000	0.2023	0.2117	0.2566	0.2508	0.2005	0.1493	0.1772	0.1318	0.1710	0.1700	0.1370
80961	0.1594	0.1429	0.1527	0.2658	0.2634	0.1443	0.1364	0.1776	0.0843	0.1719	0.1681	0.1151
29336	0.1841	0.1889	0.2173	0.1763	0.1763	0.1913	0.0917	0.0727	0.0755	0.1027	0.1014	0.0723
23222	0.2073	0.2083	0.2635	0.3572	0.3703	0.2397	0.1715	0.1630	0.2967	0.3440	0.3403	0.2617
30938	0.1619	0.1611	0.2935	0.1567	0.1561	0.1576	0.0846	0.0819	1.8221	0.0961	0.0959	0.0806
71480	0.1525	0.1399	0.1429	0.2304	0.2300	0.1411	0.0811	0.0527	0.0618	0.0991	0.0980	0.0710
88292	0.1787	0.1325	0.1528	0.2531	0.2580	0.1295	0.0662	0.0493	0.0701	0.0910	0.0888	0.0494
1618693	0.1759	0.1727	0.1809	0.2255	0.2288	0.1716	0.0263	0.0625	0.1125	0.0248	0.0243	0.1411
56737	0.1624	0.2029	0.2021	0.2128	0.2198	0.1841	0.1415	0.1919	0.2786	0.1720	0.1704	0.2418
71478	0.1776	0.1741	0.3678	0.1363	0.1394	0.1971	0.1207	0.1207	1.4871	0.1421	0.1411	0.2248
49049	0.3122	0.2362	0.2850	0.2955	0.2970	0.2470	0.1311	0.2439	1.9452	0.1452	0.1439	0.2219
72526	0.1736	0.1718	0.3175	0.1872	0.1850	0.1677	0.1042	0.1127	1.6985	0.1088	0.1068	0.1917
32391	0.1556	0.1764	0.1631	0.2414	0.2436	0.1564	0.0709	0.2248	0.0826	0.0836	0.0834	0.0994
72527	0.1618	0.1728	0.3190	0.1561	0.1556	0.2001	0.1127	0.0915	1.4054	0.1105	0.1102	0.1300
69278	0.1965	0.2000	0.2237	0.2048	0.2017	0.2056	0.1419	0.1370	0.1200	0.1581	0.1581	0.1303
28245	0.1823	0.1833	0.2044	0.1788	0.1787	0.1788	0.1961	0.1937	0.3396	0.2086	0.2067	0.2586
30814	0.1393	0.1397	0.1061	0.2027	0.2019	0.1388	0.1676	0.1711	0.1524	0.2019	0.1997	0.2035
16311	0.1669	0.1697	0.1729	0.1944	0.1949	0.1667	0.0622	0.0595	0.0681	0.0694	0.0688	0.0541
2753	0.2854	0.2519	0.3369	0.3436	0.3436	0.1758	0.1746	0.1635	1.7875	0.2184	0.2162	0.1245
3150	0.1049	0.0968	0.1125	0.0943	0.0921	0.1248	0.0944	0.1034	0.1319	0.1040	0.1030	0.1478
4171	0.1462	0.4089	0.2347	0.2394	0.2402	0.2225	0.1678	0.4182	0.3939	0.2325	0.2315	0.3675
5107	0.1888	0.2070	0.2438	0.2053	0.2008	0.2081	0.1089	0.2359	0.2132	0.1088	0.1082	0.1589
5452	0.1663	0.1767	0.1852	0.2115	0.2115	0.1666	0.0654	0.0607	0.0672	0.0741	0.0738	0.0539
5603	0.2693	0.2685	0.3466	0.2626	0.2675	0.2678	0.3099	0.3094	1.8491	0.3317	0.3306	0.3116
6712	0.1827	0.6316	0.4906	0.2023	0.2075	0.2448	0.1999	0.3687	1.3335	0.2107	0.2095	0.2732
7297	0.3237	0.3074	0.4024	0.4526	0.4537	0.3200	0.0778	0.0696	2.1536	0.1271	0.1260	0.1854
9474	0.1482	0.1521	0.1491	0.1619	0.1608	0.1492	0.1534	0.1352	0.1512	0.2117	0.2099	0.1503
11626	0.0924	0.0985	0.1394	0.0860	0.0863	0.1159	0.0856	0.0890	0.1279	0.0985	0.0979	0.1044
12043	0.1016	0.1054	0.1109	0.1095	0.1107	0.1239	0.1093	0.1086	0.1016	0.1547	0.1540	0.1056
17184	0.1548	0.1519	0.1639	0.1377	0.1372	0.1479	0.1189	0.1051	0.0937	0.1169	0.1173	0.0878
18201	0.1032	0.1186	0.4694	0.1983	0.1945	0.2377	0.1083	0.0974	2.4380	0.1682	0.1661	0.2197
20131	0.1467	0.1688	0.2761	0.3086	0.3072	0.1054	0.1348	0.1421	1.4164	0.1840	0.1817	0.1157
21148	0.2881	0.2836	0.3072	0.3804	0.3834	0.2778	0.1669	0.1633	0.1802	0.2136	0.2121	0.1672
21509	0.3038	0.2990	0.2651	0.4282	0.4314	0.2870	0.1498	0.1446	0.1476	0.1979	0.1969	0.1459
22510	0.0819	0.0835	0.3513	0.2133	0.2133	0.1231	0.0631	0.0544	1.7399	0.1404	0.1389	0.1083
22809	0.1735	0.1389	0.1387	0.2394	0.2370	0.1289	0.1627	0.2578	0.1767	0.1843	0.1815	0.1578
26561	0.1531	0.1808	0.1644	0.1836	0.1806	0.1510	0.0927	0.2425	0.0748	0.0923	0.0912	0.1124
28977	0.2475	0.2397	0.2983	0.1975	0.1981	0.2508	0.1268	0.1251	0.2639	0.1411	0.1401	0.2008
44981	0.1061	0.1382	0.1652	0.1484	0.1422	0.1385	0.0707	0.1807	0.1495	0.0887	0.0873	0.1200
46254	0.1090	0.1088	0.1146	0.1535	0.1568	0.1170	0.0597	0.0605	0.0784	0.0743	0.0741	0.0880
49063	0.1994	0.1949	0.2108	0.2863	0.2781	0.2008	0.1362	0.1284	0.1084	0.1975	0.1937	0.1165
49232	0.1981	0.1964	0.1891	0.2914	0.2855	0.1859	0.2176	0.2231	0.3138	0.2812	0.2784	0.3014
49435	0.1849	0.1908	0.1790	0.2834	0.2830	0.1732	0.1465	0.2279	0.1548	0.1839	0.1835	0.1587
50327	0.1771	0.1780	0.1942	0.1579	0.1584	0.1859	0.0792	0.0691	0.0858	0.1066	0.1056	0.0784
51324	0.1803	0.1747	0.1702	0.4068	0.4011	0.1634	0.1017	0.0980	0.0702	0.1477	0.1435	0.0814
52165	0.2529	0.2694	0.2721	0.2648	0.2619	0.2743	0.1347	0.2002	0.2796	0.1534	0.1525	0.2819
52950	0.2311	0.2536	0.4703	0.2203	0.2194	0.2726	0.1732	0.1626	2.1436	0.2008	0.1997	0.1403
53795	0.2546	0.2694	0.2852	0.2019	0.2039	0.2634	0.2379	0.2506	0.3010	0.2340	0.2327	0.2765
55106	0.1103	0.1105	0.1192	0.1295	0.1255	0.1104	0.1047	0.1018	0.1063	0.1094	0.1099	0.0989
55830	0.1910	0.2017	0.2001	0.1848	0.1838	0.1985	0.1935	0.1830	0.2516	0.2494	0.2476	0.2491

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
60195	0.1675	0.1531	0.1701	0.1908	0.1921	0.1696	0.0764	0.1902	0.1404	0.0769	0.0765	0.1622
60639	0.2977	0.2864	0.2807	0.3123	0.3171	0.1915	0.2482	0.2419	1.4199	0.2393	0.2369	0.1748
63798	0.2270	0.2224	0.1917	0.3289	0.3272	0.2131	0.0861	0.0891	0.0896	0.1031	0.1019	0.1115
66207	0.1639	0.1596	0.4331	0.1618	0.1612	0.1531	0.1216	0.1201	1.2345	0.1440	0.1419	0.0980
66215	0.1390	0.1412	0.5379	0.1159	0.1159	0.1609	0.1174	0.1207	1.7633	0.1206	0.1197	0.1890
66287	0.2069	0.2062	0.2086	0.2025	0.2025	0.2041	0.1425	0.1408	0.1367	0.1584	0.1587	0.1304
66865	0.2186	0.2154	0.3126	0.2873	0.2886	0.1980	0.1204	0.1187	0.9954	0.1407	0.1395	0.1100
69925	0.6241	0.4391	0.7389	0.5263	0.5290	0.2546	0.2744	0.2369	2.7712	0.2905	0.2853	0.1737
71821	0.0953	0.0871	0.1695	0.1577	0.1555	0.0844	0.0644	0.0608	1.6072	0.0785	0.0778	0.0611
72202	0.0858	0.0803	0.0889	0.1818	0.1736	0.0885	0.0817	0.0754	0.0854	0.1237	0.1225	0.0853
72532	0.1643	0.1646	0.1659	0.1728	0.1713	0.1662	0.2647	0.2629	0.2611	0.3000	0.2984	0.2622
72603	0.0803	0.2313	0.0761	0.0943	0.0985	0.0776	0.0822	0.1206	0.0785	0.1053	0.1043	0.0649
72723	0.1926	0.1841	0.1905	0.2067	0.2046	0.1926	0.1448	0.1664	0.2969	0.1415	0.1410	0.3013
72869	0.2039	0.2033	0.2010	0.2039	0.2040	0.2023	0.1719	0.1931	0.2827	0.2272	0.2255	0.2880
73163	0.1319	0.1349	0.1665	0.1913	0.1949	0.1315	0.1517	0.1523	0.1581	0.2024	0.2015	0.1608
73186	0.0702	0.0694	0.0859	0.1367	0.1370	0.0922	0.0681	0.0679	0.0580	0.0797	0.0792	0.0914
74983	0.1730	0.1449	0.1150	0.1812	0.1808	0.1794	0.1474	0.1266	0.1417	0.1338	0.1319	0.2184
78546	0.2627	0.2027	0.6174	0.3787	0.3805	0.1702	0.1170	0.2537	2.3775	0.1789	0.1760	0.1166
79499	0.2173	0.2332	0.2645	0.1573	0.1568	0.2423	0.2245	0.2559	0.2913	0.1528	0.1511	0.2685
79769	0.2347	0.2373	0.2307	0.4336	0.4374	0.2369	0.1284	0.1300	0.1314	0.1996	0.1973	0.1353
83878	0.2887	0.2624	0.4346	0.4322	0.4334	0.2372	0.1629	0.1574	1.9957	0.1796	0.1793	0.1975
84052	0.2293	0.2300	0.2325	0.2830	0.2827	0.2323	0.0808	0.0832	0.1329	0.0834	0.0835	0.1439
93877	0.2173	0.2167	0.2492	0.2941	0.2899	0.2119	0.1684	0.1696	0.2075	0.1852	0.1837	0.1738
1605378	0.1229	0.1228	0.1248	0.1374	0.1361	0.1236	0.1035	0.1037	0.0991	0.1274	0.1266	0.0992
1605628	0.2128	0.2567	0.6030	0.1881	0.1897	0.1451	0.1460	0.1670	1.5917	0.1408	0.1398	0.1315
1607268	0.2219	0.2291	0.1641	0.3059	0.3084	0.1780	0.2341	0.2342	0.2024	0.2617	0.2593	0.2243
1608443	0.1613	0.1467	0.1363	0.1876	0.1854	0.1234	0.0873	0.0666	0.1732	0.1253	0.1219	0.1509
1609587	0.1364	0.1368	0.1352	0.1387	0.1376	0.1395	0.1148	0.1134	0.1495	0.1486	0.1474	0.1139
1612648	0.1807	0.1474	0.2089	0.2089	0.2089	0.1388	0.1908	0.1585	0.2442	0.2188	0.2167	0.1459
1614660	0.4574	0.4004	0.6023	0.5881	0.5838	0.3724	0.4233	0.4161	1.8022	0.5052	0.4977	0.4161
1617321	0.2534	0.2588	0.3003	0.2066	0.2072	0.2667	0.1057	0.1023	0.0975	0.1409	0.1397	0.0977
1620855	0.1859	0.1839	0.1776	0.2471	0.2374	0.1823	0.1469	0.1490	0.2420	0.2047	0.2015	0.2479
1621632	0.3065	0.3072	0.3059	0.2494	0.2509	0.3086	0.1690	0.1717	0.3077	0.1917	0.1908	0.3110
1621699	0.2632	0.2682	0.4541	0.2345	0.2345	0.2891	0.2580	0.2648	1.4636	0.2454	0.2427	0.3108
1623624	0.3007	0.2987	0.3148	0.3047	0.3027	0.2986	0.0696	0.0670	0.0730	0.0721	0.0725	0.0648
1624666	0.0506	0.0405	0.4172	0.1179	0.1245	0.0776	0.0507	0.0439	1.3346	0.0992	0.0983	0.0693
1633110	0.4694	0.2770	0.4884	0.4675	0.4692	0.2077	0.1558	0.3194	2.0136	0.2588	0.2535	0.1837
85207	0.2248	0.2088	0.2194	0.2404	0.2408	0.2080	0.1804	0.1804	0.2175	0.2061	0.2040	0.2021
85282	0.0888	0.0821	0.0976	0.1066	0.1069	0.0806	0.1071	0.0970	0.1335	0.1183	0.1173	0.0961
93804	0.1467	0.1450	0.3351	0.1378	0.1408	0.1418	0.1335	0.1378	1.6618	0.1452	0.1440	0.1667
60051	0.1620	0.1764	0.1704	0.2040	0.1982	0.2005	0.0990	0.1038	0.1546	0.1415	0.1395	0.1837
44248	0.2077	0.2035	0.2104	0.2588	0.2549	0.1873	0.1835	0.1809	0.1963	0.1840	0.1818	0.1720
69982	0.2046	0.1978	0.2440	0.2144	0.2125	0.2456	0.2175	0.1922	0.2636	0.2601	0.2600	0.2627
76106	0.1112	0.1232	0.1034	0.1087	0.1083	0.1277	0.1251	0.1207	0.1276	0.1619	0.1608	0.1137
1601410	0.1486	0.1503	0.1741	0.1907	0.1927	0.1746	0.1420	0.1436	0.1620	0.2454	0.2434	0.1594
11869	0.1227	0.1260	0.2429	0.1617	0.1657	0.1602	0.0865	0.0855	2.2413	0.1041	0.1039	0.1472
78511	0.1640	0.1633	0.1883	0.2446	0.2482	0.1698	0.1684	0.1684	0.2077	0.1998	0.1982	0.1754
82421	0.1497	0.1627	0.1583	0.1605	0.1600	0.1590	0.1141	0.1163	0.1320	0.1241	0.1219	0.1245
15618	0.0974	0.1169	0.5780	0.0683	0.0649	0.1007	0.0840	0.1031	2.0493	0.0758	0.0737	0.0894
36953	0.1188	0.0947	0.0873	0.1526	0.1538	0.0861	0.0955	0.1724	0.0758	0.1131	0.1127	0.0734
66666	0.1431	0.1355	0.1721	0.3002	0.3103	0.1741	0.1394	0.1280	0.1565	0.2411	0.2386	0.1782
94870	0.0943	0.0937	0.1103	0.1031	0.1031	0.1009	0.0694	0.0715	0.1045	0.1096	0.1104	0.0890
91836	0.2429	0.2740	0.4799	0.4809	0.4809	0.3285	0.2514	0.2715	2.1027	0.4037	0.4016	0.3638
22349	0.0797	0.2870	0.2259	0.0758	0.0760	0.1463	0.0828	0.2444	0.2304	0.0874	0.0876	0.1505

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
83648	0.1549	0.1633	0.1901	0.1892	0.1913	0.2234	0.1733	0.1485	0.2200	0.2719	0.2695	0.2799
44983	0.0831	0.0941	0.1520	0.1154	0.1134	0.0939	0.0686	0.2590	0.1395	0.0850	0.0843	0.0734
71801	0.2000	0.1995	0.1903	0.1918	0.1923	0.1986	0.0814	0.1299	0.1703	0.1124	0.1116	0.1242
58776	0.0825	0.0805	0.0819	0.1541	0.1472	0.0796	0.0850	0.0843	0.0602	0.1196	0.1177	0.0911
74359	0.1024	0.1025	0.1050	0.1049	0.1061	0.1011	0.0811	0.0846	0.1021	0.0655	0.0649	0.0981
1603059	0.1699	0.1671	0.4399	0.2413	0.2373	0.1544	0.1110	0.1106	1.8106	0.1042	0.1025	0.1204
76500	0.1580	0.1573	0.1546	0.1844	0.1820	0.1530	0.0895	0.0890	0.1022	0.1116	0.1115	0.1062
66306	0.1375	0.2400	0.1455	0.1480	0.1492	0.1951	0.1114	0.2091	0.1341	0.1455	0.1447	0.1920
1626293	0.2078	0.2372	0.2690	0.2117	0.2106	0.3123	0.1583	0.1646	0.2550	0.1964	0.1944	0.3011
1629646	0.5209	0.5175	0.4529	0.5404	0.5434	0.2433	0.2680	0.2689	1.6918	0.3803	0.3755	0.2910
90928	0.1954	0.2071	0.6585	0.3831	0.3884	0.2564	0.2028	0.2279	2.3289	0.2624	0.2611	0.2509
average	**0.1878**	**0.1937**	**0.2505**	**0.2267**	**0.2267**	**0.1830**	**0.1353**	**0.1547**	**0.6251**	**0.1648**	**0.1634**	**0.1642**

Table A.13. Out-of-Sample Predictive Performance Results (MAE)

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
5106	0.1388	0.1504	0.1408	0.1385	0.1385	0.1500	0.1299	0.1287	0.2026	0.1620	0.1592	0.1808
72528	0.1193	0.1181	0.1301	0.1147	0.1145	0.1148	0.0548	0.0520	0.1108	0.0766	0.0758	0.0637
73154	0.1582	0.1620	0.1679	0.2086	0.2028	0.1582	0.1243	0.1533	0.1112	0.1485	0.1472	0.1146
80961	0.1334	0.1160	0.1185	0.2369	0.2345	0.1181	0.1197	0.1600	0.0674	0.1569	0.1532	0.1043
29336	0.1350	0.1379	0.1588	0.1324	0.1324	0.1391	0.0762	0.0578	0.0623	0.0891	0.0877	0.0530
23222	0.1594	0.1615	0.1824	0.3256	0.3370	0.1704	0.1486	0.1414	0.2459	0.3034	0.2990	0.2125
30938	0.1370	0.1358	0.2074	0.1287	0.1277	0.1299	0.0730	0.0709	1.3207	0.0857	0.0854	0.0666
71480	0.1226	0.1169	0.1131	0.1883	0.1878	0.1187	0.0678	0.0430	0.0481	0.0869	0.0857	0.0599
88292	0.1514	0.1007	0.1121	0.2325	0.2374	0.0961	0.0539	0.0391	0.0614	0.0805	0.0783	0.0382
1618693	0.1390	0.1376	0.1439	0.1766	0.1793	0.1370	0.0167	0.0569	0.1120	0.0128	0.0123	0.1399
56737	0.1299	0.1503	0.1545	0.1764	0.1832	0.1422	0.1187	0.1744	0.2619	0.1511	0.1494	0.2226
71478	0.1475	0.1443	0.2558	0.1066	0.1082	0.1670	0.1027	0.1038	1.1529	0.1269	0.1257	0.2074
49049	0.2546	0.1941	0.2227	0.2434	0.2448	0.2019	0.1086	0.2203	1.4304	0.1260	0.1245	0.2036
72526	0.1466	0.1448	0.2336	0.1610	0.1587	0.1415	0.0870	0.0950	1.2088	0.0974	0.0952	0.1810
32391	0.1284	0.1483	0.1379	0.2130	0.2152	0.1301	0.0562	0.2063	0.0709	0.0668	0.0668	0.0881
72527	0.1241	0.1302	0.2050	0.1246	0.1240	0.1486	0.0890	0.0750	0.9763	0.0877	0.0873	0.1152
69278	0.1568	0.1604	0.1843	0.1654	0.1623	0.1661	0.1155	0.1109	0.0979	0.1345	0.1340	0.1071
28245	0.1490	0.1496	0.1651	0.1470	0.1469	0.1462	0.1653	0.1635	0.3167	0.1821	0.1800	0.2298
30814	0.1096	0.1106	0.0864	0.1751	0.1741	0.1125	0.1463	0.1500	0.1346	0.1799	0.1771	0.1852
16311	0.1285	0.1289	0.1285	0.1448	0.1451	0.1286	0.0526	0.0505	0.0587	0.0609	0.0604	0.0484
2753	0.2481	0.2237	0.2895	0.2995	0.2995	0.1506	0.1510	0.1411	1.3232	0.1968	0.1941	0.1083
3150	0.0850	0.0813	0.0917	0.0809	0.0795	0.0981	0.0830	0.0902	0.1184	0.0862	0.0850	0.1315
4171	0.1126	0.2649	0.1608	0.2198	0.2207	0.1570	0.1426	0.3917	0.3631	0.2032	0.2014	0.3357
5107	0.1483	0.1655	0.1956	0.1624	0.1576	0.1671	0.0871	0.2124	0.2008	0.0873	0.0865	0.1422
5452	0.1319	0.1400	0.1490	0.1740	0.1740	0.1322	0.0543	0.0492	0.0575	0.0640	0.0636	0.0458
5603	0.2057	0.2052	0.2600	0.2043	0.2063	0.2063	0.2486	0.2485	1.3256	0.2719	0.2705	0.2547
6712	0.1523	0.5813	0.4183	0.1704	0.1763	0.2169	0.1746	0.3378	1.0359	0.1884	0.1877	0.2555
7297	0.2477	0.2409	0.3115	0.3647	0.3657	0.2697	0.0653	0.0582	1.6321	0.1125	0.1111	0.1721
9474	0.1152	0.1138	0.1157	0.1363	0.1350	0.1149	0.1307	0.1129	0.1354	0.1893	0.1869	0.1344
11626	0.0702	0.0776	0.1225	0.0644	0.0647	0.0987	0.0663	0.0712	0.1181	0.0803	0.0797	0.0929
12043	0.0763	0.0820	0.0930	0.0835	0.0849	0.1085	0.0862	0.0874	0.0845	0.1303	0.1296	0.0932
17184	0.1223	0.1181	0.1288	0.1077	0.1072	0.1157	0.0987	0.0868	0.0754	0.0992	0.0995	0.0740
18201	0.0872	0.0986	0.3974	0.1791	0.1756	0.2151	0.0939	0.0818	1.7158	0.1535	0.1513	0.1984
20131	0.1215	0.1447	0.1691	0.2908	0.2894	0.0839	0.1122	0.1190	1.0248	0.1619	0.1593	0.0943
21148	0.2165	0.2123	0.2402	0.3064	0.3093	0.2189	0.1316	0.1288	0.1498	0.1787	0.1765	0.1323
21509	0.2284	0.2240	0.2005	0.3655	0.3688	0.2202	0.1185	0.1136	0.1256	0.1665	0.1651	0.1148
22510	0.0689	0.0669	0.1830	0.1934	0.1934	0.0930	0.0527	0.0434	1.4000	0.1286	0.1268	0.0889
22809	0.1438	0.1120	0.1158	0.2079	0.2053	0.1091	0.1420	0.2382	0.1610	0.1667	0.1634	0.1434
26561	0.1324	0.1583	0.1201	0.1635	0.1611	0.1261	0.0749	0.2190	0.0632	0.0789	0.0775	0.0903
28977	0.2023	0.1968	0.2455	0.1674	0.1678	0.2054	0.1120	0.1089	0.2488	0.1254	0.1242	0.1816
44981	0.0776	0.1108	0.1427	0.1287	0.1228	0.1149	0.0555	0.1645	0.1403	0.0744	0.0730	0.1096
46254	0.0867	0.0858	0.0932	0.1175	0.1211	0.0997	0.0461	0.0473	0.0673	0.0585	0.0581	0.0798
49063	0.1787	0.1734	0.1760	0.2404	0.2315	0.1714	0.1169	0.1106	0.0950	0.1750	0.1708	0.1050
49232	0.1580	0.1566	0.1497	0.2340	0.2284	0.1467	0.1863	0.1918	0.2781	0.2474	0.2439	0.2642
49435	0.1308	0.1365	0.1393	0.2265	0.2260	0.1362	0.1144	0.1916	0.1319	0.1460	0.1455	0.1373
50327	0.1527	0.1499	0.1596	0.1363	0.1364	0.1539	0.0661	0.0575	0.0741	0.0938	0.0927	0.0667
51324	0.1472	0.1428	0.1308	0.3758	0.3700	0.1286	0.0863	0.0828	0.0529	0.1350	0.1306	0.0645
52165	0.2024	0.2164	0.2192	0.2130	0.2105	0.2209	0.1104	0.1794	0.2646	0.1329	0.1317	0.2671
52950	0.1988	0.2121	0.3564	0.1908	0.1894	0.2352	0.1455	0.1382	1.5510	0.1733	0.1715	0.1274
53795	0.1928	0.2079	0.2248	0.1538	0.1550	0.2026	0.2005	0.2108	0.2716	0.2035	0.2018	0.2459
55106	0.0885	0.0888	0.0965	0.1041	0.1009	0.0891	0.0864	0.0837	0.0886	0.0916	0.0922	0.0836
55830	0.1444	0.1604	0.1489	0.1385	0.1392	0.1470	0.1619	0.1550	0.2186	0.2185	0.2162	0.2160

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
60195	0.1365	0.1272	0.1523	0.1511	0.1512	0.1546	0.0595	0.1715	0.1277	0.0626	0.0620	0.1508
60639	0.2549	0.2456	0.2107	0.2813	0.2868	0.1614	0.2084	0.2039	0.9687	0.2109	0.2080	0.1509
63798	0.1888	0.1823	0.1416	0.3004	0.2984	0.1683	0.0676	0.0713	0.0676	0.0886	0.0876	0.0935
66207	0.1499	0.1477	0.3800	0.1506	0.1499	0.1051	0.1009	0.0993	1.0043	0.1295	0.1266	0.0628
66215	0.1105	0.1122	0.4169	0.0985	0.0985	0.1299	0.0986	0.1022	1.2825	0.1055	0.1044	0.1743
66287	0.1765	0.1760	0.1788	0.1629	0.1629	0.1745	0.1122	0.1106	0.1077	0.1309	0.1310	0.1015
66865	0.1860	0.1830	0.2728	0.2537	0.2550	0.1685	0.0978	0.0965	0.7060	0.1196	0.1180	0.0907
69925	0.5817	0.4080	0.4661	0.4943	0.4973	0.1864	0.2414	0.2085	2.0521	0.2667	0.2611	0.1388
71821	0.0808	0.0736	0.1117	0.1425	0.1400	0.0736	0.0516	0.0489	1.0674	0.0674	0.0666	0.0523
72202	0.0686	0.0644	0.0752	0.1591	0.1494	0.0760	0.0675	0.0613	0.0758	0.1071	0.1056	0.0750
72532	0.1238	0.1228	0.1230	0.1255	0.1250	0.1231	0.2253	0.2237	0.2264	0.2647	0.2624	0.2274
72603	0.0683	0.2003	0.0636	0.0801	0.0847	0.0618	0.0707	0.1091	0.0674	0.0945	0.0932	0.0537
72723	0.1401	0.1352	0.1395	0.1819	0.1797	0.1409	0.1239	0.1406	0.2851	0.1251	0.1246	0.2897
72869	0.1675	0.1678	0.1681	0.1664	0.1662	0.1673	0.1534	0.1764	0.2570	0.2035	0.2013	0.2621
73163	0.1056	0.1085	0.1448	0.1666	0.1698	0.1100	0.1191	0.1199	0.1379	0.1651	0.1639	0.1491
73186	0.0543	0.0537	0.0673	0.1123	0.1124	0.0711	0.0566	0.0566	0.0498	0.0680	0.0676	0.0750
74983	0.1429	0.1107	0.0875	0.1543	0.1538	0.1521	0.1238	0.1060	0.1246	0.1194	0.1172	0.2033
78546	0.2307	0.1644	0.3869	0.3531	0.3548	0.1422	0.1032	0.2370	1.7432	0.1618	0.1585	0.1044
79499	0.1890	0.2101	0.2472	0.1358	0.1354	0.2232	0.2030	0.2399	0.2805	0.1352	0.1335	0.2569
79769	0.1960	0.1985	0.1998	0.3901	0.3944	0.2038	0.1053	0.1073	0.1145	0.1733	0.1707	0.1172
83878	0.2194	0.1964	0.3047	0.3659	0.3673	0.1983	0.1328	0.1300	1.5206	0.1477	0.1472	0.1767
84052	0.1799	0.1817	0.1875	0.2248	0.2246	0.1879	0.0656	0.0674	0.1173	0.0718	0.0718	0.1293
93877	0.1800	0.1790	0.2105	0.2598	0.2549	0.1754	0.1325	0.1335	0.1755	0.1504	0.1489	0.1388
1605378	0.0912	0.0911	0.0944	0.1045	0.1032	0.0937	0.0815	0.0818	0.0849	0.1052	0.1041	0.0840
1605628	0.1888	0.2279	0.4451	0.1646	0.1664	0.1153	0.1248	0.1437	1.1039	0.1238	0.1226	0.1121
1607268	0.1830	0.1885	0.1341	0.2717	0.2742	0.1435	0.2025	0.2028	0.1704	0.2350	0.2321	0.1929
1608443	0.1201	0.1111	0.1014	0.1524	0.1495	0.0982	0.0714	0.0582	0.1600	0.1093	0.1059	0.1401
1609587	0.1169	0.1171	0.1169	0.1190	0.1183	0.1191	0.0971	0.0959	0.1297	0.1317	0.1303	0.0989
1612648	0.1527	0.1220	0.1788	0.1786	0.1786	0.1159	0.1516	0.1310	0.2156	0.1847	0.1822	0.1386
1614660	0.3924	0.3424	0.4808	0.5196	0.5146	0.3292	0.3724	0.3700	1.3050	0.4592	0.4508	0.3815
1617321	0.2123	0.2152	0.2403	0.1787	0.1789	0.2145	0.0853	0.0820	0.0717	0.1195	0.1180	0.0798
1620855	0.1533	0.1514	0.1470	0.2148	0.2045	0.1505	0.1233	0.1277	0.2256	0.1793	0.1756	0.2319
1621632	0.2484	0.2490	0.2472	0.1963	0.1975	0.2499	0.1383	0.1434	0.2854	0.1650	0.1639	0.2890
1621699	0.2233	0.2280	0.3661	0.1971	0.1971	0.2504	0.2283	0.2351	1.1856	0.2254	0.2220	0.2851
1623624	0.1758	0.1740	0.1897	0.1803	0.1782	0.1742	0.0566	0.0546	0.0588	0.0605	0.0607	0.0538
1624666	0.0396	0.0319	0.2850	0.1038	0.1095	0.0625	0.0401	0.0347	0.8489	0.0862	0.0849	0.0590
1633110	0.4440	0.2199	0.2950	0.4449	0.4465	0.1816	0.1378	0.2925	1.4003	0.2372	0.2316	0.1690
85207	0.1697	0.1599	0.1658	0.1779	0.1781	0.1597	0.1508	0.1560	0.1937	0.1805	0.1781	0.1826
85282	0.0745	0.0682	0.0810	0.0868	0.0869	0.0676	0.0880	0.0795	0.1175	0.0997	0.0985	0.0831
93804	0.1252	0.1238	0.2188	0.1184	0.1210	0.1218	0.1136	0.1182	1.2461	0.1263	0.1248	0.1483
60051	0.1405	0.1567	0.1499	0.1875	0.1810	0.1834	0.0820	0.0836	0.1301	0.1264	0.1242	0.1639
44248	0.1749	0.1716	0.1750	0.2212	0.2169	0.1579	0.1563	0.1548	0.1764	0.1633	0.1609	0.1552
69982	0.1727	0.1654	0.2074	0.1796	0.1771	0.2089	0.1852	0.1627	0.2249	0.2238	0.2235	0.2248
76106	0.0955	0.1100	0.0856	0.0754	0.0751	0.1168	0.1055	0.1054	0.1168	0.1398	0.1384	0.1041
1601410	0.1211	0.1237	0.1471	0.1679	0.1696	0.1445	0.1159	0.1186	0.1296	0.2193	0.2166	0.1269
11869	0.0996	0.1027	0.1693	0.1431	0.1472	0.1381	0.0721	0.0719	1.5150	0.0902	0.0897	0.1364
78511	0.1446	0.1442	0.1528	0.2008	0.2044	0.1439	0.1350	0.1354	0.1758	0.1713	0.1693	0.1446
82421	0.1083	0.1170	0.1225	0.1232	0.1191	0.1167	0.0842	0.0892	0.1028	0.0988	0.0964	0.0929
15618	0.0743	0.0971	0.4278	0.0626	0.0586	0.0801	0.0618	0.0849	1.4891	0.0590	0.0572	0.0720
36953	0.0991	0.0766	0.0717	0.1312	0.1321	0.0720	0.0786	0.1535	0.0578	0.0970	0.0965	0.0586
66666	0.1176	0.1150	0.1422	0.2653	0.2754	0.1394	0.1125	0.1080	0.1335	0.2099	0.2075	0.1468
94870	0.0742	0.0739	0.0908	0.0804	0.0804	0.0798	0.0569	0.0589	0.0836	0.0934	0.0944	0.0680
91836	0.2043	0.2347	0.2863	0.4233	0.4233	0.2500	0.2142	0.2318	1.5449	0.3547	0.3518	0.3183
22349	0.0604	0.2607	0.2154	0.0571	0.0570	0.1295	0.0656	0.2220	0.2205	0.0724	0.0730	0.1350

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
83648	0.1254	0.1297	0.1452	0.1553	0.1590	0.1727	0.1439	0.1218	0.1842	0.2316	0.2281	0.2426
44983	0.0643	0.0741	0.1273	0.0900	0.0882	0.0730	0.0545	0.2349	0.1255	0.0718	0.0710	0.0582
71801	0.1606	0.1600	0.1622	0.1615	0.1612	0.1600	0.0677	0.1157	0.1586	0.1002	0.0996	0.1083
58776	0.0656	0.0645	0.0671	0.1370	0.1292	0.0653	0.0731	0.0724	0.0507	0.1073	0.1052	0.0790
74359	0.0936	0.0938	0.0962	0.0965	0.0975	0.0926	0.0708	0.0752	0.0933	0.0550	0.0546	0.0893
1603059	0.1476	0.1451	0.3081	0.2203	0.2158	0.1352	0.0919	0.0922	1.2139	0.0938	0.0921	0.1106
76500	0.1289	0.1285	0.1295	0.1429	0.1417	0.1284	0.0709	0.0708	0.0961	0.0886	0.0883	0.0991
66306	0.1138	0.2113	0.1216	0.1194	0.1205	0.1638	0.0894	0.1875	0.1106	0.1266	0.1258	0.1768
1626293	0.1744	0.2030	0.2311	0.1764	0.1753	0.2773	0.1337	0.1397	0.2263	0.1711	0.1688	0.2763
1629646	0.4732	0.4694	0.3537	0.4987	0.5016	0.1946	0.2310	0.2340	1.2511	0.3352	0.3296	0.2561
90928	0.1591	0.1656	0.3891	0.3147	0.3204	0.2167	0.1675	0.2007	1.6305	0.2275	0.2256	0.2180
average	**0.1537**	**0.1582**	**0.1914**	**0.1922**	**0.1921**	**0.1488**	**0.1129**	**0.1330**	**0.4722**	**0.1433**	**0.1415**	**0.1448**

Table A.14. Out-of-Sample Predictive Performance Results (SMAPE)

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
5106	0.1766	0.1931	0.1798	0.1743	0.1743	0.1930	0.2760	0.2736	0.3945	0.3297	0.3256	0.3695
72528	0.1329	0.1315	0.1459	0.1276	0.1274	0.1277	0.0532	0.0507	0.1179	0.0732	0.0724	0.0629
73154	0.1621	0.1650	0.1731	0.2244	0.2166	0.1620	0.1752	0.2123	0.1638	0.1989	0.1972	0.1663
80961	0.2734	0.2521	0.2606	0.3752	0.3733	0.2562	0.3223	0.3639	0.2116	0.3978	0.3910	0.2852
29336	0.1748	0.1790	0.2128	0.1714	0.1714	0.1807	0.1459	0.1070	0.1164	0.1742	0.1718	0.0983
23222	0.2443	0.2464	0.2728	0.3735	0.3800	0.2504	0.1780	0.1696	0.2985	0.2912	0.2879	0.2402
30938	0.2116	0.2096	0.3062	0.1986	0.1969	0.2007	0.0941	0.0918	0.6573	0.1085	0.1081	0.0869
71480	0.2421	0.2184	0.2166	0.4362	0.4344	0.2219	0.1510	0.0855	0.0960	0.2038	0.2004	0.1142
88292	0.2337	0.1732	0.1994	0.3145	0.3185	0.1664	0.0906	0.0724	0.1225	0.1248	0.1219	0.0722
1618693	0.3596	0.3567	0.3800	0.5184	0.5321	0.3534	0.1051	0.3309	0.5110	0.0650	0.0616	0.5640
56737	0.2478	0.2734	0.3048	0.2912	0.2973	0.2744	0.1270	0.1998	0.3939	0.1577	0.1560	0.3113
71478	0.2036	0.1982	0.3479	0.1380	0.1402	0.2344	0.1173	0.1181	0.6809	0.1362	0.1349	0.2723
49049	0.2270	0.2047	0.2429	0.2214	0.2222	0.2182	0.1010	0.2313	0.6404	0.1164	0.1151	0.2135
72526	0.2199	0.2164	0.3429	0.2480	0.2433	0.2105	0.1095	0.1207	0.6574	0.1238	0.1205	0.2547
32391	0.1861	0.2189	0.1972	0.2677	0.2695	0.1883	0.0649	0.2256	0.0824	0.0747	0.0747	0.1045
72527	0.1701	0.1790	0.2683	0.1696	0.1690	0.2099	0.1352	0.1185	0.6345	0.1332	0.1326	0.1845
69278	0.1886	0.1938	0.2309	0.2017	0.1971	0.2026	0.2243	0.2171	0.2016	0.2510	0.2506	0.2141
28245	0.1938	0.1944	0.2199	0.1919	0.1918	0.1916	0.1442	0.1426	0.3207	0.1596	0.1577	0.2082
30814	0.2328	0.2344	0.2034	0.3093	0.3083	0.2378	0.3832	0.3931	0.3877	0.4262	0.4223	0.4460
16311	0.1406	0.1411	0.1404	0.1609	0.1613	0.1407	0.0601	0.0574	0.0669	0.0704	0.0698	0.0546
2753	0.4038	0.3469	0.4658	0.5378	0.5378	0.2197	0.2726	0.2444	0.6888	0.3906	0.3842	0.1690
3150	0.0887	0.0846	0.0963	0.0842	0.0827	0.1038	0.0825	0.0921	0.1206	0.0836	0.0825	0.1362
4171	0.1246	0.3284	0.1692	0.2224	0.2230	0.1649	0.1067	0.3189	0.3161	0.1408	0.1395	0.2837
5107	0.1666	0.1907	0.2329	0.1764	0.1726	0.1928	0.0880	0.2420	0.2439	0.0864	0.0856	0.1588
5452	0.1506	0.1612	0.1738	0.2106	0.2106	0.1509	0.0746	0.0657	0.0773	0.0891	0.0884	0.0605
5603	0.2472	0.2467	0.3205	0.2459	0.2487	0.2486	0.2395	0.2386	0.6542	0.2746	0.2727	0.2473
6712	0.1459	0.8333	0.5242	0.1671	0.1744	0.2271	0.1642	0.3536	0.6187	0.1787	0.1780	0.2624
7297	0.2122	0.2084	0.2816	0.3603	0.3618	0.2414	0.1079	0.0947	0.7452	0.1851	0.1826	0.2314
9474	0.1428	0.1418	0.1433	0.1652	0.1640	0.1424	0.1884	0.1688	0.1959	0.2424	0.2399	0.1948
11626	0.0779	0.0865	0.1442	0.0707	0.0710	0.1127	0.0769	0.0826	0.1423	0.0939	0.0933	0.1088
12043	0.0957	0.1031	0.1175	0.1004	0.1018	0.1390	0.1205	0.1226	0.1190	0.1834	0.1825	0.1314
17184	0.1908	0.1824	0.2026	0.1643	0.1635	0.1781	0.1837	0.1592	0.1382	0.1808	0.1816	0.1348
18201	0.1314	0.1435	0.3911	0.3308	0.3219	0.2575	0.1561	0.1343	0.6651	0.2734	0.2691	0.2403
20131	0.1788	0.2037	0.2362	0.3288	0.3278	0.1331	0.1835	0.1913	0.6422	0.2340	0.2323	0.1650
21148	0.1921	0.1872	0.2219	0.3115	0.3160	0.1981	0.1355	0.1316	0.1573	0.2019	0.1991	0.1364
21509	0.2070	0.2017	0.1793	0.4057	0.4113	0.2006	0.1349	0.1269	0.1407	0.2198	0.2173	0.1287
22510	0.0759	0.0730	0.1623	0.1820	0.1820	0.0989	0.0509	0.0419	0.6960	0.1141	0.1126	0.0892
22809	0.1483	0.1489	0.1256	0.1969	0.1950	0.1196	0.1659	0.3530	0.1876	0.1836	0.1804	0.1716
26561	0.1221	0.1571	0.1105	0.1470	0.1453	0.1165	0.0826	0.2358	0.0712	0.0856	0.0844	0.0968
28977	0.1957	0.1896	0.2495	0.1586	0.1590	0.1993	0.1060	0.1023	0.2657	0.1176	0.1163	0.1784
44981	0.0947	0.1264	0.1563	0.1717	0.1624	0.1315	0.0665	0.2016	0.1480	0.0930	0.0911	0.1204
46254	0.0794	0.0788	0.0860	0.1123	0.1164	0.0917	0.0436	0.0448	0.0636	0.0572	0.0569	0.0749
49063	0.1842	0.1808	0.1847	0.2193	0.2136	0.1794	0.1477	0.1429	0.1289	0.2016	0.1978	0.1412
49232	0.2063	0.2053	0.1997	0.2567	0.2531	0.1978	0.4035	0.4234	0.4892	0.5475	0.5505	0.4798
49435	0.1717	0.1809	0.1898	0.3770	0.3756	0.1853	0.1588	0.2611	0.1924	0.2297	0.2282	0.2003
50327	0.2615	0.2547	0.2628	0.2496	0.2500	0.2579	0.0963	0.0818	0.1012	0.1442	0.1423	0.0925
51324	0.1544	0.1510	0.1411	0.3043	0.3013	0.1385	0.0949	0.0923	0.0642	0.1318	0.1283	0.0768
52165	0.1995	0.2165	0.2201	0.2123	0.2092	0.2223	0.0929	0.1572	0.2602	0.1133	0.1121	0.2635
52950	0.1387	0.1491	0.2269	0.1331	0.1321	0.1667	0.1262	0.1289	0.6019	0.1470	0.1457	0.1144
53795	0.2483	0.2768	0.3108	0.1862	0.1878	0.2666	0.2422	0.2579	0.3643	0.2480	0.2456	0.3137
55106	0.0595	0.0597	0.0653	0.0709	0.0686	0.0599	0.0592	0.0573	0.0609	0.0632	0.0637	0.0573
55830	0.1155	0.1313	0.1195	0.1106	0.1111	0.1178	0.1141	0.1110	0.1565	0.1459	0.1444	0.1543

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
60195	0.1944	0.1854	0.2134	0.2070	0.2068	0.2177	0.0570	0.1630	0.1324	0.0587	0.0581	0.1607
60639	0.3442	0.3379	0.3563	0.3668	0.3709	0.2680	0.2921	0.2887	0.6087	0.2963	0.2937	0.2414
63798	0.2096	0.2045	0.1711	0.2864	0.2852	0.1932	0.0687	0.0716	0.0691	0.0876	0.0872	0.0918
66207	0.1291	0.1276	0.3895	0.1297	0.1293	0.0939	0.0928	0.0916	0.6153	0.1177	0.1152	0.0592
66215	0.1054	0.1072	0.4361	0.0933	0.0932	0.1271	0.0900	0.0934	0.6124	0.0968	0.0957	0.1678
66287	0.1648	0.1644	0.1665	0.1538	0.1538	0.1633	0.1080	0.1062	0.1008	0.1286	0.1287	0.0958
66865	0.3084	0.3018	0.4836	0.4874	0.4914	0.2719	0.2067	0.2020	0.6570	0.2737	0.2695	0.1816
69925	0.2424	0.1858	0.2701	0.2154	0.2164	0.1018	0.1165	0.1037	0.5774	0.1281	0.1256	0.0784
71821	0.1129	0.1018	0.1766	0.2237	0.2187	0.1026	0.0741	0.0696	0.5883	0.1022	0.1007	0.0748
72202	0.1812	0.1731	0.1949	0.3081	0.2965	0.1961	0.1778	0.1666	0.1976	0.2394	0.2372	0.1950
72532	0.1672	0.1658	0.1661	0.1697	0.1691	0.1663	0.2478	0.2436	0.2437	0.3064	0.3033	0.2450
72603	0.0487	0.1249	0.0453	0.0565	0.0596	0.0441	0.0513	0.0827	0.0496	0.0665	0.0656	0.0398
72723	0.2358	0.2255	0.2339	0.2753	0.2732	0.2367	0.1382	0.1552	0.3675	0.1367	0.1361	0.3759
72869	0.1902	0.1905	0.1913	0.1890	0.1888	0.1902	0.2393	0.2705	0.3500	0.3538	0.3517	0.3538
73163	0.1028	0.1061	0.1481	0.1743	0.1783	0.1102	0.1254	0.1264	0.1507	0.1894	0.1877	0.1626
73186	0.0529	0.0524	0.0658	0.1164	0.1165	0.0690	0.0585	0.0585	0.0519	0.0723	0.0717	0.0755
74983	0.1568	0.1146	0.0872	0.1720	0.1713	0.1689	0.1232	0.1029	0.1205	0.1177	0.1153	0.2180
78546	0.1430	0.1087	0.2159	0.2008	0.2015	0.0954	0.0640	0.1895	0.5766	0.0980	0.0963	0.0663
79499	0.2371	0.2566	0.2876	0.1879	0.1875	0.2683	0.2654	0.3001	0.3336	0.2000	0.1980	0.3151
79769	0.3488	0.3506	0.3547	0.4670	0.4693	0.3569	0.1480	0.1505	0.1601	0.2117	0.2094	0.1629
83878	0.1671	0.1456	0.2280	0.3358	0.3377	0.1505	0.1231	0.1203	0.6443	0.1423	0.1416	0.1626
84052	0.2135	0.2155	0.2225	0.2766	0.2762	0.2223	0.1012	0.1030	0.1620	0.1146	0.1146	0.1746
93877	0.2085	0.2069	0.2567	0.3471	0.3374	0.2018	0.1521	0.1532	0.2192	0.1831	0.1806	0.1600
1605378	0.0841	0.0840	0.0872	0.0972	0.0960	0.0865	0.0802	0.0805	0.0839	0.1002	0.0992	0.0830
1605628	0.3134	0.4106	0.4809	0.2600	0.2638	0.1459	0.1864	0.2238	0.6172	0.1850	0.1829	0.1385
1607268	0.1532	0.1567	0.1203	0.2081	0.2095	0.1271	0.1921	0.1923	0.1703	0.2107	0.2088	0.1868
1608443	0.1853	0.1768	0.1658	0.2185	0.2157	0.1633	0.1471	0.1137	0.2677	0.2182	0.2128	0.2448
1609587	0.1194	0.1196	0.1199	0.1221	0.1214	0.1214	0.0841	0.0827	0.1145	0.1195	0.1182	0.0850
1612648	0.3008	0.2321	0.3670	0.3661	0.3661	0.2178	0.2495	0.2240	0.4000	0.3028	0.2984	0.2350
1614660	0.4334	0.3631	0.4759	0.7127	0.6955	0.3539	0.3639	0.3508	0.6200	0.5148	0.5012	0.3627
1617321	0.1985	0.2020	0.2325	0.1636	0.1638	0.2012	0.1114	0.1068	0.0939	0.1562	0.1545	0.1046
1620855	0.1561	0.1546	0.1512	0.2011	0.1940	0.1539	0.1460	0.1510	0.2459	0.2205	0.2169	0.2508
1621632	0.1897	0.1902	0.1884	0.1423	0.1433	0.1910	0.0956	0.1000	0.2209	0.1103	0.1095	0.2244
1621699	0.2430	0.2498	0.4735	0.2070	0.2070	0.2837	0.2403	0.2486	0.6838	0.2370	0.2332	0.3163
1623624	0.1829	0.1801	0.2054	0.1901	0.1868	0.1804	0.0955	0.0919	0.0996	0.1028	0.1030	0.0907
1624666	0.0642	0.0516	0.4081	0.1937	0.2072	0.1011	0.0647	0.0554	0.5475	0.1570	0.1545	0.0932
1633110	0.3232	0.1905	0.2591	0.3239	0.3246	0.1702	0.1192	0.2948	0.6299	0.2110	0.2068	0.1594
85207	0.2437	0.2267	0.2366	0.2584	0.2587	0.2263	0.2151	0.2164	0.2740	0.2575	0.2539	0.2543
85282	0.1204	0.1096	0.1319	0.1424	0.1427	0.1087	0.1228	0.1107	0.1715	0.1403	0.1385	0.1142
93804	0.2118	0.2101	0.3048	0.2039	0.2070	0.2079	0.2137	0.2224	0.7231	0.2312	0.2292	0.2665
60051	0.0945	0.1070	0.1014	0.1310	0.1258	0.1277	0.0534	0.0552	0.0883	0.0805	0.0790	0.1149
44248	0.1615	0.1592	0.1614	0.1922	0.1894	0.1493	0.1436	0.1430	0.1603	0.1494	0.1475	0.1450
69982	0.1394	0.1341	0.1695	0.1451	0.1430	0.1708	0.1323	0.1190	0.1708	0.1536	0.1532	0.1705
76106	0.1993	0.2217	0.1828	0.1615	0.1606	0.2343	0.1831	0.1685	0.2038	0.2553	0.2525	0.1804
1601410	0.1634	0.1674	0.1968	0.2268	0.2293	0.1927	0.1437	0.1491	0.1658	0.3172	0.3129	0.1619
11869	0.1075	0.1127	0.1825	0.1365	0.1397	0.1587	0.0780	0.0787	0.6181	0.0899	0.0897	0.1609
78511	0.1627	0.1623	0.1667	0.2412	0.2470	0.1602	0.1774	0.1777	0.2104	0.2203	0.2180	0.1847
82421	0.0868	0.0940	0.0986	0.0991	0.0958	0.0938	0.0740	0.0783	0.0899	0.0870	0.0851	0.0819
15618	0.0919	0.1152	0.4348	0.0815	0.0762	0.0981	0.0762	0.1002	0.6575	0.0744	0.0721	0.0872
36953	0.1518	0.1293	0.1163	0.1879	0.1889	0.1168	0.1155	0.2634	0.0878	0.1402	0.1397	0.0891
66666	0.0918	0.0893	0.1068	0.1787	0.1839	0.1048	0.0752	0.0746	0.0877	0.1279	0.1265	0.0974
94870	0.1022	0.1018	0.1256	0.1104	0.1104	0.1097	0.0715	0.0739	0.1074	0.1116	0.1124	0.0849
91836	0.1760	0.1967	0.2368	0.2930	0.2930	0.2024	0.1526	0.1616	0.5782	0.2170	0.2157	0.2230
22349	0.0514	0.2907	0.2213	0.0484	0.0483	0.1205	0.0560	0.1944	0.2254	0.0619	0.0624	0.1252

Tour ID	Long Term						Short Term					
	SDE	ARMA	GARCH	MA	ES	NM	SDE	ARMA	GARCH	MA	ES	NM
83648	0.1227	0.1260	0.1384	0.1471	0.1501	0.1681	0.1165	0.0992	0.1482	0.1733	0.1708	0.2087
44983	0.0568	0.0661	0.1222	0.0820	0.0801	0.0650	0.0489	0.2241	0.1235	0.0660	0.0653	0.0524
71801	0.2444	0.2434	0.2449	0.2447	0.2444	0.2434	0.1694	0.2930	0.3145	0.2559	0.2558	0.2447
58776	0.0769	0.0756	0.0781	0.1799	0.1674	0.0769	0.0835	0.0820	0.0560	0.1291	0.1262	0.0892
74359	0.1230	0.1232	0.1259	0.1261	0.1272	0.1218	0.0945	0.1001	0.1216	0.0757	0.0755	0.1171
1603059	0.2989	0.2960	0.4405	0.3756	0.3713	0.2843	0.1788	0.1799	0.6724	0.1791	0.1766	0.2118
76500	0.2079	0.2072	0.2090	0.2329	0.2307	0.2070	0.1439	0.1437	0.1924	0.1897	0.1889	0.1969
66306	0.0946	0.1949	0.1005	0.0989	0.0998	0.1295	0.0755	0.1842	0.0929	0.1111	0.1104	0.1394
1626293	0.1164	0.1392	0.1607	0.1167	0.1159	0.2015	0.0884	0.0934	0.1579	0.1072	0.1057	0.2020
1629646	0.5101	0.5035	0.3076	0.5593	0.5646	0.1557	0.1843	0.1861	0.5890	0.2695	0.2635	0.1915
90928	0.1272	0.1342	0.2520	0.2077	0.2103	0.1692	0.1193	0.1538	0.5619	0.1505	0.1494	0.1602
average	**0.1793**	**0.1886**	**0.2223**	**0.2234**	**0.2231**	**0.1761**	**0.1377**	**0.1615**	**0.3103**	**0.1734**	**0.1715**	**0.1755**

Table A.15. F1 Score of Out-of-Sample Prediction

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
5106	0.40	0.40	0.40	0.40	0.40	0.39	0.39	0.11	0.34	0.34
72528	0.25	0.25	0.25	0.40	0.41	0.46	0.44	0.00	0.45	0.46
73154	0.54	0.41	0.24	0.24	0.24	0.28	0.35	0.34	0.23	0.24
80961	0.33	0.41	0.33	0.33	0.33	0.37	0.39	0.47	0.32	0.32
29336	0.49	0.44	0.30	0.37	0.37	0.41	0.43	0.43	0.34	0.33
23222	0.34	0.34	0.33	0.34	0.34	0.43	0.43	0.16	0.43	0.43
30938	0.30	0.30	0.36	0.40	0.37	0.30	0.30	0.52	0.33	0.30
71480	0.39	0.35	0.39	0.39	0.39	0.46	0.46	0.46	0.46	0.46
88292	0.34	0.42	0.33	0.34	0.34	0.36	0.54	0.32	0.37	0.37
1618693	0.28	0.28	0.28	0.28	0.28	0.49	0.49	0.49	0.49	0.49
56737	0.38	0.36	0.28	0.38	0.38	0.50	0.44	0.01	0.50	0.50
71478	0.57	0.65	0.20	0.46	0.46	0.50	0.50	0.32	0.50	0.50
49049	0.44	0.42	0.24	0.44	0.44	0.49	0.40	0.35	0.49	0.48
72526	0.17	0.17	0.29	0.17	0.17	0.46	0.46	0.32	0.43	0.43
32391	0.26	0.21	0.26	0.26	0.26	0.46	0.39	0.46	0.46	0.46
72527	0.44	0.44	0.46	0.44	0.44	0.45	0.45	0.41	0.45	0.45
69278	0.45	0.45	0.15	0.44	0.45	0.32	0.32	0.42	0.25	0.26
28245	0.35	0.34	0.34	0.32	0.32	0.53	0.52	0.12	0.41	0.40
30814	0.52	0.51	0.45	0.16	0.16	0.36	0.36	0.47	0.31	0.31
16311	0.33	0.33	0.33	0.33	0.33	0.31	0.34	0.21	0.30	0.31
2753	0.27	0.27	0.27	0.27	0.27	0.31	0.30	0.39	0.29	0.29
3150	0.48	0.48	0.48	0.48	0.48	0.50	0.41	0.50	0.50	0.50
4171	0.36	0.50	0.30	0.36	0.36	0.49	0.41	0.03	0.49	0.49
5107	0.41	0.45	0.24	0.41	0.41	0.45	0.39	0.03	0.49	0.49
5452	0.45	0.21	0.21	0.21	0.21	0.34	0.38	0.29	0.33	0.33
5603	0.41	0.41	0.31	0.39	0.42	0.33	0.33	0.37	0.29	0.30
6712	0.50	0.04	0.21	0.50	0.50	0.50	0.35	0.30	0.50	0.50
7297	0.31	0.31	0.71	0.31	0.31	0.50	0.50	0.34	0.48	0.48
9474	0.27	0.54	0.28	0.27	0.27	0.30	0.38	0.32	0.25	0.27
11626	0.47	0.47	0.10	0.47	0.47	0.45	0.44	0.08	0.38	0.37
12043	0.47	0.47	0.47	0.47	0.47	0.42	0.39	0.43	0.31	0.31
17184	0.36	0.30	0.25	0.40	0.40	0.29	0.29	0.38	0.26	0.26
18201	0.50	0.50	0.15	0.50	0.50	0.50	0.50	0.30	0.50	0.50
20131	0.28	0.28	0.37	0.28	0.28	0.32	0.33	0.44	0.31	0.31
21148	0.23	0.23	0.23	0.23	0.23	0.32	0.31	0.22	0.25	0.25
21509	0.20	0.20	0.43	0.20	0.20	0.37	0.37	0.25	0.37	0.38
22510	0.38	0.61	0.44	0.38	0.38	0.46	0.48	0.34	0.46	0.46
22809	0.27	0.70	0.27	0.27	0.27	0.29	0.41	0.12	0.26	0.26
26561	0.28	0.25	0.38	0.28	0.28	0.38	0.35	0.44	0.40	0.40

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
28977	0.44	0.44	0.18	0.44	0.44	0.47	0.49	0.01	0.45	0.45
44981	0.48	0.48	0.08	0.48	0.48	0.49	0.40	0.03	0.49	0.49
46254	0.65	0.42	0.42	0.42	0.42	0.47	0.47	0.47	0.47	0.47
49063	0.37	0.37	0.29	0.37	0.37	0.33	0.33	0.45	0.26	0.28
49232	0.29	0.29	0.29	0.29	0.29	0.40	0.40	0.01	0.36	0.36
49435	0.30	0.49	0.30	0.30	0.30	0.36	0.38	0.36	0.36	0.36
50327	0.41	0.41	0.24	0.41	0.41	0.43	0.43	0.19	0.43	0.43
51324	0.30	0.30	0.36	0.30	0.30	0.33	0.33	0.46	0.30	0.29
52165	0.44	0.44	0.45	0.44	0.44	0.50	0.45	0.42	0.48	0.49
52950	0.42	0.42	0.54	0.42	0.42	0.37	0.42	0.49	0.27	0.27
53795	0.47	0.21	0.10	0.47	0.47	0.43	0.40	0.05	0.39	0.37
55106	0.42	0.30	0.21	0.21	0.21	0.31	0.31	0.25	0.38	0.33
55830	0.64	0.33	0.24	0.42	0.42	0.37	0.51	0.26	0.45	0.45
60195	0.35	0.35	0.35	0.35	0.35	0.49	0.42	0.49	0.49	0.49
60639	0.25	0.25	0.77	0.25	0.25	0.23	0.23	0.57	0.23	0.23
63798	0.03	0.03	0.49	0.03	0.03	0.47	0.47	0.49	0.39	0.39
66207	0.39	0.39	0.27	0.39	0.39	0.37	0.36	0.47	0.33	0.33
66215	0.45	0.45	0.34	0.45	0.45	0.47	0.47	0.37	0.48	0.46
66287	0.33	0.32	0.29	0.38	0.38	0.32	0.32	0.25	0.35	0.35
66865	0.13	0.13	0.12	0.13	0.13	0.33	0.33	0.39	0.28	0.28
69925	0.48	0.48	0.28	0.48	0.48	0.47	0.47	0.41	0.47	0.47
71821	0.16	0.16	0.67	0.16	0.16	0.34	0.33	0.45	0.34	0.34
72202	0.31	0.31	0.36	0.31	0.31	0.32	0.32	0.31	0.32	0.32
72532	0.35	0.35	0.47	0.31	0.31	0.31	0.31	0.47	0.25	0.26
72603	0.34	0.34	0.34	0.34	0.34	0.40	0.40	0.24	0.30	0.30
72723	0.55	0.39	0.39	0.39	0.39	0.48	0.50	0.46	0.50	0.50
72869	0.37	0.37	0.30	0.37	0.37	0.46	0.46	0.47	0.43	0.43
73163	0.23	0.23	0.23	0.23	0.23	0.34	0.34	0.34	0.34	0.34
73186	0.78	0.52	0.46	0.46	0.46	0.45	0.46	0.49	0.49	0.49
74983	0.37	0.50	0.50	0.00	0.00	0.50	0.50	0.50	0.48	0.48
78546	0.20	0.31	0.27	0.20	0.20	0.47	0.48	0.39	0.28	0.28
79499	0.50	0.50	0.00	0.50	0.50	0.50	0.50	0.00	0.50	0.50
79769	0.24	0.24	0.41	0.24	0.24	0.31	0.35	0.38	0.26	0.26
83878	0.32	0.32	0.50	0.32	0.32	0.41	0.41	0.46	0.41	0.41
84052	0.35	0.35	0.35	0.35	0.35	0.50	0.50	0.50	0.50	0.50
93877	0.39	0.32	0.17	0.17	0.17	0.33	0.33	0.17	0.41	0.37
1605378	0.56	0.61	0.31	0.31	0.31	0.37	0.37	0.37	0.35	0.35
1605628	0.46	0.46	0.35	0.46	0.46	0.48	0.48	0.39	0.48	0.48
1607268	0.20	0.20	0.43	0.20	0.20	0.29	0.28	0.48	0.30	0.31
1608443	0.30	0.30	0.30	0.30	0.30	0.44	0.44	0.09	0.40	0.42

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
1609587	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.19	0.27	0.27
1612648	0.30	0.30	0.30	0.30	0.30	0.32	0.32	0.22	0.25	0.25
1614660	0.16	0.16	0.54	0.16	0.16	0.32	0.32	0.36	0.24	0.24
1617321	0.43	0.43	0.19	0.43	0.43	0.29	0.29	0.37	0.28	0.28
1620855	0.44	0.44	0.43	0.20	0.20	0.49	0.49	0.47	0.43	0.44
1621632	0.42	0.43	0.46	0.46	0.46	0.48	0.48	0.43	0.50	0.50
1621699	0.48	0.48	0.46	0.48	0.48	0.48	0.48	0.37	0.45	0.45
1623624	0.22	0.36	0.19	0.19	0.19	0.31	0.30	0.29	0.31	0.31
1624666	0.46	0.46	0.34	0.46	0.46	0.46	0.46	0.46	0.46	0.46
1633110	0.16	0.41	0.56	0.16	0.16	0.37	0.52	0.42	0.32	0.32
85207	0.34	0.59	0.29	0.29	0.29	0.35	0.39	0.22	0.30	0.30
85282	0.32	0.32	0.32	0.32	0.32	0.33	0.33	0.17	0.29	0.29
93804	0.16	0.16	0.36	0.51	0.57	0.37	0.37	0.35	0.33	0.34
60051	0.50	0.50	0.50	0.03	0.48	0.50	0.50	0.50	0.48	0.50
44248	0.10	0.10	0.21	0.21	0.21	0.29	0.29	0.16	0.27	0.28
69982	0.41	0.41	0.64	0.41	0.41	0.45	0.45	0.39	0.45	0.45
76106	0.44	0.38	0.44	0.44	0.44	0.32	0.48	0.27	0.27	0.28
1601410	0.31	0.31	0.31	0.31	0.31	0.30	0.30	0.30	0.29	0.29
11869	0.49	0.49	0.47	0.49	0.49	0.49	0.49	0.35	0.49	0.49
78511	0.36	0.36	0.31	0.36	0.36	0.52	0.52	0.15	0.32	0.32
82421	0.81	0.34	0.33	0.33	0.29	0.50	0.36	0.30	0.33	0.35
15618	0.49	0.10	0.36	0.49	0.49	0.49	0.12	0.35	0.50	0.50
36953	0.31	0.37	0.35	0.31	0.31	0.33	0.43	0.38	0.27	0.28
66666	0.34	0.34	0.34	0.34	0.34	0.42	0.44	0.41	0.41	0.41
94870	0.35	0.35	0.32	0.32	0.32	0.36	0.36	0.29	0.33	0.33
91836	0.36	0.36	0.52	0.36	0.36	0.41	0.41	0.44	0.41	0.41
22349	0.50	0.06	0.00	0.50	0.50	0.50	0.31	0.00	0.50	0.50
83648	0.42	0.42	0.42	0.42	0.42	0.47	0.47	0.47	0.47	0.47
44983	0.56	0.49	0.28	0.28	0.28	0.40	0.35	0.24	0.32	0.32
71801	0.40	0.41	0.36	0.36	0.36	0.54	0.50	0.06	0.36	0.36
58776	0.31	0.24	0.44	0.19	0.19	0.34	0.34	0.47	0.24	0.24
74359	0.39	0.06	0.06	0.03	0.06	0.46	0.45	0.05	0.40	0.41
1603059	0.15	0.15	0.70	0.15	0.15	0.37	0.37	0.42	0.33	0.35
76500	0.33	0.33	0.33	0.33	0.33	0.41	0.41	0.40	0.41	0.41
66306	0.43	0.43	0.43	0.43	0.43	0.44	0.45	0.44	0.44	0.44
1626293	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1629646	0.21	0.21	0.29	0.21	0.21	0.34	0.34	0.44	0.30	0.30
90928	0.43	0.39	0.46	0.43	0.43	0.47	0.41	0.40	0.46	0.46
average	**0.37**	**0.36**	**0.34**	**0.33**	**0.34**	**0.41**	**0.40**	**0.33**	**0.38**	**0.38**

Table A.16. T-test Results of Model Comparison (F1 Score)

Comparison	Long Term	Short Term
SDE versue ARMA	0.9492 (0.1717)	0.4113 (0.3406)
SDE versus GARCH	1.8493**(0.0328)	5.0706***(0)
SDE versus MA	2.3966***(0.0087)	2.451***(0.0075)
SDE versus ES	2.1369**(0.0168)	2.4048***(0.0085)

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.17. Accuracy of Out-of-Sample Prediction

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
5106	67.9%	67.9%	67.9%	67.9%	67.9%	71.3%	71.3%	13.8%	60.8%	61.3%
72528	34.0%	34.0%	34.0%	66.0%	65.1%	90.3%	84.3%	0.3%	89.5%	89.5%
73154	64.8%	68.6%	31.4%	31.4%	31.4%	43.8%	56.0%	55.8%	31.8%	33.3%
80961	49.8%	53.2%	50.2%	49.8%	49.8%	69.0%	67.5%	91.0%	58.8%	59.3%
29336	62.5%	60.2%	42.4%	57.6%	57.6%	71.3%	70.0%	77.3%	55.0%	54.5%
23222	51.7%	51.7%	48.3%	51.7%	51.7%	77.5%	77.5%	22.5%	77.5%	77.5%
30938	41.8%	41.8%	45.1%	59.2%	58.2%	48.8%	48.8%	55.8%	48.8%	48.3%
71480	63.8%	35.9%	63.8%	63.8%	63.8%	84.3%	82.8%	84.3%	84.3%	84.3%
88292	51.6%	48.8%	48.4%	51.6%	51.6%	59.0%	68.5%	50.5%	61.0%	60.8%
1618693	38.4%	38.4%	38.4%	38.4%	38.4%	96.3%	96.3%	96.3%	96.3%	96.3%
56737	60.3%	55.7%	39.7%	60.3%	60.3%	99.3%	83.8%	0.8%	99.3%	99.3%
71478	62.6%	71.6%	20.6%	85.2%	85.2%	99.9%	99.9%	46.6%	99.9%	99.9%
49049	79.2%	71.2%	25.5%	79.2%	79.2%	96.3%	74.7%	51.0%	96.1%	96.0%
72526	21.1%	21.1%	29.4%	21.1%	21.1%	90.8%	90.8%	47.5%	85.8%	85.8%
32391	35.2%	24.9%	35.2%	35.2%	35.2%	84.5%	69.3%	84.5%	84.5%	84.5%
72527	79.8%	79.8%	69.3%	79.8%	79.8%	84.5%	84.5%	50.0%	84.5%	84.5%
69278	82.0%	82.0%	18.0%	79.8%	81.4%	52.5%	52.5%	75.0%	38.0%	39.3%
28245	48.9%	48.3%	52.2%	47.8%	47.8%	83.5%	83.3%	14.5%	74.3%	73.8%
30814	65.0%	63.6%	81.4%	18.6%	18.6%	66.3%	66.3%	88.8%	55.8%	56.3%
16311	50.0%	50.0%	50.0%	50.0%	50.0%	49.8%	49.3%	28.3%	46.5%	47.3%
2753	36.2%	36.2%	35.1%	36.2%	36.2%	49.0%	48.8%	44.0%	47.5%	47.5%
3150	92.9%	92.9%	92.9%	92.9%	92.9%	99.3%	80.5%	99.3%	99.3%	99.3%
4171	56.5%	50.0%	43.5%	56.5%	56.5%	96.3%	75.3%	3.8%	96.3%	96.3%
5107	68.6%	67.6%	31.4%	68.6%	68.6%	88.0%	74.5%	3.5%	96.5%	96.5%
5452	56.8%	27.0%	27.0%	27.0%	27.0%	54.0%	52.5%	44.5%	51.3%	51.3%
5603	46.7%	46.7%	39.5%	61.0%	62.1%	55.0%	55.0%	43.0%	47.3%	48.3%
6712	100.0%	3.7%	25.9%	100.0%	100.0%	100.0%	65.0%	45.0%	100.0%	100.0%
7297	45.7%	45.7%	73.9%	45.7%	45.7%	99.0%	99.0%	49.8%	95.7%	95.7%
9474	37.4%	64.1%	36.9%	37.4%	37.4%	49.8%	60.5%	39.0%	39.5%	43.5%
11626	88.4%	88.4%	11.6%	88.4%	88.4%	81.0%	80.0%	9.5%	68.8%	68.0%
12043	88.7%	88.7%	88.7%	88.7%	88.7%	65.0%	62.0%	76.5%	51.8%	51.8%
17184	36.8%	31.6%	32.9%	67.1%	67.1%	42.3%	42.3%	60.0%	36.5%	36.8%
18201	100.0%	100.0%	17.5%	100.0%	100.0%	100.0%	100.0%	43.0%	100.0%	100.0%
20131	38.5%	38.5%	38.5%	38.5%	38.5%	53.8%	55.3%	52.3%	52.3%	52.5%
21148	30.5%	30.5%	30.5%	30.5%	30.5%	53.3%	49.5%	34.0%	39.3%	39.5%
21509	24.2%	24.2%	75.8%	24.2%	24.2%	63.3%	63.3%	37.8%	61.8%	63.8%
22510	60.2%	62.4%	53.8%	60.2%	60.2%	85.8%	86.3%	43.5%	85.8%	85.8%
22809	36.8%	70.1%	36.8%	36.8%	36.8%	49.3%	70.3%	15.0%	44.3%	44.5%
26561	39.2%	27.2%	60.8%	39.2%	39.2%	66.8%	58.5%	78.8%	69.8%	70.3%
28977	77.9%	77.9%	22.1%	77.9%	77.9%	94.3%	94.8%	1.3%	89.3%	89.0%
44981	91.5%	55.3%	8.5%	91.5%	91.5%	96.0%	75.8%	4.0%	96.0%	96.0%
46254	65.8%	72.2%	72.2%	72.2%	72.2%	91.0%	91.0%	90.8%	91.0%	91.0%
49063	58.9%	58.9%	41.1%	58.9%	58.9%	53.5%	53.3%	79.8%	43.3%	47.3%
49232	40.0%	40.0%	40.0%	40.0%	40.0%	75.8%	75.8%	0.8%	70.5%	70.5%
49435	43.5%	49.1%	43.5%	43.5%	43.5%	61.8%	62.8%	61.8%	61.8%	61.8%
50327	70.5%	70.5%	30.4%	70.5%	70.5%	75.5%	75.5%	24.8%	75.5%	75.5%

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
51324	43.8%	43.8%	56.2%	43.8%	43.8%	56.5%	55.0%	85.0%	51.0%	49.0%
52165	79.0%	79.0%	78.5%	78.8%	79.0%	98.8%	86.5%	72.8%	96.4%	96.5%
52950	71.4%	71.4%	71.4%	71.4%	71.4%	47.8%	58.5%	52.8%	39.0%	40.8%
53795	88.8%	21.4%	11.2%	88.8%	88.8%	82.8%	75.0%	5.3%	75.0%	71.5%
55106	73.3%	43.3%	26.7%	26.7%	26.7%	46.0%	46.0%	34.0%	49.5%	45.0%
55830	67.4%	48.4%	29.5%	71.6%	71.6%	67.1%	77.1%	30.8%	83.8%	83.8%
60195	54.7%	54.7%	54.7%	54.7%	54.7%	97.5%	80.8%	97.0%	97.5%	97.5%
60639	33.3%	33.3%	81.5%	33.3%	33.3%	30.0%	30.0%	62.9%	30.0%	30.0%
63798	3.1%	3.1%	96.9%	3.1%	3.1%	92.0%	86.0%	96.5%	76.8%	77.0%
66207	64.3%	64.3%	28.6%	64.3%	64.3%	51.9%	51.3%	47.5%	50.0%	50.0%
66215	80.4%	80.4%	33.9%	80.4%	80.4%	87.0%	87.5%	46.8%	91.8%	87.8%
66287	35.1%	34.3%	39.9%	60.1%	60.1%	50.5%	50.5%	34.8%	57.3%	57.3%
66865	14.3%	14.3%	12.2%	14.3%	14.3%	57.8%	57.8%	44.3%	47.5%	47.8%
69925	90.9%	90.9%	30.9%	90.9%	90.9%	90.5%	90.5%	53.5%	90.5%	90.5%
71821	19.6%	19.6%	76.5%	19.6%	19.6%	59.3%	55.5%	53.8%	59.3%	59.8%
72202	44.7%	44.7%	55.3%	44.7%	44.7%	52.5%	52.5%	48.5%	52.5%	52.5%
72532	54.4%	54.4%	58.8%	45.6%	45.6%	50.5%	50.5%	61.0%	39.5%	40.8%
72603	52.4%	52.4%	52.4%	52.4%	52.4%	48.8%	58.3%	32.8%	46.5%	46.8%
72723	55.0%	64.4%	64.4%	64.4%	64.4%	94.8%	96.3%	85.8%	99.3%	99.3%
72869	57.5%	57.5%	42.5%	57.5%	57.5%	88.3%	88.3%	88.8%	83.0%	83.3%
73163	30.4%	30.4%	30.4%	30.4%	30.4%	57.3%	57.3%	57.3%	57.3%	57.3%
73186	84.5%	74.8%	84.5%	84.5%	84.5%	86.4%	87.0%	97.2%	97.2%	97.2%
74983	58.2%	100.0%	100.0%	0.0%	0.0%	100.0%	100.0%	100.0%	95.0%	95.0%
78546	24.2%	34.8%	27.3%	24.2%	24.2%	54.3%	63.0%	43.8%	42.8%	44.0%
79499	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%
79769	30.8%	30.8%	69.2%	30.8%	30.8%	50.3%	50.8%	64.8%	39.8%	39.8%
83878	47.4%	47.4%	52.6%	47.4%	47.4%	72.8%	72.8%	51.8%	72.8%	72.8%
84052	53.1%	53.1%	53.1%	53.1%	53.1%	99.6%	99.6%	98.7%	99.6%	99.6%
93877	39.0%	32.2%	20.3%	20.3%	20.3%	55.5%	55.5%	23.5%	63.8%	56.5%
1605378	56.9%	61.5%	45.4%	45.4%	45.4%	59.0%	59.0%	45.3%	54.0%	55.3%
1605628	84.8%	84.8%	54.5%	84.8%	84.8%	91.5%	91.5%	58.1%	91.5%	91.5%
1607268	24.6%	24.6%	75.4%	24.6%	24.6%	49.0%	50.0%	93.0%	56.5%	57.0%
1608443	43.8%	42.2%	43.8%	43.8%	43.8%	83.8%	83.8%	11.5%	74.0%	78.3%
1609587	53.9%	53.9%	53.9%	53.9%	53.9%	53.0%	53.3%	25.0%	43.0%	43.3%
1612648	41.9%	41.9%	41.9%	41.9%	41.9%	52.5%	52.5%	32.5%	38.8%	39.0%
1614660	18.8%	18.8%	62.5%	18.8%	18.8%	58.8%	57.1%	50.0%	40.8%	41.3%
1617321	76.4%	76.4%	23.6%	76.4%	76.4%	44.5%	44.5%	62.0%	41.3%	42.3%
1620855	44.3%	44.3%	74.3%	25.1%	25.1%	94.8%	94.8%	84.8%	84.5%	84.8%
1621632	70.0%	70.7%	85.3%	86.1%	86.1%	95.3%	95.3%	75.5%	99.8%	99.8%
1621699	91.4%	91.4%	57.1%	91.4%	91.4%	86.0%	86.0%	47.7%	85.7%	86.0%
1623624	22.1%	47.1%	23.5%	23.5%	23.5%	44.3%	44.8%	42.3%	44.5%	43.3%
1624666	86.2%	86.2%	51.7%	86.2%	86.2%	85.0%	85.0%	55.0%	85.0%	85.0%
1633110	19.2%	50.0%	76.9%	19.2%	19.2%	66.3%	78.3%	59.8%	57.5%	58.0%
85207	43.2%	61.6%	41.8%	41.8%	41.8%	56.0%	58.5%	29.8%	48.3%	49.8%
85282	46.6%	46.6%	46.6%	46.6%	46.6%	55.5%	55.5%	22.5%	46.5%	47.0%
93804	18.5%	18.5%	37.0%	82.7%	84.0%	69.0%	69.0%	48.0%	59.5%	60.5%
60051	100.0%	100.0%	100.0%	2.6%	92.3%	100.0%	100.0%	100.0%	95.0%	99.2%

	Long Term					Short Term				
Tour ID	SDE	ARMA	GARCH	MA	ES	SDE	ARMA	GARCH	MA	ES
44248	10.5%	10.5%	26.3%	26.3%	26.3%	46.9%	46.9%	21.4%	43.1%	44.4%
69982	69.4%	69.4%	66.7%	69.4%	69.4%	82.8%	82.8%	50.0%	82.8%	82.8%
76106	80.0%	38.0%	80.0%	80.0%	80.0%	52.5%	64.0%	43.0%	43.0%	43.5%
1601410	44.4%	44.4%	44.4%	44.4%	44.4%	44.7%	44.7%	44.4%	44.1%	44.1%
11869	96.2%	96.2%	88.6%	96.2%	96.2%	95.0%	95.0%	49.3%	95.0%	95.0%
78511	55.0%	55.0%	45.0%	55.0%	55.0%	75.3%	75.3%	20.3%	54.3%	55.3%
82421	81.5%	51.9%	48.1%	48.1%	40.7%	61.4%	56.4%	43.6%	50.0%	54.3%
15618	96.6%	10.3%	44.8%	96.6%	96.6%	93.9%	13.9%	50.6%	98.3%	98.3%
36953	45.9%	37.7%	54.1%	45.9%	45.9%	41.3%	56.5%	59.8%	36.3%	37.0%
66666	52.2%	52.2%	52.2%	52.2%	52.2%	68.0%	63.3%	69.0%	69.0%	69.0%
94870	53.6%	53.6%	46.4%	46.4%	46.4%	57.5%	57.5%	42.5%	51.3%	51.3%
91836	56.5%	56.5%	56.5%	56.5%	56.5%	73.8%	73.8%	50.0%	73.8%	73.8%
22349	100.0%	6.9%	0.0%	100.0%	100.0%	100.0%	52.2%	0.0%	100.0%	100.0%
83648	73.7%	73.7%	73.7%	73.7%	73.7%	88.0%	88.0%	88.0%	88.0%	88.0%
44983	61.3%	52.0%	38.7%	38.7%	38.7%	58.5%	56.8%	32.3%	49.3%	49.8%
71801	46.0%	48.0%	57.0%	57.0%	57.0%	87.8%	74.0%	7.0%	67.0%	67.0%
58776	31.4%	25.7%	77.1%	22.9%	22.9%	62.0%	62.0%	90.0%	41.3%	42.0%
74359	45.2%	6.5%	6.5%	3.2%	6.5%	87.7%	85.9%	5.5%	77.7%	79.1%
1603059	18.2%	18.2%	77.3%	18.2%	18.2%	69.0%	69.0%	58.3%	60.0%	64.0%
76500	48.1%	48.1%	48.1%	48.1%	48.1%	71.8%	71.8%	69.0%	71.8%	71.8%
66306	75.4%	75.4%	75.4%	75.4%	75.4%	83.0%	78.1%	83.0%	83.0%	83.0%
1626293	98.3%	98.3%	98.3%	98.3%	98.3%	99.8%	99.8%	99.8%	99.8%	99.8%
1629646	26.1%	26.1%	30.4%	26.1%	26.1%	56.3%	55.8%	47.3%	50.0%	50.3%
90928	75.0%	63.9%	45.8%	75.0%	75.0%	89.3%	73.5%	52.3%	88.8%	88.8%
average	**56.2%**	**52.2%**	**48.9%**	**54.3%**	**55.1%**	**72.1%**	**69.7%**	**50.5%**	**68.3%**	**68.5%**

Table A.18. T-test Results of Model Comparison (Accuracy)

Comparison	Long Term	Short Term
SDE versue ARMA	1.3103*(0.0957)	0.9703 (0.1665)
SDE versus GARCH	2.4379***(0.0078)	6.9037***(0)
SDE versus MA	0.6012 (0.2742)	1.4201*(0.0785)
SDE versus ES	0.3672 (0.3569)	1.3353*(0.0915)

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.19. The Parameter Estimation Results with n=10

Tour ID	$\hat{\lambda}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\sigma}$	γ_1	γ_2
5106	1.595***(0.036)	0.298***(0.204)	0.207***(0.045)	0.023**(0.012)	0.037***(0.005)	0.544***(0.068)	8.12	0.05
72528	1.533***(0.042)	0.42***(0.227)	0.079***(0.016)	0.043***(0.016)	0.02***(0.002)	0.177***(0.006)	2.07	0.23
73154	2.22***(0.050)	0.358***(0.318)	0.144**(0.086)	0.039***(0.009)	0.025***(0.004)	1.142***(0.077)	2.58	0.16
80961	6.377***(0.009)	0.193***(0.298)	0.4**(0.227)	0.018***(0.004)	0.02***(0.001)	2.143***(0.073)	4.35	0.04
29336	1.941***(0.022)	0.266***(0.162)	0.231**(0.108)	0.033***(0.012)	0.036***(0.003)	1.333***(0.07)	4.98	0.06
23222	1.633***(0.036)	0.248***(0.248)	0.173***(0.060)	0.018*(0.013)	0.03***(0.004)	0.574***(0.102)	7.95	0.04
30938	1.54***(0.045)	0.279***(0.25)	0.093*(0.071)	0.002 (0.006)	0.016***(0.003)	0.683***(0.097)	42.98	0.01
71480	3.428***(0.024)	0.261***(0.32)	0.161**(0.074)	0.015***(0.007)	0.015***(0.001)	0.878***(0.089)	4.25	0.07
88292	3.711***(0.017)	0.216***(0.284)	0.244*(0.179)	0.012 (0.012)	0.021***(0.002)	1.679***(0.091)	7.23	0.03
1618693	19.72***(0.005)	0.163***(0.558)	1.347**(0.651)	0.031***(0.003)	0.023***(0.001)	4.933***(0.072)	2.60	0.05
56737	7.437***(0.011)	0.199***(0.4)	0.395*(0.297)	0.028***(0.005)	0.021***(0.001)	2.534***(0.08)	2.41	0.07
71478	0.624***(0.043)	0.365***(0.098)	0.057***(0.011)	0.045**(0.024)	0.031***(0.004)	0.116***(0.006)	3.16	0.14
49049	1.565***(0.045)	0.333***(0.227)	0.124***(0.034)	0.021*(0.016)	0.024***(0.003)	0.461***(0.052)	5.76	0.07
72526	0.932***(0.054)	0.258***(0.142)	0.001(0.046)	0.236***(0.026)	0.046***(0.008)	0.325***(0.018)	0.01	0.36
32391	1.735***(0.047)	0.309***(0.29)	0.094***(0.034)	0.024***(0.011)	0.02***(0.003)	0.433***(0.043)	3.34	0.10
72527	0.819***(0.1)	0.302***(0.272)	0.049*(0.037)	0.001 (0.031)	0.013***(0.004)	0.308***(0.079)	118.79	0.00
69278	1.199***(0.047)	0.278***(0.202)	0.095*(0.073)	0.005 (0.017)	0.02***(0.003)	0.731***(0.109)	21.04	0.02
28245	1.052***(0.057)	0.357***(0.182)	0.088***(0.027)	0.023**(0.011)	0.028***(0.005)	0.342***(0.035)	5.62	0.08
30814	0.802***(0.071)	0.288***(0.201)	0.021 (0.035)	0.045***(0.015)	0.013***(0.003)	0.357***(0.072)	0.81	0.23
16311	2.268***(0.060)	0.372***(0.447)	0.083*(0.057)	0.039***(0.01)	0.016***(0.003)	0.773***(0.043)	1.49	0.24
2753	0.275***(0.025)	0.187***(0.014)	0.195***(0.034)	0.503***(0.017)	0.548***(0.007)	0.525***(0.025)	1.73	0.08
3150	0.164***(0.066)	0.288***(0.077)	-0.003 (0.009)	0.165***(0.036)	0.036***(0.006)	0.114***(0.01)	-0.16	0.48
4171	0.212***(0.047)	0.313***(0.038)	0.002 (0.004)	0.131***(0.014)	0.03***(0.002)	0.094***(0.004)	0.1	0.41
5107	1.326***(0.191)	0.172***(0.024)	0.004 (0.008)	0.005 (0.007)	0.003***(0)	0.193***(0.003)	0.68	0.12
5452	0.464***(0.062)	0.292***(0.031)	0.007*(0.004)	0.04***(0.014)	0.011***(0.001)	0.122***(0.003)	0.5	0.28
5603	0.406***(0.096)	0.184***(0.03)	0.011**(0.006)	0.039***(0.014)	0.025***(0.004)	0.102***(0.004)	0.83	0.12
6712	0.12**(0.055)	0.264***(0.096)	0.005 (0.009)	0.095**(0.048)	0.051***(0.01)	0.108***(0.013)	0.64	0.22
7297	0.547***(0.096)	0.244***(0.033)	0.022**(0.011)	0.065***(0.017)	0.038***(0.004)	0.164***(0.007)	0.81	0.18
9474	0.255***(0.087)	0.282***(0.082)	0.001 (0.003)	0.015 (0.024)	0.005***(0.002)	0.073***(0.003)	0.53	0.26
11626	0.175***(0.074)	0.297***(0.058)	0.009*(0.007)	0.127**(0.059)	0.04***(0.01)	0.077***(0.007)	0.59	0.27
12043	0.108***(0.044)	0.121***(0.045)	0.003*(0.003)	0.009 (0.015)	0.026***(0.006)	0.049***(0.003)	3.95	0.03
17184	0.236***(0.031)	0.293***(0.035)	0.013**(0.007)	0.074***(0.021)	0.019***(0.001)	0.135***(0.005)	1.1	0.2
18201	0.188***(0.036)	0.2***(0.028)	0.013**(0.006)	0.096***(0.011)	0.095***(0.005)	0.11***(0.007)	0.86	0.13
20131	0.27***(0.057)	0.296***(0.042)	0.054***(0.022)	0.258***(0.085)	0.07***(0.009)	0.127***(0.014)	1.1	0.2
21148	0.154***(0.062)	0.321***(0.053)	0.005*(0.003)	0.088***(0.036)	0.024***(0.006)	0.053***(0.004)	0.5	0.32
21509	0.457 (0.384)	0.36*(0.275)	0 (0.011)	0.014 (0.029)	0.002*(0.002)	0.125***(0.008)	-0.12	0.64
22510	0.347***(0.056)	0.422***(0.035)	0.022*(0.016)	0.419***(0.033)	0.064***(0.003)	0.212***(0.012)	0.26	0.58
22809	0.387***(0.038)	0.293***(0.011)	0.014*(0.009)	0.581***(0.024)	0.176***(0.006)	0.146***(0.008)	0.09	0.38
26561	0.436***(0.08)	0.327***(0.023)	0.009**(0.005)	0.128***(0.018)	0.04***(0.004)	0.098***(0.004)	0.23	0.39
28977	0.33***(0.029)	0.361***(0.02)	0.016***(0.006)	0.206***(0.018)	0.042***(0.002)	0.169***(0.004)	0.37	0.41
44981	0.214***(0.039)	0.375***(0.033)	-0.012***(0.003)	0.194***(0.021)	0.021***(0.004)	0.058***(0.003)	-0.48	1.16
46254	0.295***(0.103)	0.418***(0.04)	-0.004 (0.005)	0.19***(0.027)	0.03***(0.004)	0.08***(0.005)	-0.13	0.82
49063	0.184***(0.044)	0.366***(0.049)	0.011**(0.007)	0.179***(0.042)	0.042***(0.005)	0.094***(0.007)	0.55	0.37
49232	0.273***(0.038)	0.303***(0.019)	-0.032***(0.007)	0.514***(0.022)	0.083***(0.005)	0.146***(0.006)	-0.33	0.65
49435	0.411***(0.135)	0.262***(0.077)	0.002 (0.009)	0.027 (0.025)	0.009***(0.002)	0.108***(0.006)	0.28	0.28
50327	0.384***(0.033)	0.35***(0.017)	0.026**(0.011)	0.424***(0.018)	0.085***(0.002)	0.244***(0.008)	0.24	0.43
51324	0.533***(0.093)	0.219***(0.035)	0.004 (0.005)	0.017*(0.012)	0.01***(0.002)	0.133***(0.003)	0.6	0.18
52165	2.027***(0.141)	0.236***(0.008)	0.02*(0.016)	0.174***(0.007)	0.057***(0.002)	0.276***(0.005)	0.08	0.29
52950	0.254***(0.029)	0.206***(0.006)	0.017**(0.008)	0.855***(0.028)	0.538***(0.015)	0.134***(0.008)	0.1	0.24
53795	0.173***(0.025)	0.236***(0.025)	0.019**(0.011)	0.306***(0.035)	0.07***(0.004)	0.167***(0.008)	0.47	0.21
55106	0.165*(0.121)	0.642***(0.054)	0.001 (0.004)	0.288***(0.065)	0.026***(0.005)	0.05***(0.006)	0.07	1.68
55830	0.421***(0.051)	0.455***(0.018)	0.013**(0.007)	0.385***(0.04)	0.075***(0.006)	0.108***(0.006)	0.14	0.73
60195	0.162***(0.018)	0.152***(0.012)	0.028***(0.004)	0.106***(0.008)	0.168***(0.004)	0.076***(0.005)	1.89	0.06
60639	0.109***(0.028)	0.189***(0.017)	-0.006 (0.005)	0.561***(0.039)	0.246***(0.018)	0.071***(0.008)	-0.12	0.27
63798	0.157***(0.026)	0.532***(0.018)	0.004*(0.003)	0.461***(0.034)	0.058***(0.004)	0.055***(0.003)	0.12	1.01
66207	0.166*(0.103)	0.311***(0.103)	0.004 (0.005)	0.061**(0.031)	0.026***(0.007)	0.06***(0.007)	0.57	0.29
66215	0.224***(0.055)	0.267***(0.036)	0.007 (0.006)	0.138***(0.049)	0.052***(0.011)	0.083***(0.006)	0.33	0.27

Notes. The estimation is based on the in-sample data for each tour.

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Tour ID	$\hat{\lambda}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\sigma}$	γ_1	γ_2
66287	0.295***(0.031)	0.361***(0.019)	0.006**(0.004)	0.176***(0.015)	0.049***(0.003)	0.122***(0.004)	0.18	0.48
66865	0.607***(0.142)	0.103***(0.024)	-0.019 (0.017)	0.055**(0.032)	0.009***(0.002)	0.145***(0.007)	-0.63	0.31
69925	0.361***(0.067)	0.554***(0.038)	0.014**(0.007)	0.239***(0.034)	0.05***(0.004)	0.1***(0.008)	0.36	0.91
71821	0.249***(0.042)	0.333***(0.025)	0.023***(0.009)	0.336***(0.03)	0.061***(0.003)	0.083***(0.007)	0.41	0.35
72202	0.454***(0.088)	0.191***(0.029)	0.062**(0.033)	0.204***(0.037)	0.066***(0.005)	0.233***(0.014)	0.83	0.13
72532	0.668**(0.391)	0.225**(0.118)	0.014*(0.009)	0.003 (0.02)	0.009**(0.005)	0.113***(0.005)	7.89	0.03
72603	0.124**(0.062)	0.525***(0.134)	0.004**(0.002)	0.032 (0.028)	0.014***(0.004)	0.041***(0.003)	1.93	0.38
72723	1.505***(0.093)	0.141***(0.009)	0.059**(0.026)	0.023*(0.017)	0.021***(0.001)	0.323***(0.007)	2	0.05
72869	0.132***(0.018)	0.514***(0.018)	0.013***(0.005)	0.974***(0.035)	0.13***(0.004)	0.12***(0.007)	0.22	0.87
73163	0.205***(0.033)	0.364***(0.029)	0.001 (0.004)	0.205***(0.016)	0.04***(0.002)	0.097***(0.004)	0.04	0.55
73186	0.283***(0.051)	0.37***(0.053)	0.02***(0.006)	0.043**(0.021)	0.032***(0.003)	0.121***(0.006)	2.68	0.16
74983	0.42***(0.105)	0.333***(0.051)	0.005 (0.018)	0.211***(0.033)	0.035***(0.004)	0.163***(0.011)	0.08	0.46
78546	0.225**(0.102)	0.564***(0.057)	0.004*(0.002)	0.09***(0.026)	0.02***(0.004)	0.055***(0.003)	0.4	0.92
79499	1.486 (1.951)	0.38***(0.137)	-0.012 (0.052)	0.102**(0.054)	0.016**(0.009)	0.184***(0.021)	-0.13	0.7
79769	0.731***(0.142)	0.366***(0.061)	0.122*(0.093)	0.257***(0.055)	0.059***(0.004)	0.542***(0.036)	1.02	0.29
83878	0.349***(0.063)	0.515***(0.034)	0.001 (0.007)	0.286***(0.027)	0.038***(0.003)	0.117***(0.006)	0.03	1.03
84052	0.249***(0.015)	0.5***(0.015)	0.052***(0.007)	0.554***(0.015)	0.095***(0.001)	0.238***(0.006)	0.76	0.57
93877	0.557***(0.084)	0.216***(0.02)	0.09**(0.053)	0.655***(0.029)	0.218***(0.006)	0.42***(0.026)	0.31	0.21
1605378	2.411***(0.564)	0.421***(0.026)	0.01 (0.029)	0.183***(0.02)	0.031***(0.003)	0.223***(0.01)	0.04	0.7
1605628	0.47***(0.147)	0.485***(0.077)	0.043**(0.022)	0.182**(0.081)	0.023***(0.004)	0.103***(0.012)	0.98	0.47
1607268	1.327***(0.294)	0.522***(0.03)	0.032**(0.019)	0.215***(0.019)	0.036***(0.002)	0.192***(0.009)	0.24	0.88
1608443	0.418***(0.04)	0.339***(0.026)	1.58*(0.963)	15.498***(2.028)	2.628***(0.211)	3.426***(0.167)	0.37	0.37
1609587	1.732***(0.665)	0.436***(0.056)	0.062*(0.041)	0.153***(0.039)	0.026***(0.005)	0.197***(0.015)	0.42	0.55
1612648	0.539***(0.044)	0.158***(0.01)	2.053**(0.98)	13.65***(1.287)	6.987***(0.447)	3.736***(0.155)	0.33	0.14
1614660	0.316***(0.084)	0.196***(0.05)	-0.03 (0.033)	0.246***(0.044)	0.062***(0.008)	0.277***(0.026)	-0.49	0.47
1617321	0.593***(0.134)	0.222***(0.045)	0.029*(0.019)	0.036*(0.026)	0.031***(0.004)	0.188***(0.011)	1.74	0.1
1620855	3.759***(0.691)	0.414***(0.029)	0.004 (0.055)	0.19***(0.021)	0.038***(0.003)	0.343***(0.016)	0.01	0.7
1621632	2.47***(0.272)	0.399***(0.025)	0.074**(0.037)	0.168***(0.015)	0.037***(0.002)	0.395***(0.012)	0.3	0.51
1621699	0.552 (2.003)	0.196 (0.444)	0.005 (0.021)	0.007 (0.059)	0.006 (0.018)	0.123***(0.012)	1.45	0.1
1623624	1.447**(0.653)	0.227**(0.102)	0.004 (0.11)	0.02 (0.072)	0.006**(0.002)	0.313***(0.034)	0.16	0.25
1624666	1.1***(0.414)	0.309***(0.049)	0.09**(0.047)	0.178***(0.057)	0.038***(0.007)	0.139***(0.017)	0.66	0.27
1633110	0.469*(0.313)	0.258***(0.107)	-0.01 (0.011)	0.07***(0.019)	0.019***(0.007)	0.134***(0.009)	-0.42	0.6
85207	0.894***(0.167)	0.176***(0.033)	0.032*(0.023)	0.005 (0.055)	0.015***(0.003)	0.196***(0.009)	9.26	0.02
85282	0.415***(0.041)	0.281***(0.022)	0.04***(0.016)	0.217***(0.012)	0.063***(0.001)	0.287***(0.009)	0.62	0.24
93804	0.378***(0.067)	0.3***(0.04)	0.035***(0.011)	0.078***(0.022)	0.042***(0.003)	0.147***(0.008)	1.69	0.16
60051	0.151***(0.021)	0.4***(0.017)	0.009*(0.006)	0.897***(0.03)	0.222***(0.006)	0.113***(0.009)	0.11	0.6
44248	0.325***(0.129)	0.32***(0.077)	0.012**(0.005)	0.027 (0.095)	0.029 (0.032)	0.066***(0.005)	2	0.16
69982	0.353***(0.089)	0.24***(0.056)	0.026***(0.008)	0.007 (0.041)	0.048***(0.016)	0.084***(0.007)	13.93	0.02
76106	0.311***(0.037)	0.2***(0.019)	0.049***(0.02)	0.307***(0.054)	0.131***(0.01)	0.298***(0.012)	0.64	0.15
1601410	0.508***(0.122)	0.231***(0.031)	0.078**(0.037)	0.329***(0.038)	0.152***(0.01)	0.279***(0.022)	0.61	0.19
11869	0.481***(0.059)	0.25***(0.022)	0.023**(0.014)	0.185***(0.024)	0.058***(0.004)	0.227***(0.008)	0.35	0.25
78511	0.256***(0.037)	0.56***(0.019)	0.008*(0.005)	0.504***(0.02)	0.051***(0.002)	0.107***(0.005)	0.14	1.12
82421	0.26***(0.038)	0.4***(0.03)	0.053***(0.01)	0.266***(0.045)	0.107***(0.008)	0.1***(0.009)	1.28	0.29
15618	0.15**(0.072)	0.32***(0.043)	0 (0.004)	0.165***(0.018)	0.04***(0.004)	0.047***(0.005)	0.02	0.46
36953	0.106***(0.015)	0.333***(0.035)	0.023**(0.011)	0.473***(0.037)	0.1***(0.004)	0.2***(0.011)	0.68	0.3
66666	0.105***(0.008)	0.5***(0.01)	0.004**(0.003)	1.161***(0.023)	0.193***(0.003)	0.107***(0.004)	0.07	0.93
94870	0.436***(0.079)	0.28***(0.037)	0.074***(0.015)	0.075**(0.042)	0.056***(0.007)	0.094***(0.009)	3.16	0.09
91836	0.293***(0.101)	0.4***(0.041)	0.001 (0.005)	0.155***(0.018)	0.031***(0.003)	0.084***(0.005)	0.02	0.65
22349	0.166***(0.039)	0.28***(0.033)	0.013***(0.004)	0.125***(0.024)	0.065***(0.006)	0.053***(0.005)	0.87	0.21
83648	1.054***(0.17)	0.379***(0.054)	0.079**(0.039)	0.069***(0.02)	0.029***(0.002)	0.351***(0.016)	1.75	0.22
44983	0.134***(0.029)	0.174***(0.024)	0.006***(0.002)	0.054***(0.016)	0.065***(0.008)	0.055***(0.003)	1.07	0.1
71801	0.129***(0.016)	0.321***(0.022)	0.013***(0.005)	0.438***(0.029)	0.077***(0.003)	0.121***(0.005)	0.34	0.35
58776	0.153***(0.028)	0.24***(0.021)	0.032***(0.012)	0.651***(0.045)	0.236***(0.011)	0.159***(0.013)	0.42	0.22
74359	0.167***(0.041)	0.429***(0.018)	0.01**(0.005)	0.64***(0.043)	0.09***(0.005)	0.059***(0.005)	0.17	0.64
1603059	0.305***(0.053)	0.214***(0.019)	0.021**(0.011)	0.298***(0.039)	0.095***(0.008)	0.103***(0.008)	0.29	0.21
76500	4.306***(0.714)	0.373***(0.042)	0.559**(0.28)	0.311***(0.037)	0.044***(0.002)	0.827***(0.042)	0.67	0.36
66306	0.673***(0.089)	0.586***(0.03)	0.071***(0.014)	0.283***(0.051)	0.055***(0.005)	0.146***(0.009)	0.9	0.74
1626293	0.6***(0.07)	0.348***(0.016)	0.033**(0.016)	0.467***(0.031)	0.127***(0.006)	0.168***(0.01)	0.18	0.45
1629646	1.267***(0.343)	0.72***(0.045)	0.018 (0.031)	0.368***(0.065)	0.018***(0.002)	0.191***(0.013)	0.14	2.25
90928	0.386***(0.059)	0.68***(0.025)	0.02***(0.007)	0.463***(0.04)	0.038***(0.002)	0.113***(0.006)	0.35	1.57

Notes. The estimation is based on the in-sample data for each tour. *** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.20. Predictive Performance of SDE without response

Method	Long Term	Short Term
SDE without response	0.2559 (0.1730)	0.1503 (0.0740)
SDE	0.1878 (0.0868)	0.1353 (0.0618)
ARMA	0.1937 (0.0861)	0.1547 (0.076)
GARCH	0.2505 (0.1383)	0.6251 (0.7716)

Table A.21. Predictive Performance Comparison of SDE without response

Comparison	Long Term	Short Term
SDE without response versus SDE	3.8082***(0.0001)	1.6761**(0.0475)
SDE without response versus ARMA	3.4829***(0.0003)	-0.4475 (0.3275)
SDE without response versus GARCH	0.2639 (0.396)	-6.6259***(0)

*** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.22. ARMA Estimate for Ctrip Data

Tour ID	Constant	AR{1}	AR{2}	MA{1}	MA{2}
5106	0.046 (3.762)	0.076 (2.791)	0.775 (30.64)	0.925 (22.880)	-0.038 (-1.027)
72528	0.059 (4.323)	0.866 (32.443)	N/A	0.107 (2.279)	0.226 (4.449)
73154	0.016 (2.965)	1.737 (22.843)	-0.774 (-11.862)	-0.771 (-8.388)	N/A
80961	0.009 (7.068)	1.732 (47.400)	-0.769 (-14.684)	-0.679 (-14.684)	N/A
29336	0.058 (5.352)	-0.002 (-0.093)	0.823 (43.429)	0.997 (313.814)	N/A
23222	0.008 (3.152)	1.813 (28.596)	-0.842 (-14.839)	-0.808 (-11.171)	N/A
30938	0.061 (3.161)	0.012 (0.342)	0.825 (29.439)	0.989 (115.356)	N/A
71478	0.054 (5.881)	0.039 (2.405)	0.792 (59.479)	0.988 (202.526)	N/A
88292	0.046 (3.926)	0.063 (3.959)	0.786 (44.934)	1.012 (31.914)	0.020 (0.625)
1618693	0.007 (5.810)	1.715 (39.246)	-0.750 (-19.549)	-0.742 (-14.651)	N/A
56737	0.007 (5.095)	1.759 (40.064)	-0.789 (-20.139)	-0.757 (-13.881)	N/A
71478	0.010 (2.408)	1.671 (13.893)	-0.711 (-6.564)	-0.620 (-4.042)	N/A
49049	0.040 (3.425)	0.044 (2.073)	0.832 (41.306)	0.992 (172.259)	N/A
72526	0.029 (2.438)	0.885 (25.129)	N/A	0.002 (0.029)	N/A
32391	0.067 (4.534)	0.037 (1.531)	0.789 (49.048)	0.995 (27.503)	-0.001 (-0.003)
72527	0.029 (1.909)	0.033 (0.870)	0.824 (20.015)	1 (89.927)	N/A
69278	0.012 (2.073)	1.697 (12.410)	-0.734 (-6.079)	-0.694 (-4.319)	N/A
28245	0.038 (4.313)	0.062 (2.488)	0.833 (33.982)	1 (193.147)	N/A
30814	0.013 (2.865)	1.749 (18.315)	-0.789 (-9.176)	-0.742 (-6.878)	N/A
16311	0.014 (2.695)	1.730 (19.305)	-0.764 (-9.737)	-0.763 (-7.343)	N/A
2753	0.006 (0.703)	0.980 (58.916)	N/A	0.062 (0.828)	N/A
3150	0.007 (1.751)	1.691 (14.342)	-0.707 (-6.459)	-1 (6.742)	N/A
4171	0.003 (1.488)	1.918 (48.415)	-0.925 (-25.414)	-0.951 (-18.351)	N/A
5107	0.003 (3.231)	1.866 (47.091)	-0.873 (-23.252)	-0.862 (-18.104)	N/A
5452	0.037 (5.303)	0.001 (0.065)	0.894 (65.747)	1 (349.629)	N/A
5603	0.025 (3.533)	0.92 (44.454)	N/A	0.138 (2.617)	N/A
6712	0.002 (0.632)	1.972 (19.701)	-0.98 (-10.486)	-1 (-5.272)	N/A
7297	0.019 (2.032)	0.959 (54.207)	N/A	0.15 (2.927)	N/A
9474	0.033 (3.048)	0.034 (1.609)	0.874 (31.636)	1.071 (21.025)	0.071 (1.374)
11626	0.036 (1.784)	0.913 (17.869)	N/A	0.145 (1.183)	N/A
12043	0.034 (3.309)	0.909 (30.682)	N/A	0.203 (2.176)	N/A
17184	0.011 (2.048)	0.957 (65.037)	N/A	0.108 (1.886)	N/A
18201	0.018 (1.107)	0.967 (38.855)	N/A	-0.038 (-0.432)	0.274 (3.022)
20131	0.016 (1.544)	0.962 (33.816)	N/A	-0.091 (-0.795)	N/A
21148	0.033 (1.902)	0.92 (23.184)	N/A	0.161 (1.497)	N/A
21509	0.02 (1.952)	0.952 (41.662)	N/A	0.072 (0.936)	N/A
22510	0.056 (2.777)	-0.019 (-0.642)	0.885 (23.107)	1 (32.799)	N/A
22809	0.003 (3.608)	1.948 (78.347)	-0.955 (-39.025)	-1 (-58.257)	N/A
26561	0.005 (2.871)	1.921 (52.278)	-0.93 (-26.771)	-0.903 (-18.342)	N/A
28977	0.036 (4.001)	0.026 (1.319)	0.887 (50.766)	1 (111.656)	N/A
44981	0.004 (1.821)	1.93 (51.422)	-0.938 (-26.546)	-0.935 (-17.952)	N/A

Tour ID	Constant	AR{1}	AR{2}	MA{1}	MA{2}
46254	0.051 (2.721)	0.91 (30.029)	N/A	0.254 (2.587)	N/A
49063	0.024 (1.677)	0.945 (28.458)	N/A	0.187 (1.497)	N/A
49232	0.043 (2.654)	0.909 (25.848)	N/A	0.006 (0.081)	N/A
49435	0.001 (1.982)	1.956 (100.281)	-0.958 (-50.44)	-1 (-59.297)	N/A
50327	0.014 (2.135)	0.965 (75.59)	N/A	0.065 (1.302)	N/A
51324	0.02 (3.53)	0.958 (88.708)	N/A	0.048 (1.443)	N/A
52165	0.002 (2.462)	1.836 (30.362)	-0.843 (-14.487)	-0.84 (-12.46)	N/A
52950	0.007 (3.232)	1.854 (48.092)	-0.866 (-24.311)	-1 (-19.646)	N/A
53795	0.022 (1.8)	0.13 (1.459)	0.77 (8.309)	0.925 (16.03)	N/A
55106	0.096 (1.953)	0.861 (12.26)	N/A	0.195 (1.471)	N/A
55830	0.013 (2.056)	1.805 (25.965)	-0.828 (-13.452)	-0.806 (-8.793)	0.045 (0.561)
60195	0.003 (1.234)	1.901 (41.652)	-0.907 (-21.627)	-1 (-13.673)	N/A
60639	0.019 (0.716)	0.968 (18.644)	N/A	-0.106 (-0.581)	N/A
63798	0.058 (2.244)	0.9 (20.731)	N/A	0.195 (2.666)	N/A
66207	0.009 (0.71)	0.986 (38.579)	N/A	-0.03 (-0.205)	N/A
66215	0.017 (1.088)	0.961 (30.262)	N/A	0.029 (0.316)	N/A
66287	0.023 (3.264)	0.961 (85.626)	N/A	-0.039 (-0.988)	N/A
66865	0.006 (1.059)	0.973 (54.452)	N/A	0.033 (0.562)	N/A
69925	0.017 (0.631)	0.986 (28.778)	N/A	0.088 (0.853)	0.345 (3.953)
71821	0.036 (2.297)	0.889 (18.336)	N/A	0.038 (0.406)	0.289 (3.073)
72202	0.014 (1.892)	0.932 (28.831)	N/A	0.125 (1.494)	N/A
72532	0.017 (2.356)	0.951 (50.621)	N/A	-0.019 (-0.26)	N/A
72603	0.005 (0.629)	1.888 (24.351)	-0.896 (-13.071)	-1 (-7.721)	N/A
72723	0.013 (2.62)	0.036 (1.651)	0.912 (45.414)	0.98 (83.445)	N/A
72869	0.067 (2.796)	0.876 (20.227)	N/A	0.135 (1.776)	N/A
73163	0.015 (1.845)	0.969 (73.475)	N/A	0.096 (1.968)	N/A
73186	0.017 (1.779)	0.969 (62.752)	N/A	0.048 (0.799)	N/A
74983	0.022 (1.112)	0.059 (1.535)	0.884 (16.824)	1 (39.133)	N/A
78546	0.009 (3.933)	1.917 (67.963)	-0.927 (-34.842)	-1 (-36.853)	N/A
79499	0.229 (3.998)	0.58 (5.533)	N/A	0.544 (12.077)	0.93 (18.499)
79769	0.044 (2.61)	0.897 (23.638)	N/A	-0.017 (-0.173)	N/A
83878	0.021 (1.48)	0.964 (50.494)	N/A	0.168 (2.694)	N/A
84052	0.041 (3.374)	0.162 (1.304)	0.746 (6.304)	0.868 (6.942)	0.128 (3.178)
93877	0.017 (1.292)	0.948 (27.897)	N/A	-0.012 (-0.123)	N/A
1605378	0.031 (2.62)	0.94 (44.506)	N/A	-0.017 (-0.301)	N/A
1605628	-0.002 (-0.137)	0.996 (32.839)	N/A	-0.15 (-1.117)	N/A
1607268	0.037 (2.74)	0.946 (48.19)	N/A	-0.037 (-0.556)	N/A
1608443	0.019 (1.218)	0.949 (26.621)	N/A	0.173 (1.882)	N/A
1609587	0.041 (2.526)	0.924 (32.163)	N/A	0.029 (0.335)	N/A
1612648	0.014 (1.255)	0.956 (41.808)	N/A	0.146 (2.668)	N/A

Tour ID	Constant	AR{1}	AR{2}	MA{1}	MA{2}
1614660	0.034 (0.884)	0.895 (8.295)	N/A	0.271 (2.523)	N/A
1617321	0.027 (1.851)	0.933 (29.147)	N/A	0.113 (1.523)	N/A
1620855	0.033 (2.739)	0.945 (51.17)	N/A	0.048 (0.847)	N/A
1621632	0.022 (3.414)	0.955 (80.481)	N/A	0.061 (1.474)	N/A
1621699	0.026 (1.64)	0.914 (18.235)	N/A	0.069 (0.652)	N/A
1623624	0.03 (2.475)	0.893 (22.95)	N/A	0.303 (3.85)	N/A
1624666	0.029 (1.377)	0.914 (15.991)	N/A	0.243 (1.596)	N/A
1633110	0.005 (3.153)	1.961 (97.547)	-0.968 (-46.688)	-1 (-18.63)	N/A
85207	0.014 (2.133)	0.946 (45.237)	N/A	0.161 (2.741)	N/A
85282	0.008 (1.689)	0.974 (86.254)	N/A	0.104 (2.763)	N/A
93804	0.029 (2.093)	0.923 (27.019)	N/A	0.044 (0.736)	N/A
60051	0.046 (1.625)	0.924 (20.392)	N/A	0.215 (1.823)	N/A
44248	0.027 (1.309)	0.961 (31.382)	N/A	0.041 (0.422)	N/A
69982	0.034 (1.367)	0.942 (20.277)	N/A	0.225 (1.927)	N/A
76106	0.007 (2.248)	1.769 (31.23)	-0.791 (-14.859)	-0.673 (-7.84)	N/A
1601410	0.045 (1.815)	0.891 (15.165)	N/A	-0.037 (-0.25)	N/A
11869	0.016 (2.333)	0.956 (52.47)	N/A	0.032 (0.63)	N/A
78511	0.034 (2.029)	0.931 (30.125)	N/A	-0.093 (-1.186)	N/A
82421	0.113 (2.158)	0.288 (2.428)	0.514 (3.879)	1 (13.759)	N/A
15618	0.182 (2.425)	0.233 (2.186)	0.399 (3.808)	1 (29.657)	N/A
36953	0.004 (4.791)	1.923 (91.202)	-0.936 (-42.106)	-0.818 (-9.479)	-0.182 (-2.048)
66666	0.019 (2.431)	1.68 (11.08)	-0.71 (-4.976)	-0.589 (-3.146)	N/A
94870	0.03 (1.739)	0.92 (18.737)	N/A	-0.027 (-0.176)	N/A
91836	0.019 (1.066)	0.969 (29.462)	N/A	0.032 (0.372)	N/A
22349	0.007 (2.755)	1.902 (74.796)	-0.918 (-33.167)	-1 (-12.094)	N/A
83648	0.026 (1.96)	0.94 (33.6)	N/A	0.209 (3.19)	N/A
44983	0.002 (3.71)	1.955 (84.293)	-0.959 (-42.178)	-1 (-41.506)	N/A
71801	0.005 (1.462)	1.867 (28.054)	-0.883 (-15.13)	-0.904 (-9.668)	N/A
58776	0.03 (1.214)	0.924 (16.814)	N/A	0.14 (0.992)	N/A
74359	0.096 (1.811)	0.78 (6.49)	N/A	0.001 (0.006)	N/A
1603059	0.026 (1.353)	0.923 (15.839)	N/A	0.006 (0.044)	N/A
76500	0.012 (1.527)	0.958 (42.705)	N/A	-0.088 (-1.189)	N/A
66306	0.002 (0.747)	1.827 (35.075)	-0.83 (-16.466)	-0.766 (-8.151)	-0.234 (-2.478)
1626293	0.03 (1.399)	0.945 (25.11)	N/A	0.098 (1.005)	N/A
1629646	0.007 (0.549)	0.973 (49.293)	N/A	-0.049 (-0.481)	N/A
90928	0.009 (1.676)	1.828 (16.707)	-0.843 (-8.109)	-0.756 (-4.897)	N/A

Table A.23. GARCH Estimate for Ctrip Data

Tour ID	Constant	GARCH{1}	ARCH{1}	Tour ID	Constant	GARCH{1}	ARCH{1}
5106	0.006 (5.160)	0.012 (0.152)	0.835 (6.878)	69925	0.0017 (1.1863)	0.0531 (0.2276)	0.9469 (2.1162)
72528	0.007 (5.017)	N/A	0.818 (4.319)	71821	0.0007 (3.2921)	N/A	1 (3.3277)
73154	0.008 (7.357)	N/A	0.862 (7.465)	72202	0.0006 (3.1514)	0.1246 (1.2138)	0.8308 (3.7012)
80961	0.003 (8.564)	0.030 (1.301)	0.969 (11.318)	72532	0.0009 (3.6252)	0.1225 (1.1569)	0.8208 (3.3828)
29336	0.004 (5.135)	0.027 (0.475)	0.912 (6.118)	72603	0.0015 (1.5559)	N/A	0.9568 (2.607)
23222	0.004 (5.255)	0.044 (0.499)	0.885 (2.223)	72723	0.0011 (6.2937)	N/A	0.9596 (6.9912)
30938	0.006 (6.508)	N/A	0.926 (6.008)	72869	0.0013 (1.7477)	0.0005 (0.0039)	0.9043 (1.9989)
71478	0.004 (9,755)	0.044 (1.136)	0.947 (9.200)	73163	0.0021 (3.3621)	N/A	0.9043 (3.0468)
88292	0.006 (9.229)	N/A	0.886 (8.406)	73186	0.0012 (2.8555)	0.1638 (1.542)	0.7954 (2.6794)
1618693	0.003 (13.963)	N/A	0.971 (14.744)	74983	0.0011 (2.0603)	N/A	0.9137 (1.6786)
56737	0.004 (14.018)	N/A	0.869 (12.676)	78546	0.0017 (1.6543)	0.0464 (0.2101)	0.9536 (2.7679)
71478	0.004 (5.285)	N/A	0.913 (4.834)	79499	0.0038 (1.1013)	N/A	0.7536 (1.3415)
49049	0.006 (7.563)	0.050 (1.623)	0.826 (6.893)	79769	0.0017 (3.1812)	N/A	0.8324 (3.048)
72526	0.004 (4.751)	0.007 (0.119)	0.980 (4.363)	83878	0.0009 (1.4128)	0.1759 (1.1241)	0.8241 (2.4968)
32391	0.006 (7.799)	0.059 (1.753)	0.851 (7.116)	84052	0.0024 (5.4055)	N/A	0.9119 (5.4233)
72527	0.002 (4.235)	0.077 (0.793)	0.923 (4.393)	93877	0.0014 (1.678)	0.0373 (0.1455)	0.9238 (1.8103)
69278	0.005 (4.842)	0.036 (0.542)	0.799 (5.319)	1605378	0.0021 (2.7346)	N/A	0.8492 (2.6402)
28245	0.005 (4.453)	0.025 (0.535)	0.939 (6.209)	1605628	0.0009 (1.2774)	N/A	1 (2.0255)
30814	0.003 (3.478)	N/A	0.991 (3.572)	1607268	0.0019 (2.5127)	0.0603 (0.3925)	0.884 (3.0703)
16311	0.008 (6.033)	N/A	0.798 (6.977)	1608443	0.0035 (1.3152)	N/A	0.935 (1.937)
2753	0.001 (1.882)	N/A	1.000 (2.313)	1609587	0.0016 (2.353)	N/A	0.9447 (2.728)
3150	0.002 (1.82)	N/A	0.792 (1.719)	1612648	0.0001 (0.6483)	N/A	1 (6.4531)
4171	0.001 (4.23)	N/A	0.948 (3.838)	1614660	0.0016 (1.8106)	N/A	1 (2.4261)
5107	0.002 (10.2882)	N/A	0.9234 (10.1634)	1617321	0.0025 (2.442)	N/A	0.9199 (2.9504)
5452	0.0019 (7.0986)	N/A	0.9077 (7.8691)	1620855	0.0033 (3.7142)	N/A	0.9126 (3.7206)
5603	0.0009 (6.1515)	N/A	1 (6.2136)	1621632	0.0017 (3.8826)	0.0283 (0.4139)	0.9378 (5.0486)
6712	0.0006 (2.0965)	0.0297 (0.1961)	0.9703 (2.0889)	1621699	0.0005 (2.0757)	0.2262 (1.9441)	0.7738 (3.026)
7297	0.0011 (2.3348)	0.1562 (1.6258)	0.8438 (2.5729)	1623624	0.001 (1.8025)	N/A	0.9581 (2.9801)
9474	0.0017 (5.156)	N/A	0.8833 (4.6582)	1624666	0.001 (1.7386)	N/A	1 (2.1344)
11626	0.0022 (1.6381)	N/A	0.9377 (1.9449)	1633110	0.0019 (1.7924)	N/A	1 (3.2014)
12043	0.0014 (1.1293)	N/A	0.9367 (2.7765)	85207	0.0007 (2.5508)	0.0795 (0.6781)	0.9167 (3.6131)
17184	0.0008 (5.3356)	N/A	0.995 (5.4798)	85282	0.0011 (2.8981)	N/A	0.9748 (3.0237)
18201	0.0019 (2.3706)	0.0032 (0.0281)	0.9968 (3.3236)	93804	0.0008 (1.8116)	0.0489 (0.3691)	0.9511 (2.4345)
20131	0.0009 (2.5783)	0.0099 (0.101)	0.9901 (3.3902)	60051	0.0029 (2.45)	0.0078 (0.0461)	0.8301 (1.5357)
21148	0.0013 (1.8452)	N/A	0.9066 (1.8333)	44248	0.0015 (0.8866)	0.0571 (0.1468)	0.882 (1.2694)
21509	0.0018 (2.5927)	N/A	0.9703 (2.4633)	69982	0.0017 (0.8935)	N/A	0.9974 (1.2846)
22510	0.0009 (2.4014)	0.1557 (1.1087)	0.8443 (3.6561)	76106	0.0014 (3.2442)	N/A	0.9822 (3.9906)
22809	0.0018 (2.7544)	N/A	0.9741 (2.8198)	1601410	0.0015 (2.5261)	N/A	0.9236 (1.8951)
26561	0.0018 (3.7797)	N/A	0.9538 (4.0525)	11869	0.0013 (4.441)	N/A	1 (4.0114)
28977	0.0016 (3.7397)	0.0309 (0.367)	0.9505 (6.3591)	78511	0.0035 (2.2583)	N/A	0.7385 (2.3533)
44981	0.0016 (1.8887)	N/A	0.9847 (2.8808)	82421	0.0039 (1.172)	N/A	0.8484 (1.1465)
46254	0.0028 (2.8363)	N/A	0.8227 (2.3823)	15618	0.0011 (1.6072)	0.0559 (0.1982)	0.9441 (2.1878)
49063	0.0016 (1.4502)	N/A	0.9766 (2.5992)	36953	0.0019 (1.5647)	N/A	0.924 (1.3806)
49232	0.0014 (2.7849)	0.0329 (0.2037)	0.8994 (3.4561)	66666	0.0028 (3.1064)	N/A	0.9284 (3.1295)
49435	0.0013 (1.9412)	N/A	0.9456 (2.4738)	94870	0.001 (1.3454)	N/A	0.9452 (1.6202)
50327	0.0021 (3.9601)	N/A	0.9469 (3.6537)	91836	0.0013 (2.3764)	N/A	1 (3.7621)
51324	0.0027 (5.9723)	0.0316 (0.5494)	0.8799 (5.1756)	22349	0.0012 (1.0835)	N/A	0.9609 (1.1298)
52165	0.0016 (8.7276)	N/A	0.922 (10.136)	83648	0.0014 (2.2177)	N/A	0.9839 (2.7102)
52950	0.0019 (2.6109)	N/A	1 (3.2394)	44983	0.001 (1.7763)	N/A	0.9906 (2.2808)
53795	0.001 (3.9234)	N/A	0.9577 (3.0225)	71801	0.001 (5.8558)	N/A	0.9402 (3.9024)
55106	0.0015 (1.0498)	0.0247 (0.0737)	0.7127 (1.0407)	58776	0.0019 (1.4994)	N/A	0.9524 (1.7148)
55830	0.0023 (2.7369)	N/A	0.9053 (2.5917)	74359	0.0015 (0.9503)	N/A	0.5003 (1.4076)
60195	0.0018 (1.3274)	N/A	0.9645 (2.4051)	1603059	0.0011 (3.1878)	N/A	1 (2.986)
60639	0.0005 (2.7896)	N/A	1 (4.0139)	76500	0.0008 (1.5869)	0.1829 (0.9449)	0.7943 (2.0277)
63798	0.0021 (1.7739)	0.0837 (0.372)	0.7645 (2.1159)	66306	0.0018 (2.4888)	N/A	0.9505 (2.3495)
66207	0.0003 (0.5155)	0.2565 (0.6184)	0.7435 (0.8259)	1626293	0.0017 (1.4162)	N/A	0.9786 (1.6179)

Tour ID	Constant	GARCH{1}	ARCH{1}	Tour ID	Constant	GARCH{1}	ARCH{1}
66215	0.0013 (2.2661)	N/A	1 (2.1052)	1629646	0.0007 (2.1847)	0.0292 (0.1469)	0.9708 (3.2737)
66287	0.0031 (4.9058)	N/A	0.9106 (4.3512)	90928	0.0016 (2.4831)	N/A	1 (3.5633)
66865	0.0004 (6.9002)	N/A	1 (4.1362)				

APPENDIX B

APPENDIX FOR CHAPTER 3

Proofs for Theorem 1 and 2: Solving the Stochastic Optimal Control Problem

The proofs for Theorem 1 and 2 are presented in this subsection. In this part, we solve the stochastic optimal control problem in equation (B.1).

$$\begin{aligned} \max_{u} \quad & \mathbb{E}[\int_0^\infty e^{-\rho t}(\eta S(t) - cu^2)dt] \\ \text{subject to} \quad & dS(t) = (\beta(x(t) - \mu) + \gamma)dt + \sigma(S(t))dZ(t) \\ & dx(t) = (u\sqrt{b - x(t)} - \mu x(t))dt + \zeta(x(t))dW(t) \end{aligned} \tag{B.1}$$

where u is the control parameter corresponding to any form of effort exerted to improve the online reputation, $S(t)$ is firm's sales rate at time t, and $x(t)$ is the rating at time t. To solve the stochastic optimal control problem, the control u at optimal should satisfy the following Hamilton-Jacobi-Bellman Equation

$$\begin{aligned} \rho V(S,x) = \max_{u}\{(\eta S(t) - cu^2) + V_s[\beta(x(t) - \mu) + \gamma] + V_x[u\sqrt{b - x(t)} - \mu x(t)] \\ + \frac{1}{2}V_{ss}[\sigma(S(t))]^2 + \frac{1}{2}V_{xx}[\zeta(x(t))]^2\} \end{aligned} \tag{B.2}$$

Taking the first order condition with respect to u, we solve the optimal control as

$$u^* = \frac{V_x\sqrt{b - x(t)}}{2c}.$$

We verify a value function following a linear form, that is, $V(S,x) = \lambda_s S + \lambda_x x + \lambda_0$, where $\lambda_s > 0$ and $\lambda_x > 0$. Next, we derive the expressions of λ_s, λ_x, and λ_0 in terms of parameters defined in the model, e.g., η, ρ, β, c, and μ. The partial derivatives of the value function with respect to the state variables are: $V_s = \lambda_s$, $V_x = \lambda_x$, $V_{ss} = 0$, and $V_{xx} = 0$. Substituting u^*, V_s, V_x, V_{ss}, and V_{xx} into equation (B.2) and get

$$\begin{aligned} \rho V(S,x) &= [\eta S(t) - \frac{(V_x)^2(b - x(t))}{4c}] + V_s[\beta(x(t) - \mu) + \gamma] + V_x[\frac{V_x(\sqrt{b - x(t)})^2}{2c} - \mu x(t)] \\ &= \eta S(t) + [-\frac{(\lambda_x)^2}{4} + \lambda_s\beta - \lambda_x\mu]x(t) - \lambda_s(\beta\mu + \gamma) + \frac{(\lambda_x)^2 b}{4} \end{aligned}$$

To summarize, we have

$$\rho V(S,x) = \rho\lambda_s S(t) + \rho\lambda_x x(t) + \rho\lambda_0 = \eta S(t) + [-\frac{(\lambda_x)^2}{4c} + \lambda_s\beta - \lambda_x\mu]x(t) - \lambda_s(\beta\mu+\gamma) + \frac{(\lambda_x)^2 b}{4c}$$

The above equation should hold for arbitrary values of S and x, hence we have

$$\rho\lambda_s = \eta \tag{B.3a}$$

$$\rho\lambda_x = -\frac{(\lambda_x)^2}{4c} + \lambda_s\beta - \lambda_x\mu \tag{B.3b}$$

$$\rho\lambda_0 = -\lambda_s(\beta\mu+\gamma) + \frac{(\lambda_x)^2 b}{4c} \tag{B.3c}$$

Solving the above three equations (taking the positive root for λ_x), we get

$$\lambda_s = \frac{\eta}{\rho} \tag{B.4a}$$

$$\lambda_x = 2c\sqrt{(\rho+\mu)^2 + \frac{\eta\beta}{\rho c}} - 2c(\rho+\mu) \tag{B.4b}$$

$$\lambda_0 = -\frac{\eta\beta\mu}{\rho^2} + \frac{\eta\gamma}{\rho^2} + \frac{bc}{\rho}[\frac{\eta\beta}{\rho c} - 2(\rho+\mu)\sqrt{(\rho+\mu)^2 + \frac{\eta\beta}{\rho c}} + 2(\rho+\mu)^2] \tag{B.4c}$$

Therefore, the optimal control is

$$u^* = \frac{\lambda_x\sqrt{b-x(t)}}{2c} = \alpha\sqrt{b-x(t)},$$

where $\alpha = \sqrt{(\rho+\mu)^2 + \frac{\eta\beta}{\rho c}} - (\rho+\mu)$.

Substituting equation (B.4) into the value function $V(S,x)$, we get the optimal value function $V(S,x)$ as

$$V(S,x) = \lambda_s S + \lambda_x x + \lambda_0 = \frac{\eta}{\rho}S + 2c(A-\rho-\mu)x - \frac{\eta\beta\mu}{\rho^2} + \frac{\eta\gamma}{\rho^2} + \frac{bc}{\rho}[\frac{\eta\beta}{\rho c} - 2(\rho+\mu)A + 2(\rho+\mu)^2] \tag{B.5}$$

where $A = \sqrt{(\rho+\mu)^2 + \frac{\eta\beta}{\rho c}}$. ■

Proof for Theorem 3: Market Equilibrium

A unique equilibrium exists for the mean market rating $\mu = g(\eta_i, \beta_i, c_i, \rho, N, b)$***.***

The mean market rating in equilibrium should satisfy the following equation:

$$\mu = \frac{\sum_{i=1}^{N} \nu_i}{N} = b - \frac{b}{N} \sum_{i=1}^{N} \frac{\mu}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho}.$$

For notation simplicity, define $g(\mu) = b - \frac{b}{N} \sum_{i=1}^{N} \frac{\mu}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho}$, thus we have $g(\mu) = \mu$.

We consider a range of $\mu \in [0,\ b]$. When $\mu = 0$, we have $g(\mu) = b$, and thus $g(\mu) > \mu$. When $\mu = b$, $0 < g(\mu) < b$, thus $g(\mu) < \mu$. Clearly, both functions $g(\mu)$ and μ are continuous in the range $[0,\ b]$. Thus, there exists at least one point where $g(\mu) = \mu$. Therefore, a equilibrium exists.

To prove uniqueness, we use the fact that $g(\mu)$ is monotonically decreasing in μ, but μ increases in μ.

The first derivative of $g(\mu)$ with respect to μ is

$$g'(\mu) = -\frac{b}{N} \sum_{i=1}^{N} \frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}}{\left(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho\right)^2}.$$

In the numerator, we have

$$\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}} > \sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \mu > 0.$$

The denominator is also clearly positive, implying that each term of the summation is positive. Hence, $g'(\mu) < 0$.

To summarize, $g(\mu)$ monotonically decreases with respect to μ. Therefore, there exists only one point where $g(\mu) = \mu$, thus proving the uniqueness of the equilibrium.■

Proof for Proposition 1

The mean market rating in equilibrium changes as the cost of control effort, sales margin, and customers sensitivity change. The parameters (sales margin, customer sensitivity, and cost of control) for the other firms in the market are held constant.

(i) ***For any firm i, as its cost of control effort c_i increases, the mean market rating in equilibrium μ decreases*** $(\frac{d\mu}{dc_i} < 0)$.

(ii) ***For any firm i, as its sales margin η_i increases, the mean market rating in equilibrium μ increases*** $(\frac{d\mu}{d\eta_i} > 0)$.

(iii) ***For any firm i, as the customers are more sensitive to its ratings (higher β_i), the mean market rating in equilibrium μ increases*** $(\frac{d\mu}{d\beta_i} > 0)$.

Define $f(\mu) = b - \mu - \frac{b}{N}\sum_{i=1}^{N} \frac{\mu}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} - \rho}$. According to the implicit function theorem, we have

$$\frac{d\mu}{dc_i} = -\frac{\frac{\partial f}{\partial c_i}}{\frac{\partial f}{\partial \mu}},$$

$$\frac{d\mu}{d\eta_i} = -\frac{\frac{\partial f}{\partial \eta_i}}{\frac{\partial f}{\partial \mu}},$$

$$\frac{d\mu}{d\beta_i} = -\frac{\frac{\partial f}{\partial \beta_i}}{\frac{\partial f}{\partial \mu}},$$

where

$$\frac{\partial f}{\partial \mu} = -1 - \frac{b}{N}\sum_{i=1}^{N} \frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} - \rho)^2} < 0,$$

$$\frac{\partial f}{\partial \beta_i} = -\frac{b\mu}{N} \frac{-\frac{1}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}} \frac{\eta_i}{2\rho c_i}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} - \rho)^2} > 0,$$

$$\frac{\partial f}{\partial \eta_i} = -\frac{b\mu}{N} \frac{-\frac{1}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}\frac{\beta_i}{2\rho c_i}}{(\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho)^2} > 0,$$

$$\frac{\partial f}{\partial c_i} = -\frac{b\mu}{N} \frac{-\frac{1}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}\frac{-\beta_i\eta_i}{2\rho c_i^2}}{(\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho)^2} < 0.$$

Therefore, we have

$$\frac{d\mu}{dc_i} < 0, \frac{d\mu}{d\eta_i} > 0, \frac{d\mu}{d\beta_i} > 0. \blacksquare$$

Proof for Corollary 1

The magnitude of the self-impact of an event affecting a firm is greater than the market-impact of the same event.

Since $N\mu = \sum_{k=1}^{N} \nu_k$, we differentiate both sides with respect to ϕ_i, where ϕ_i is η_i, c_i, or β_i, we get

$$N\frac{d\mu}{d\phi_i} = \frac{d\nu_i}{d\phi_i} + \sum_{k\neq i}^{N} \frac{d\nu_k}{d\phi_i}.$$

Let $\phi_i = \eta_i$. From Proposition 1, we know that $\frac{d\mu}{d\eta_i} > 0$. For $k \neq i$, $\frac{d\nu_k}{d\eta_i} = \frac{\partial \nu_k}{\partial \mu}\frac{d\mu}{d\eta_i}$, where $\frac{d\mu}{d\eta_i} > 0$ (from Proposition 1), and $\frac{\partial \nu_k}{\partial \mu} = -\frac{1}{(\sqrt{(\rho+\mu)^2+\frac{\beta_k\eta_k}{\rho c_k}}-\rho)^2}[(\sqrt{(\rho+\mu)^2+\frac{\beta_k\eta_k}{\rho c_k}} - \rho)\sqrt{(\rho+\mu)^2+\frac{\beta_k\eta_k}{\rho c_k}} - \mu(\rho+\mu)] < 0$. Therefore, $\frac{d\nu_k}{d\eta_i} < 0$ and $\sum_{k\neq i}^{N} \frac{d\nu_k}{d\eta_i} < 0$.

Since $N\frac{d\mu}{d\phi_i} = \frac{d\nu_i}{d\phi_i} + \sum_{k\neq i}^{N} \frac{d\nu_k}{d\phi_i}$, $N\frac{d\mu}{d\phi_i} > 0$, and $\sum_{k\neq i}^{N} \frac{d\nu_k}{d\phi_i} < 0$, we know $\frac{d\nu_i}{d\phi_i}$ has to be positive, and its magnitude has to be greater than the magnitude of $N\frac{d\mu}{d\phi_i}$.

The proofs for $\phi_i = \beta_i$ and $\phi_i = c_i$ follow the same logic. $\blacksquare$

Proof for Proposition 2

The focal firm's profit in equilibrium changes as the parameters (cost of control effort, sales margin, and the customer sensitivity parameter) change. The parameters for the other firms in the market are held constant.

(i) ***For any firm i, as its sales margin η_i increases, its equilibrium profit V_i increases if and only if $N - f_{\eta_i}(N) > 0$.***

(ii) ***For any firm i, as the customers are more sensitive to ratings (higher β_i), its equilibrium profit V_i increases if and only if $N - f_{\beta_i}(N) > 0$.***

(iii) ***For any firm i, as its cost of control effort c_i increases, its equilibrium profit V_i increases if and only if $N - f_{c_i}(N) < 0$.***

The first derivative of V_i with respect to any parameter p_i (c_i, η_i, or β_i) is

$$\frac{dV_i}{dp_i} = \frac{\partial V_i}{\partial p_i} + \frac{\partial V_i}{\partial \mu}\frac{d\mu}{dp_i}.$$

We already know $\frac{d\mu}{dc_i} < 0$, $\frac{d\mu}{d\eta_i} > 0$, and $\frac{d\mu}{d\beta_i} > 0$. Next, we derive and prove $\frac{\partial V_i}{\partial \mu}$. Firm i's equilibrium profit is

$$V_i(S,x) = \frac{\eta_i}{\rho}S + 2c_i(A_i - \rho - \mu)x - \frac{\eta_i\beta_i\mu}{\rho^2} + \frac{\eta_i\gamma_i}{\rho^2} + \frac{bc_i}{\rho}\left[\frac{\eta_i\beta_i}{\rho c_i} - 2(\rho+\mu)A_i + 2(\rho+\mu)^2\right],$$

where $A_i = \sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}$.

$$\frac{\partial V_i}{\partial \mu} = 2c_i x(A_i' - 1) - \frac{\eta_i\beta_i}{\rho^2} + \frac{2bc_i}{\rho}(-A_i - (\rho+\mu)A_i' + 2(\rho+\mu)),$$

where $A_i' = \frac{(\rho+\mu)}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}$. Hence, $(A_i' - 1) < 0$. For notation simplicity, we define

$$f = -A_i - (\rho+\mu)A_i' + 2(\rho+\mu) = -\frac{2(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}} + 2(\rho+\mu).$$

The first derivative of f with respect to η_i is

$$\frac{df}{d\eta_i} = -\frac{\beta_i}{\rho c_i}\frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} + \frac{2(\rho+\mu)^2+\frac{\eta_i\beta_i}{\rho c_i}}{2\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}}{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}} < 0.$$

Thus, the value of f decreases as η_i increases. As $\eta_i = 0$, we know $f = 0$. As $\eta_i > 0$, we have $f < 0$.

Therefore, $\frac{\partial V}{\partial \mu} = 2c_i x(A_i' - 1) - \frac{\eta_i \beta_i}{\rho^2} + \frac{2bc_i}{\rho} f < 0$. ■

Next, we derive the partial first derivative of V_i with respect to c_i. The value function can be rewritten as

$$V_i(S, x) = \frac{\eta_i}{\rho} S + \frac{\beta_i \eta_i}{\rho^2}(b - \mu) + 2c_i(A_i - \rho - \mu)(x - b\frac{\rho + \mu}{\rho}) + \frac{\eta_i \gamma_i}{\rho},$$

where $A_i = \sqrt{(\mu + \rho)^2 + \frac{\beta_i \eta_i}{\rho c_i}}$. The partial first derivative of V_i with respect to c_i is

$$\frac{\partial V_i}{\partial c_i} = 2(x - b\frac{\rho + \mu}{\rho})\frac{1}{2A_i}(2A_i(A_i - \rho - \mu) - \frac{\beta_i \eta_i}{\rho c_i}).$$

Since $(x - b\frac{\rho+\mu}{\rho}) < 0$ and $(2A_i(A_i - \rho - \mu) - \frac{\beta_i \eta_i}{\rho c_i}) > 0$, we thus get $\frac{\partial V_i}{\partial c_i} < 0$. ■

To summarize, the first derivative of V_i with respect to c_i is

$$\frac{dV_i}{dc_i} = \frac{\partial V_i}{\partial c_i} + \frac{\partial V_i}{\partial \mu}\frac{d\mu}{dc_i},$$

where $\frac{\partial V_i}{\partial c_i} = 2(x - b\frac{\rho+\mu}{\rho})\frac{1}{2A_i}(2A_i(A_i - \rho - \mu) - \frac{\beta_i \eta_i}{\rho c_i})$, $\frac{\partial V_i}{\partial \mu} = 2c_i x(A_i' - 1) - \frac{\eta_i \beta_i}{\rho^2} + \frac{2bc_i}{\rho}(-A_i - (\rho + \mu)A_i' + 2(\rho + \mu))$, $\frac{d\mu}{dc_i} = -(\frac{b\mu}{N}\frac{\frac{1}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}\frac{\beta_i \eta_i}{2\rho c_i^2}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho)^2})/(1 + \frac{b}{N}\sum_{i=1}^{N}\frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho)^2})$, and $A_i' = \frac{(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}$.

We have proved that $\frac{\partial V_i}{\partial c_i} < 0$ and $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{dc_i} > 0$. Depending on the magnitude of the term $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{dc_i}$, the value of $\frac{dV_i}{dc_i}$ can stay at negative or become positive.

The firm i's value function is increasing with respect to c_i if and only if $N < f_{c_i}(N)$,

where $f_{c_i}(N) = \frac{b\mu BD}{C} - b\sum_{i=1}^{N}\frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho)^2}$,

$C = 2(x - b\frac{\rho+\mu}{\rho})\frac{1}{2A_i}(2A_i(A_i - \rho - \mu) - \frac{\beta_i \eta_i}{\rho c_i})$,

$B = 2c_i x(A_i' - 1) - \frac{\eta_i \beta_i}{\rho^2} + \frac{2bc_i}{\rho}(-A_i - (\rho + \mu)A_i' + 2(\rho + \mu))$,

and $D = \frac{\frac{1}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}\frac{\beta_i \eta_i}{2\rho c_i^2}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho)^2}$. ■

Next, we derive the partial first derivative of V_i with respect to η_i. The partial first derivative of V_i with respect to η_i is

$$\frac{\partial V_i}{\partial \eta_i} = \frac{S}{\rho} + \frac{\beta_i(b-\mu)}{\rho^2} + \frac{\beta_i}{\rho A_i}(x - b\frac{\rho+\mu}{\rho}) + \frac{\gamma_i}{\rho},$$

where $A_i = \sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}$. Since a firm's value function should always be non-negative, we have $\lambda_0 > 0$, which can be rewritten as

$$\eta_i(\frac{\beta_i(b-\mu)}{\rho^2} - \frac{2c_i}{\eta_i} b\frac{\rho+\mu}{\rho}(A_i - \rho - \mu) + \frac{\gamma_i}{\rho}) > 0.$$

Since the initial state variables $x \geq 0$, $S \geq 0$, and $\frac{\partial V_i}{\partial \eta_i}$ is increasing with respect to x and S separately, we have

$$\frac{\partial V_i}{\partial \eta_i} = \frac{S}{\rho} + \frac{\beta_i(b-\mu)}{\rho^2} + \frac{\beta_i}{\rho A_i}(x - b\frac{\rho+\mu}{\rho}) + \frac{\eta_i \gamma_i}{\rho} \geq \frac{\beta_i(b-\mu)}{\rho^2} - b\frac{\rho+\mu}{\rho}\frac{\beta_i}{\rho A_i} + \frac{\gamma_i}{\rho}.$$

Define $g(\beta_i) = 2c_i \frac{A_i - \rho - \mu}{\eta_i} - \frac{\beta_i}{\rho A}$. The first derivative of $g(\beta_i)$ with respect to β_i is

$$g'(\beta_i) \sim (2\rho c_i A_i' A_i - \eta_i) A_i + \beta_i \eta_i A_i' = \eta_i A_i + \beta_i \eta_i A_i' > 0.$$

Since $g(\beta_i = 0) = 0$, we have $g(\beta_i) \geq 0$ and thus $2c_i \frac{A_i - \rho - \mu}{\eta_i} \geq \frac{\beta_i}{\rho A_i}$. Therefore,

$$\frac{\beta_i(b-\mu)}{\rho^2} - b\frac{\rho+\mu}{\rho}\frac{\beta_i}{\rho A_i} + \frac{\gamma_i}{\rho} \geq \frac{\beta_i(b-\mu)}{\rho^2} - b\frac{\rho+\mu}{\rho} 2c_i \frac{A_i - \rho - \mu}{\eta} + \frac{\gamma_i}{\rho} > 0.$$

Thus, we have $\frac{\partial V_i}{\partial \eta_i} > 0$. To summarize, the first derivative of V_i with respect to η_i is

$$\frac{dV_i}{d\eta_i} = \frac{\partial V_i}{\partial \eta_i} + \frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\eta_i}.$$

We have proved that $\frac{\partial V_i}{\partial \eta_i} > 0$ and $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\eta_i} < 0$. Depending on the magnitude of the term $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\eta_i}$, the value of $\frac{dV_i}{d\eta_i}$ can stay at positive or become negative.

The firm i's value function is increasing with respect to η_i if and only if $N > f_{\eta_i}(N)$, where $f_{\eta_i}(N) = \frac{b\mu BE}{G} - b\sum_{i=1}^{N} \frac{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho - \frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}}}}{(\sqrt{(\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i}} - \rho)^2}$,

$G = \frac{S}{\rho} + \frac{\beta_i(b-\mu)}{\rho^2} + \frac{\beta_i}{\rho A_i}(x - b\frac{\rho+\mu}{\rho}) + \frac{\gamma_i}{\rho}$,

$B = 2c_i x(A_i' - 1) - \frac{\eta_i \beta_i}{\rho^2} + \frac{2bc_i}{\rho}(-A_i - (\rho+\mu)A_i' + 2(\rho+\mu))$,

and $E = \frac{\frac{1}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}\frac{\beta_i}{2\rho c_i}}{(\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho)^2}$. ■

Similarly, the partial first derivative of V_i with respect to β_i is

$$\frac{\partial V_i}{\partial \beta_i} = \frac{\eta_i(b-\mu)}{\rho^2} + \frac{\eta_i}{\rho A_i}(x - b\frac{\rho+\mu}{\rho}) > 0.$$

To summarize, the first derivative of V_i with respect to β_i is

$$\frac{dV_i}{d\beta_i} = \frac{\partial V_i}{\partial \beta_i} + \frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\beta_i}.$$

We have proved that $\frac{\partial V_i}{\partial \beta_i} > 0$ and $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\beta_i} < 0$. Depending on the magnitude of the term $\frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\beta_i}$, the value of $\frac{dV_i}{d\beta_i}$ can stay at positive or become negative.

The firm i's value function is increasing with respect to β_i if and only if $N > f_{\beta_i}(N)$,

where $f_{\beta_i}(N) = \frac{b\mu BF}{H} - b\sum_{i=1}^{N}\frac{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho-\frac{\mu(\rho+\mu)}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}}{(\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho)^2}$,

$H = \frac{\eta_i(b-\mu)}{\rho^2} + \frac{\eta_i}{\rho A_i}(x - b\frac{\rho+\mu}{\rho})$,

$B = 2c_i x(A_i' - 1) - \frac{\eta_i\beta_i}{\rho^2} + \frac{2bc_i}{\rho}(-A_i - (\rho+\mu)A_i' + 2(\rho+\mu))$,

and $F = \frac{\frac{1}{\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}}\frac{\eta_i}{2\rho c_i}}{(\sqrt{(\rho+\mu)^2+\frac{\beta_i\eta_i}{\rho c_i}}-\rho)^2}$. ■

Proof for Proposition 3

The entry of a firm has consequences on the incumbent firms, the platform and the consumers (mean market rating).

(i) ***The entry of a high rating firm lowers the profits of the the incumbent firms. It increases the mean market rating.***

Suppose a firm enters a market with mean μ, where $\mu = \frac{1}{N}\sum_{i=1}^{N}\nu_i = \frac{S_N}{N}$.

Its rating after entry is ν_l, where $\nu_l > \mu = \frac{S_N}{N}$, and the "new" market rating is denoted as $\mu' = \frac{S'_{N+1}}{N+1} = \frac{S'_N+\nu_l}{N+1}$. We prove $\mu' > \mu$ by contradiction.

Let $\mu' \leq \mu$. We know that $\frac{d\nu_i}{d\mu} < 0$, $\nu_i = \frac{A_i-(\rho+\mu)}{A_i-\rho}b$, and $A_i = \sqrt{(\rho+\mu)^2 + \frac{\eta_i\beta_i}{\rho c_i}}$. Therefore, we get $S'_N \geq S_N$.

Since $\mu' \leq \mu$, We have

$$\frac{S'_{N+1}}{N+1} \leq \frac{S_N}{N},$$

which can be further written as

$$\frac{S'_N + \nu_l}{N+1} \leq \frac{S_N}{N}.$$

Since $\nu_l > \frac{S_N}{N}$, we have

$$\frac{S'_N + \nu_l}{N+1} > \frac{S'_N + \frac{S_N}{N}}{N+1}.$$

By the chain rule, we get

$$\frac{S'_N + \frac{S_N}{N}}{N+1} < \frac{S_N}{N},$$

which can be simplified as

$$NS'_N + S_N < (N+1)S_N,$$

and further

$$S'_N < S_N,$$

which contradict with the fact that $S'_N \geq S_N$. $\mu' \leq \mu$ is not true, and thus we prove that $\mu' > \mu$. ■

When a high rating firm enters the market, the mean market rating must increase. As a result, the profits of the incumbent firms in the market must go down (since $\frac{\partial V_i}{\partial \mu} < 0$). However, if we include the profit of the new entrant, the total market profit could go up or down. It is easy to demonstrate this with a numerical example.

(ii) ***The entry of a low rating firm increases the total profit of the market and also the profits of the incumbent firms. However it lowers the mean market rating.***

The proof of the entry of a low rating firm lowers the mean market rating ($\nu_l < \mu = \frac{S_N}{N}$) follows the same logic of a high rating firm in (i). We are going to prove $\mu' < \mu$. Let $\mu' \geq \mu$. Due to $\frac{d\nu_i}{\mu} < 0$, we know $S'_N \leq S_N$. Since $\mu' \geq \mu$, we have

$$\frac{S'_{N+1}}{N+1} = \frac{S'_N + \nu_l}{N+1} \geq \frac{S_N}{N}.$$

Since $\nu_l < \frac{S_N}{N}$, we have

$$\frac{S'_N + \frac{S_N}{N}}{N+1} > \frac{S_N}{N},$$

which can be further simplified as

$$S'_N > S_N,$$

contradicting with the fact that $S'_N \leq S_N$. Therefore, $\mu' < \mu$. ■

As proved above, the entry of a low rating firm reduces the market equilibrium. Since $\frac{\partial V_i}{\partial \mu} < 0$, we know that this increases the profits of the incumbent firms. If we include the new entrant, it is clear that the total profit of firms in the market must increase.

(iii) ***An average firm's entry increases the total profit of the market and leaves the incumbent firms and the market mean unaffected.***

The above assertion can easily be argued to be true. Since the new entrant has average rating, it will not change the mean market rating. Thus, the profits of the incumbent firms will not be affected. However, the new entrant will bring some positive profit to increase the total market profit.

Proof for Corollary 3

The exit of a firm has consequences for the existing (incumbent) firms, the platform and the consumers.

(i) ***The exit of a high rating firm lowers the mean market rating, and increases the profits of the incumbent firms.***

The proof concerning the exit of a firm follows the same logic. The mean market rating is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \nu_i.$$

Assume a firm with higher than the mean market rating ($\nu_i > \frac{S_{N-1}}{N-1}$) exits. After the firm exits, the "new" mean market rating is μ'. We prove $\mu' < \mu$ by contradiction.

Let $\mu' \geq \mu$. We denote the summation of the remaining firms' ratings as S'_{N-1} after the exit and $\mu' = \frac{S'_{N-1}}{N-1}$.

Since $\mu' \geq \mu$ and $\frac{d\nu_i}{d\mu} < 0$, we have $S'_{N-1} \leq S_{N-1}$.

Since $\mu' \geq \mu$ and $\nu_l > \frac{S_{N-1}}{N-1}$, we have

$$\frac{S'_{N-1}}{N-1} \geq \frac{S_{N-1} + \nu_l}{N},$$

$$\frac{S'_{N-1}}{N-1} > \frac{S_{N-1} + \frac{S_{N-1}}{N-1}}{N},$$

which can be further simplified as

$$S'_{N-1} > S_{N-1},$$

contradicting with the fact that $S'_{N-1} \leq S_{N-1}$. Therefore, $\mu' < \mu$. ■

When a high rating firm exits the market, the mean market rating must decreases. As a result, the profits of the incumbent firms in the market must go up (since $\frac{\partial V_i}{\partial \mu} < 0$). However, if we exclude the profit of the exiting firm, the total market profit could go up or down. It is easy to demonstrate this with a numerical example.

(ii) ***The exit of a low rating firm increases the mean market rating, lowers the profits of incumbent firms, and lowers the total profit of the market.***

The proof for the low rating firm ($\nu_l < \frac{S_{N-1}}{N-1}$) follows the similar logic as for the high rating firm. We want to show $\mu' > \mu$.

Let $\mu' \leq \mu$. Since $\mu' \leq \mu$ and $\frac{d\nu_i}{d\mu} < 0$, we have $S'_{N-1} \geq S_{N-1}$.

Since $\mu' \leq \mu$ and $\nu_l < \frac{S_{N-1}}{N-1}$, we have

$$\frac{S'_{N-1}}{N-1} \leq \frac{S_{N-1} + \nu_l}{N} < \frac{S_{N-1} + \frac{S_{N-1}}{N-1}}{N},$$

which can be further simplified as

$$S'_{N-1} < S_{N-1},$$

contradicting with the fact that $S'_{N-1} \geq S_{N-1}$. Therefore, $\mu' > \mu$. ■

As proved above, the exit of a low rating firm increases the market rating in equilibrium. Since $\frac{\partial V_i}{\partial \mu} < 0$, we know that this decreases the profits of the incumbent firms. If we exclude the exiting firm, it is clear that the total profit of firms in the market must decrease.

(iii) ***The exit of an average firm leaves the incumbent firms and the market mean rating unaffected, and decreases the total profit of the market.***

The above assertion can easily be argued to be true. Since the exiting firm has average rating, it will not change the mean market rating. Thus, the profits of the incumbent firms will not be affected. However, the exiting firm will reduce the total market profit.

The Proof for Proposition 4

A market with more heterogeneous firms (one where the parameters of the firms are very different) leads to a lower mean market rating in equilibrium.

Define $\delta_i = \frac{\beta_i \eta_i}{\rho c_i}$. $\delta_i = \bar{\delta} + (i - \frac{N+1}{2})\alpha$, $i = 1, 2, ..., N$, and N is odd. The parameter α represents the heterogeneous parameter. As this parameter increases, the firms in the market are more heterogeneous. Let

$$f(\mu) = b - \mu - \frac{b}{N} \sum_{i=1}^{N} \frac{\mu}{\sqrt{(\rho + \mu)^2 + \delta_i} - \rho}.$$

We have

$$\frac{\partial f}{\partial \alpha} = \frac{b\mu}{N} \sum_{i=1}^{N} \frac{\frac{i - \frac{N+1}{2}}{2\sqrt{(\rho+\mu)^2 + \frac{\bar{\delta} + (i - \frac{N+1}{2})\alpha}{\rho}}}}{(\sqrt{(\rho + \mu)^2 + \frac{\bar{\delta} + (i - \frac{N+1}{2})\alpha}{\rho}} - \rho)^2}$$

$$= \frac{b\mu}{2N} \sum_{i=1}^{N} \frac{i - \frac{N+1}{2}}{(\sqrt{(\rho + \mu)^2 + \frac{\bar{\delta} + (i - \frac{N+1}{2})\alpha}{\rho}} - \rho)^2 \sqrt{(\rho + \mu)^2 + \frac{\bar{\delta} + (i - \frac{N+1}{2})\alpha}{\rho}}}.$$

In the numerator, when $i < \frac{N+1}{2}$, the ith term is negative and otherwise positive, but symmetric. That is, each term in the numerator with a negative sign has a corresponding

term of the same magnitude, but with a positive sign. However, the denominator increases for higher values of i. Therefore, the negative terms dominate the positive terms, and we have

$$\frac{\partial f}{\partial \alpha} < 0.$$

We also know that $\frac{\partial f}{\partial \mu} < 0$. Then, according to the implicit function theorem, we have

$$\frac{d\mu}{d\alpha} = -\frac{\frac{\partial f}{\partial \alpha}}{\frac{\partial f}{\partial \mu}} < 0. \blacksquare$$

The Proof for Proposition 5

The total profit of the firms in the market increases with an increase in the extent of firm heterogeneity in the market.

We first consider the case where heterogeneity comes from the sales margin. Let

$$\eta_i = \bar{\eta} + (i - \frac{N+1}{2})\alpha,$$

where $i = 1, 2, ..., N$ and N is odd. We have already seen that the mean market rating in equilibrium reduces with the heterogeneity parameter α. Since the middle firm's parameter does not change and the market mean rating reduces, its profit will increase.

Next, take a pair of firms: $i = \frac{N+1}{2} \pm k\alpha$. These firms will experience an increase and a reduction in the sales margin, i.e., denote by $\eta_1 = \bar{\eta} - k\alpha$ and $\eta_2 = \bar{\eta} + k\alpha$. We ask: What will be the impact of an increase in α on the sum of the profits of these firms, denoted by $\Pi = V_1 + V_2$?

For a given mean market rating μ, the second derivative of the profit w.r.t sales margin is

$$\frac{\partial^2 V_i}{\partial \eta_i^2} = 2c_i(x - b\frac{\rho+\mu}{\rho})(-\frac{\beta_i}{4\rho c_i})((\rho+\mu)^2 + \frac{\beta_i \eta_i}{\rho c_i})^{-1.5} > 0.$$

Thus, the profit is a convex function with respect to sales margin (η_i).

Assume the mean market rating stays the same, we have

$$\frac{\partial \Pi}{\partial \alpha} = \frac{\partial V_1}{\partial \alpha} + \frac{\partial V_2}{\partial \alpha} = \frac{\partial V_1}{\partial \eta_1}\frac{d\eta_1}{d\alpha} + \frac{\partial V_2}{\partial \eta_2}\frac{d\eta_2}{d\alpha} = \frac{\partial V_1}{\partial \eta_1}(-k) + \frac{\partial V_2}{\partial \eta_2}k$$

As the extent of heterogeneity becomes larger (higher α), η_2 increases and η_1 decreases symmetrically. We have $\frac{\partial V_2}{\partial \eta_2} > \frac{\partial V_1}{\partial \eta_1} > 0$ due to the convex property. Hence, $\frac{\partial \Pi}{\partial \alpha} > 0$ given μ.

We have proved that increasing the heterogeneity parameter α reduces the market mean μ. This would further increase the profit. To summarize, we have

$$\frac{d\Pi}{d\alpha} = \frac{\partial \Pi}{\partial \alpha} + \frac{\partial \Pi}{\partial \mu}\frac{d\mu}{d\alpha} > 0.$$

When the source of heterogeneity comes from customer sensitivity, that is, $\beta_i = \bar{\beta} + (i - \frac{N+1}{2})\alpha$, the proof is very similar to the proof on η_i. When the source of heterogeneity comes from the cost of the control, that is, $c_i = \bar{c} + (i - \frac{N+1}{2})\alpha$. The proof is similar. Take a pair of firms, denoted by $c_1 = \bar{c} - k\alpha$ and $c_2 = \bar{c} + k\alpha$.

For a given mean market rating μ, the second derivative of the profit w.r.t cost of control (c_i) is

$$\frac{\partial^2 V_i}{\partial c_i^2} = 2(x - b\frac{\rho+\mu}{\rho})(-\frac{\beta_i\eta_i}{\rho\sqrt{(\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i}}})\frac{\frac{\beta_i\eta_i}{4\rho c_i}}{c_i^2((\rho+\mu)^2 + \frac{\beta_i\eta_i}{\rho c_i})} > 0.$$

Thus, the profit is a convex function with respect to the cost of control (c_i). We have $\frac{\partial V_1}{\partial c_1} < \frac{\partial V_2}{\partial c_2} < 0$ due to the convex property. Therefore, we have

$$\frac{\partial \Pi}{\partial \alpha} = \frac{\partial V_1}{\partial \alpha} + \frac{\partial V_2}{\partial \alpha} = \frac{\partial V_1}{\partial c_1}\frac{dc_1}{d\alpha} + \frac{\partial V_2}{\partial c_2}\frac{dc_2}{d\alpha} = \frac{\partial V_1}{\partial c_1}(-k) + \frac{\partial V_2}{\partial c_2}k > 0.$$

We have proved that increasing the heterogeneity parameter α reduces the market mean μ. This would further increase the profit. To summarize, we have

$$\frac{d\Pi}{d\alpha} = \frac{\partial \Pi}{\partial \alpha} + \frac{\partial \Pi}{\partial \mu}\frac{d\mu}{d\alpha} > 0.$$

The above results hold for all values of k. That is, the sum of the profits must increase for any two firms that are symmetric across the middle. Hence, the total profit of the market

must increase with an increase with the heterogeneity in the sales margin or the cost. The impact of an increase in the heterogeneity of the sensitivity parameter is similar (in sign) to that of an increase in heterogeneity of the sales margin parameter. ■

Impact of the heterogeneity on firm i's profit

In this part, we show that $\frac{dV_i}{d\alpha} > 0$ when $i > \frac{N+1}{2}$. Define $\eta_i = \bar{\eta} + (i - \frac{N+1}{2})\alpha$, we have

$$\frac{dV_i}{d\alpha} = \frac{\partial V_i}{\partial \alpha} + \frac{\partial V_i}{\partial \mu}\frac{d\mu}{d\alpha} + \frac{\partial V_i}{\partial \nu_i}\frac{d\nu_i}{d\alpha}.$$

$V_i(S,x) = \frac{\eta_i S}{\rho} + 2c_i(A_i - \rho - \mu)x + \frac{\eta_i \beta_i}{\rho^2}(b-\mu) + \frac{bc_i}{\rho}(-2(\rho+\mu)A_i + 2(\rho+\mu)^2)$, where $A_i = \rho + \frac{\mu b}{b-\nu_i}$.

$$\frac{\partial V_i}{\partial \alpha} = \frac{S}{\rho}\left(i - \frac{N+1}{2}\right) + \frac{\beta_i}{\rho^2}(b-\mu)(i - \frac{N+1}{2}) > 0.$$

We know that

$$\frac{\partial V_i}{\partial \mu} < 0, \frac{d\mu}{d\alpha} < 0, \frac{\partial V_i}{\partial v_i} > 0, \frac{\partial \nu_i}{\partial \alpha} > 0$$

$$\nu_i = \frac{A_i - \rho - \mu}{A_i - \rho} b = (1 - \frac{\mu}{A_i - \rho})b.$$

Combining all the above results, we have

$$\frac{dV_i}{d\alpha} > 0.$$

Similarly, it can be numerically verified that when $i < \frac{N+1}{2}$, then $\frac{dV}{d\alpha} < 0$ or $\frac{dV}{d\alpha} > 0$.

Expectation of Sales Rate and Rating

In this section, we derive the mean of $S_i(t)$ given the initial states of $S_i(t_0)$ and $x_i(t_0)$. We consider the stochastic magnitudes following the formulas $\sigma_i(S_i(t)) = \sigma_i S_i(t)$ and $\zeta_i(x_i(t)) = \zeta_i\sqrt{b - x_i(t)}$. The rating equation is similar to the form of a CIR process. To apply the property of the CIR process, we define $x_i'(t) = b - x_i(t)$ and $\mu' = b - \mu$, where b is the upper bound of ratings. We consider a two-dimensional stochastic process as

$$dS_i(t) = (\beta_i(\mu' - x_i'(t)) + \gamma_i)dt + \sigma_i S_i(t) dZ_i(t) \tag{B.6a}$$

$$dx_i'(t) = \lambda_i(\nu_i' - x_i'(t))dt + \zeta_i\sqrt{x_i'(t)}dW_i(t) \tag{B.6b}$$

where $\nu_i' = b - \nu_i = b - \frac{(A_i-\rho-\mu)}{(A_i-\rho)}b$, $\lambda_i = A_i - \rho$, and $A_i = \sqrt{(\rho+\mu)^2 + \frac{\eta_i\beta_i}{\rho c_i}}$. For equation (B.6a), after integrating on both sides, we have

$$S_i(t) - S_i(t_0) = \int_{t_0}^{t} (\beta_i(\mu' - x_i'(v)) + \gamma_i)dv + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v).$$

Thus,

$$S_i(t) = S_i(t_0) + \int_{t_0}^{t} (\beta_i(\mu' - x_i'(v)) + \gamma_i)dv + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v) \tag{B.7}$$

Similarly, we integrate equation (B.6b) on both sides and get:

$$x_i'(t) = e^{-\lambda_i(t-t_0)}x_i'(t_0) + \nu_i'(1 - e^{-\lambda_i(t-t_0)}) + \zeta_i \int_{t_0}^{t} e^{-\lambda_i(t-u)}\sqrt{x_i'(u)}dW_i(u) \tag{B.8}$$

Equation (B.8) is the analytical solution for the stochastic process $x_i'(t)$. To further simplify equation (B.7), we substitute equation (B.8) into $S_i(t)$ and get:

$$\begin{aligned}
S_i(t) &= S_i(t_0) + \int_{t_0}^{t} (\beta_i(\mu' - x_i'(v)) + \gamma_i)dv + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v) \\
&= S_i(t_0) + \beta_i \int_{t_0}^{t} \mu' dv + \int_{t_0}^{t} \gamma_i dv - \beta_i \int_{t_0}^{t} x_i'(v)dv + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v) \\
&= -\beta_i \int_{t_0}^{t} [e^{-\lambda_i(v-t_0)}x_i'(t_0) + \nu_i'(1 - e^{-\lambda_i(v-t_0)}) + \zeta_i \int_{t_0}^{t} e^{-\lambda_i(v-u)}\sqrt{x_i'(u)}dW_i(u)]dv \\
&\quad + \beta_i\mu'(t-t_0) + \gamma_i(t-t_0) + S_i(t_0) + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v) \\
&= -\beta_i \int_{t_0}^{t} [e^{-\lambda_i(v-t_0)}x_i'(t_0) + \nu_i'(1 - e^{-\lambda_i(v-t_0)})]dv - \beta_i \int_{t_0}^{t} [\zeta_i \int_{t_0}^{v} e^{-\lambda_i(v-u)}\sqrt{x_i'(u)}dW_i(u)]dv \\
&\quad + \beta_i\mu'(t-t_0) + \gamma_i(t-t_0) + S_i(t_0) + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v) \\
&= -\frac{\beta_i(e^{-\lambda_i(t-t_0)} - 1)(\nu_i' - x_i'(t_0))}{\lambda_i} - \beta_i\nu_i'(t-t_0) - \beta_i\zeta_i \int_{t_0}^{t} [\int_{t_0}^{v} e^{-\lambda_i(v-u)}\sqrt{x_i'(u)}dW_i(u)]dv \\
&\quad + \beta_i\mu'(t-t_0) + \gamma_i(t-t_0) + S_i(t_0) + \sigma_i \int_{t_0}^{t} S_i(v)dZ_i(v)
\end{aligned} \tag{B.9}$$

Next, we simplify the term $\beta_i\zeta_i \int_{t_0}^{t}\int_{t_0}^{v} e^{-\lambda_i(v-u)}\sqrt{x_i'(u)}dW_i(u)dv$ in equation (B.9). By integration by parts, we have

$$\begin{aligned}
&\beta_i\zeta_i \int_{t_0}^{t}\int_{t_0}^{v} e^{-\lambda_i(v-u)}\sqrt{x_i'(u)}dW_i(u)dv \\
&= \beta_i\zeta_i \int_{t_0}^{t} e^{-\lambda_i v}\int_{t_0}^{v} e^{\lambda_i u}\sqrt{x_i'(u)}dW_i(u)dv \\
&= \frac{\beta_i\zeta_i}{-\lambda_i}\int_{t_0}^{t}\int_{t_0}^{v} e^{\lambda_i u}\sqrt{x_i'(u)}dW_i(u)d_v(e^{-\lambda_i v}) \\
&= \frac{\beta_i\zeta_i}{-\lambda_i}[e^{-\lambda_i t}\int_{t_0}^{t} e^{\lambda_i u}\sqrt{x_i'(u)}dW_i(u) - \int_{t_0}^{t} e^{-\lambda_i v}d(\int_{t_0}^{v} e^{\lambda_i u}\sqrt{x_i'(u)}dW_i(u))] \\
&= \frac{\beta_i\zeta_i}{-\lambda_i}\int_{t_0}^{t}(e^{-\lambda_i t} - e^{-\lambda_i v})d(\int_{t_0}^{v} e^{\lambda_i u}\sqrt{x_i'(u)}dW_i(u)) \\
&= \frac{\beta_i\zeta_i}{-\lambda_i}\int_{t_0}^{t}(e^{-\lambda_i t+\lambda_i v} - 1)\sqrt{x_i'(v)}dW_i(v).
\end{aligned} \tag{B.10}$$

Substituting equation (B.10) back into $S_i(t)$ in equation (B.9), we get the analytical solution for $S_i(t)$ in terms of the parameters ρ, η_i, β_i, c_i, μ, γ_i, σ_i, and ζ_i as well as the initial states of $S_i(t_0)$ and $x_i'(t_0)$.

$$\begin{aligned}
S_i(t) &= \beta_i(\mu' - \nu_i')(t-t_0) + \gamma_i(t-t_0) - \frac{\beta_i(e^{-\lambda_i(t-t_0)} - 1)(\nu_i' - x_i(t_0))}{\lambda_i} \\
&+\frac{\beta_i\zeta_i}{\lambda_i}\int_{t_0}^{t}(e^{-\lambda_i t+\lambda_i v} - 1)\sqrt{x_i'(v)}dW_i(v) + S_i(t_0) + \sigma_i\int_{t_0}^{t} S_i(v)dZ_i(v)
\end{aligned} \tag{B.11}$$

From equation (B.11), we take expectation on both sides and get the mean of sales at time t given the initial states of $S_i(t_0)$ and $x_i'(t_0)$ as

$$\begin{aligned}
\mathbb{E}[S_i(t)|S_i(t_0), x_i'(t_0)] = \beta_i(\mu' - \nu_i')(t-t_0) + \gamma_i(t-t_0) - \frac{\beta_i(e^{-\lambda_i(t-t_0)} - 1)(\nu_i' - x_i'(t_0))}{\lambda_i} \\
+S_i(t_0)
\end{aligned} \tag{B.12}$$

Substituting $x_i'(t_0)$ with $b - x_i(t_0)$, μ_1' with $b - \mu_1$, and ν_i' with $b - \nu_i$ in equation (B.12), we get the mean of sales at time t given the initial states of $S_i(t_0)$ and $x_i(t_0)$ as

$$\mathbb{E}[S_i(t)|S_i(t_0), x_i(t_0)] = \beta_i(\nu_i - \mu)(t-t_0) + \gamma_i(t-t_0) - \frac{\beta_i(e^{-\lambda_i(t-t_0)} - 1)(x_i(t_0) - \nu_i)}{\lambda_i} + S_i(t_0). \blacksquare$$

The mean of rating at time t given the initial state of $x_i(t_0)$ is

$$\mathbb{E}[x_i(t)|x_i(t_0)] = b - (b - x_i(0))e^{-\lambda_i(t-t_0)} - \nu_i'(1 - e^{-\lambda_i(t-t_0)}).\blacksquare$$

APPENDIX C

APPENDIX FOR CHAPTER 4

Proof for Theorem 4

To prove the optimal solution, let $g(\delta)$ denote the terms involving the control variable in the HJB as shown below.

$$g(\delta) = \frac{-\alpha a}{\sqrt{\delta_0 + \delta}} + \frac{\gamma v}{\delta_0 + \delta}$$

The first derivative yields $g'(\delta) = \frac{1}{2(\delta_0+\delta)^{\frac{3}{2}}}\left(\alpha a - \frac{2\gamma v}{\sqrt{\delta_0+\delta}}\right)$. The second derivative yields $g''(\delta) = \frac{1}{(\delta_0+\delta)^{\frac{5}{2}}}\left(-\frac{3}{4}\alpha a + \frac{2\gamma v}{\sqrt{\delta_0+\delta}}\right)$.

The optimal solution is a corner solution as shown below. Let the first derivative be equal to zero, we get $\delta = \delta_1$. Solving for $\delta = \delta_1$ yields a minimum because $g''(\delta_1) > 0$. The value of δ_1 is unique because it is the only solution of the first order condition. Hence, the optimal solution can be found by evaluating the value of the function $g(\delta)$ at the two extreme points $(0, \delta_m)$ and choosing the higher of the two values. Let $g(\delta_m) \geq g(0)$, after rewriting, we get $\frac{a}{v} \geq \left(\frac{\gamma}{\alpha}\right)\left(\frac{\frac{1}{\delta_0} - \frac{1}{\delta_0+\delta_m}}{\frac{1}{\sqrt{\delta_0}} - \frac{1}{\sqrt{\delta_0+\delta_m}}}\right)$.

REFERENCES

Ammann, M. and N. Schaub (2016). Social interaction and investing: Evidence from an online social trading network. *Working Paper, University of St. Gallen, Germany*.

Anderson, C. K. (2012a). The impact of social media on lodging performance. *Cornell Hospitality Report 12*(15), 6–11.

Anderson, C. K. and S. Han (2016). Hotel performance impact of socially engaging with consumers. *Cornell Hospitality Report 16*(10), 3–9.

Anderson, E. T. and D. I. Simester (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research 51*(3), 249–269.

Anderson, J. C., J. A. Narus, and W. Van Rossum (2006). Customer value propositions in business markets. *Harvard business review 84*(3), 1–8.

Anderson, K. C. (2012b). The impact of social media on lodging performance. *Cornell Hospitality Report 12*(15), 6–11.

Anderson, M. (2014). 88% of consumers trust online reviews as much as personal recommendations. [Online; accessed 25-December-2017].

Anderson, M. and J. Magruder (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal 122*, 957–989.

Banjo, S. (2012). Firms take online reviews to heart. [Online; accessed 06-October-2015].

Bass, F. M., A. Krishnamoorthy, A. Prasad, and S. P. Sethi (2005). Building brand awareness in dynamic oligopoly markets. *Marketing Science 24*(4), 556–568.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy 81*(3), 637–654.

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

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brand, T. (2017). What's going on in the social trading and investing space. [Online; accessed 3-Jun-2018].

Breitmayer, B., M. Mensmann, and M. Pelster (2017). Social recognition and investor overconfidence. *Working Paper, Leuphana University, Germany*.

BrightLocal (2015). Local consumer review survey. [Online; accessed 25-April-2016].

BrokerNotes (2018). The modern trader: A report on the 13.9 million online traders. [Online; accessed 9-Oct-2018].

Brown, R. G. (2004). *Smoothing, forecasting and prediction of discrete time series*. Prentice Hall.

Cairns, A. J. G. (2004). *Interest Rate Models: An Introduction*. Princeton University Press.

Cennamo, C. and J. Santaló (2013). Platform competition: Strategic trade-offs in platform markets. *Strategic management journal 34*(11), 1331–1350.

Cennamo, C. and J. Santaló (2015). How to avoid platform traps. *MIT Sloan Management Review 57*(1), 12.

Cennamo, C. and J. Santaló (2018). Generativity tension and value creation in platform-based technology ecosystems. *Forthcoming, Organization Science*.

Chalasani, P. and S. Jha (1997). *Steven Shreve: Stochastic Calculus and Finance*.

Chevalier, J. A., Y. Dover, and D. Mayzlin (2018). Channels of impact: user reviews when quality is dynamic and managers respond. *Working Paper*.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43*(3), 345–354.

Chintagunta, P. K. and N. J. Vilcassim (1992). An empirical investigation of advertising strategies in a dynamic duopoly. *Management Science 38*(9), 1230–1244.

Choi, S. H. and M. Chhabria (2012). Effective delays in portfolio disclosure. *Journal of Financial Regulation and Compliance 20*(2), 196–211.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

De Gooijer, J. G. and R. J. Hyndman (2006). 25 years of time series forecasting. *International Journal of Forecasting 22*(3), 443–473.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science 52*(10), 1577–1593.

DeMers, J. (2014). How one hotel ruined its reputation by penalizing negative reviews. [Online; accessed 06-October-2015].

Dixit, A. K., R. S. Pindyck, and R. S. Pindyck (1994). *Investment Under Uncertainty.* Princeton University Press.

Doering, P., S. Neumann, and S. Paul (2015). A primer on social trading networks–institutional aspects and empirical evidence. *Working Paper, Ruhr-University Bochum, Germany*.

Duan, W., B. Gu, and A. B. Whinston (2008). The dynamics of online word-of-mouth and product salesan empirical investigation of the movie industry. *Journal of retailing 84*(2), 233–242.

Eldridge, R. (2017). How social media is shaping financial services. [Online; accessed 9-Oct-2018].

Empire, F. (2017). The forex industry in 2018: Regulation, market size, and the future. [Online; accessed 3-Jun-2018].

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Erickson, G. M. (1985). A model of advertising competition. *Journal of Marketing Research*, 297–304.

Erickson, G. M. (1995). Advertising strategies in a dynamic oligopoly. *Journal of Marketing Research*, 233–237.

Erickson, G. M. (2009). Advertising competition in a dynamic oligopoly with multiple brands. *Operations research 57*(5), 1106–1113.

Erskine, R. (2018). Study: 97% of business owners say online reputation management is important – here's how to keep up. [Online; accessed 30-August-2018].

eToro (2017). [Online; accessed 9-Oct-2018].

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics 54*(1), 173–182.

Forbes (2013). Is online reputation management worth the money? [Online; accessed 2-April-2018].

Forman, C., A. Ghose, and B. Wiesenfeld (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research 19*(3), 291–313.

Fornell, C. and B. Wernerfelt (1988). A model for customer complaint management. *Marketing Science 7*(3), 287–298.

Fruchter, G. E. (1999). The many-player advertising game. *Management Science 45*(11), 1609–1611.

Fruchter, G. E. and S. Kalish (1997). Closed-loop advertising strategies in a duopoly. *Management Science 43*(1), 54–63.

Gartner (2012). Gartner says by 2014, 10-15 percent of social media reviews to be fake, paid for by companies. [Online; accessed 06-October-2015].

Gesenhues, A. (2013). Survey: 90% of customers say buying decisions are influenced by online reviews. [Online; accessed 06-October-2015].

Ghose, A. and Y. Yao (2011). Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research 22*(2), 269–288.

Glaser, F. and M. Risius (2018). Effects of transparency: analyzing social biases on trader performance in social trading. *Journal of Information Technology 33*(1), 19–30.

Glassman, N. (2011). What every social media marketer should know about revinate. [Online; accessed 22-Sep-2016].

Green, W. H. (2007). *Econometric Analysis*. Prentice Hall, New Jersey, United States.

Gu, B. and Q. Ye (2014). First step in social media: measuring the influence of online management responses on customer satisfaction. *Production and Operations Management 23*(4), 570–582.

Guéant, O., J.-M. Lasry, and P.-L. Lions (2011). Mean field games and applications. In *Paris-Princeton lectures on mathematical finance 2010*, pp. 205–266. Springer.

GuestRevu (2016). Back to basics: what is online reputation management and why does it matter to hotels. [Online; accessed 17-January-2017].

Gunarathne, P., H. Rui, and A. Seidmann (2017). When social media delivers customer service: Differential customer treatment in the airline industry. *Management Information Systems Quarterly*, To appear.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Hass, D. and D. K. Rigby (2004). Outsmarting walmart. *Harvard Business Review*.

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies 29*(11), 3177–3209.

Herzenstein, M., S. Sonenshein, and U. M. Dholakia (2011). Tell me a good story and i may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research 48*(Special Interdisciplinary Issue 2011: Consumer Financial Decision Making), S138–S149.

Ho, Y., J. Wu, and Y. Tan (2017). Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research 28*(3), 626–642.

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (2014). From lemon markets to managed markets: the evolution of ebays reputation system. *Working paper, Ohio State University*.

Huotari, P. (2017). Too big to fail? overcrowding a multi-sided platform and sustained competitive advantage. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers 116*(1), 770–799.

Jabr, W. and Z. Zheng (2014). Know yourself and know your enemy: An analysis of firm recommendations and consumer reviews in a competitive environment. *Management Information Systems Quarterly 38*(3), 635–654.

Jiang, H. and M. Verardo (2018). Does herding behavior reveal skill? an analysis of mutual fund performance. *The Journal of Finance 73*(5), 2229–2269.

Kapadia, R. (2018). The fund industry is overdue for change. [Online; accessed 9-Oct-2018].

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics 156*(1), 3–20.

Kent, K. (2014). Catching up with the competition: How to get more online reviews from your customers. [Online; accessed 25-December-2017].

Kim, P. (2017). Average mutual fund return. [Online; accessed 9-Oct-2018].

Kortekaas, V. (2013). Social trading targets savvy retail investors. [Online; accessed 1-Jun-2018].

Kumar, N., L. Qiu, and S. Kumar (2017). Exit, voice, and response in digital platforms: An empirical investigation of online management response strategies. *Information Systems Research*, To appear.

Kumar, N., L. Qiu, and S. Kumar (2018). Exit, voice, and response in digital platforms: An empirical investigation of online management response strategies. *Information Systems Research, forthcoming*.

Kwark, Y., J. Chen, and S. Raghunathan (2014). Online product reviews: Implications for retailers and competing manufacturers. *Information systems research 25*(1), 93–110.

Kwark, Y., G. M. Lee, P. A. Pavlou, and L. Qiu (2018). On the spillover effects of online product reviews on purchases: Evidence from clickstream data. *Working paper*.

Lasry, J.-M. and P.-L. Lions (2007). Mean field games. *Japanese Journal of Mathematics 2*(1), 229–260.

Lee, I. and Y. J. Shin (2018). Fintech: Ecosystem, business models, investment decisions, and challenges. *Business Horizons 61*(1), 35–46.

Lee, W. and Q. Ma (2015). Whom to follow on social trading services? a system to support discovering expert traders. In *Tenth International Conference on Digital Information Management (ICDIM)*, pp. 188–193. IEEE.

Lee, Y. J., K. Xie, and A. Besharat (2016). Management response to online wom: Helpful or detrimental? *AMCIS*.

Li, L., S. Tadelis, and X. Zhou (2016). Buying reputation as a signal of quality: Evidence from an online marketplace. *NBER Working Paper*.

Li, X. and L. M. Hitt (2008). Self-selection and information role of online product reviews. *Information Systems Research 19*(4), 456–474.

Li, X., L. M. Hitt, and Z. J. Zhang (2011). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems 27*(4), 9–42.

Little, J. D. C. (1979). Aggregate advertising models: The state of the art. *Operations Research 27*(4), 629–667.

Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Liu, D., S. Kumar, and V. S. Mookerjee (2012). Advertising strategies in electronic retailing: A differential games approach. *Information Systems Research 23*(3), 903–917.

Lizorkin, P. I. (2001). Bessel functions. *Encyclopedia of Mathematics*.

Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp. com. *Harvard Business School NOM Unit Working Paper* (12-016).

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science 62*(12), 3412–3427.

Ma, L., B. Sun, and S. Kekre (2015). The squeaky wheel gets the grease - an empirical analysis of customer voice and firm intervention on twitter. *Marketing Science 34*(5), 627–645.

Malmendier, U. and D. Shanthikumar (2014). Do security analysts speak in two tongues? *The Review of Financial Studies 27*(5), 1287–1322.

Manley, B. (2017). Hoteliers strive for control in age of online ratings. [Online; accessed 2-September-2018].

Manning, C. D., P. Raghavan, and H. Schutze (2008). Scoring, term weighting, and the vector space model. In *Introduction to Information Retrieval*. Cambridge University Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Mäschle, O. (2012). Which information should entrepreneurs on german crowdinvesting-platforms disclose? Technical report, Thünen-Series of Applied Economic Theory.

Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science 25*(2), 155–163.

Mayzlin, D., Y. Dover, and J. Chevalier (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review 104*(8), 2421–2455.

Medallia (2015). Responding to social media boosts a companys bottom line, new research finds. [Online; accessed 13-November-2017].

Merchant, K. A. (1982). The control function of management. *Sloan Management Review 23*(4), 43–55.

Meyer, R. (2013). Brands need to use social media the right way. [Online; accessed 06-October-2015].

Moe, W. W. and D. A. Schweidel (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science 31*(3), 372–386.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, 315–335.

Naik, P. A., A. Prasad, and S. P. Sethi (2008). Building brand awareness in dynamic oligopoly markets. *Management Science 54*(1), 129–138.

Oehler, A., M. Horn, and S. Wendt (2016). Benefits from social trading? empirical evidence for certificates on wikifolios. *International Review of Financial Analysis 46*, 202–210.

Page, V. (2018). How expedia makes money. [Online; accessed 2-September-2018].

Pan, W., Y. Altshuler, and A. Pentland (2012). Decoding social influence and the wisdom of the crowd in financial trading network. *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, 203–209.

Pavlou, P. A. and A. Dimoka (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research 17*(4), 392–414.

Pelster, M. and A. Hofmann (2017). About the fear of reputational loss: Social trading and the disposition effect. *Working Paper, St. Johns University, United States*.

Pool, V. K., N. Stoffman, and S. E. Yonker (2015). The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance 70*(6), 2679–2732.

Poterba, J. M. and L. H. Summers (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics 22*(1), 27–59.

Prasad, A. and S. P. Sethi (2004). Competitive advertising under uncertainty: A stochastic differential game approach. *Journal of Optimization Theory and Applications 123*(1), 163–185.

Proserpio, D. and G. Zervas (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science 36*(5), 645–665.

Reuters (2018). Israeli social trading firm etoro raises $100 million in private funding. [Online; accessed 1-Aug-2018].

Rietveld, J., M. A. Schilling, and C. Bellavitis (2019). Platform strategy: Managing ecosystem value through selective promotion of complements. *Forthcoming, Organization Science*.

Röder, F. and A. Walter (2017). What drives investment flows into social trading portfolios? *Working Paper, University of Giessen, Germany*.

Ross, S. M. (2014). *Introduction to probability models*. Academic Press.

Rossi, P. E. (2014). Even the rich can make themselves poor: A critical examination of iv methods in marketing applications. *Marketing Science 33*(5), 655–672.

Rubel, O., P. A. Naik, and S. Srinivasan (2011). Optimal advertising when envisioning a product-harm crisis. *Marketing Science 30*(1), 1048–1065.

Rushe, D. (2013). Fake online reviews crackdown in new york sees 19 companies fined. [Online; accessed 07-April-2016].

Sethi, S. P. (2004). Competitive advertising under uncertainty: A stochastic differential game approach. *Journal of Optimization Theory and Applications 123*(1), 163–185.

Sorger, G. (1989). Competitive dynamic advertising: A modification of the case game. *Journal of Economic Dynamics and Control 13*(1), 55–80.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics 8*, 321–340.

Theil, H. (1966). *Applied Economic Forecasting*. North Holland Publishing Co.

Tiwana, A., B. Konsynski, and A. A. Bush (2010). Research commentaryplatform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information systems research 21*(4), 675–687.

TripAdvisor (2012). Survey finds half of tripadvisor users will not book a hotel that has no reviews. [Online; accessed 13-November-2017].

Venkatesan, R., K. Mehta, and R. Bapna (2007). Do market characteristics impact the relationship between retailer characteristics and online prices? *Journal of Retailing 83*(3), 309–324.

Verbeek, M. and Y. Wang (2013). Better than the original? the relative success of copycat funds. *Journal of Banking & Finance 37*(9), 3454–3471.

Wang, Q. and Z. Wu (2001). A duopolistic model of dynamic competitive advertising. *European Journal of Operational Research 128*(1), 213–226.

Wang, Y. and A. Chaudhry (2018). When and how managers responses to online reviews affect subsequent reviews. *Journal of Marketing Research 55*(2), 163–177.

Wareham, J., P. B. Fox, and J. Cano Giner (2014). Technology ecosystem governance. *Organization Science 25*(4), 1195–1215.

Williams, D. (2002). Structure and competition in the us home video game industry. *International Journal on Media Management 4*(1), 41–54.

Wu, R. and C. Qiu (2016). Seller manipulation of consumer reviews under competition. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 3575–3583. IEEE.

Yang, M., E. Zheng, and V. Mookerjee (2018). Prescribing response strategies to manage customer opinions: A stochastic differential equation approach. *Information Systems Research, forthcoming*.

Ye, S., G. Gao, and S. Viswanathan (2014). Strategic behavior in online reputation systems: Evidence from revoking on ebay. *Management Information Systems Quarterly 38*(4), 1033–1056.

Zhou, Z., S. Xiao, Y.-C. C. Ho, and Y. Tan (2018). The persuasive and informative effects of information disclosure: Evidence from an online supply chain finance market. *Working Paper, University of Washington, United States*.

Zhu, F. and X. Zhang (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing 74*(2), 133–148.

BIOGRAPHICAL SKETCH

Mingwen Yang was born in Hebei Province, China, to Zhenhong Yang and Aifen Wang. She received a Bachelor of Science degree in Information Management and Information Systems in July 2010 from Tianjin University, China. She then attended Shanghai Jiao Tong University (SJTU), China and Georgia Institute of Technology (Gatech) and received her two MS degrees: one in Management Science (SJTU) and the other in Electrical and Computer Engineering (Gatech) in 2013. In Fall 2013, she came to The University of Texas at Dallas (UTD) to pursue a doctoral degree in Management Science at the Naveen Jindal School of Management. At UTD, she taught undergraduate courses in Information Systems in Fall 2017 and Summer 2018.

CURRICULUM VITAE

Mingwen Yang

April, 2019

Contact Information:

Jindal School of Management, The University of Texas at Dallas
800 West Campbell RD, SM 33, Richardson, TX, 75080
Email: `mingwen.yang@utdallas.edu`

Education:

Ph.D., Management Science, The University of Texas at Dallas, 2019
M.S., Electrical and Computer Engineering, Georgia Institute of Technology, 2013
M.S., Management Science and Engineering, Shanghai Jiao Tong University, China, 2013
B.S., Information Management and Information Systems, Tianjin University, China, 2010

Publications:

Mingwen Yang, Eric Zheng, Vijay Mookerjee. 2019. "*Prescribing Response Strategies to Manage Customer Opinions: A Stochastic Differential Equation Approach.*" Forthcoming at *Information Systems Research.*
Mingwen Yang, Guohua Wan, Eric Zheng. 2014. "*A predictive DEA model for outlier detection.*" *Journal of Management Analytics*, Vol. 1, No. 1: 20-41.

Working Papers:

Mingwen Yang, Varghese Jacob, Srinivasan Raghunathan. "*Cloud Deployment Model's Role in Provider and Users Security Investment Incentives.*"
Mingwen Yang, Eric Zheng, Vijay Mookerjee. "*The Race for Online Reputation: Implications for Firms, Platforms and Consumers.*"
Mingwen Yang, Eric Zheng, Vijay Mookerjee. "*The Transparency - Revenue Tension in Social Trading: Implications for Platforms and Investors.*"

Teaching Experience:

Instructor at The University of Texas at Dallas
Object-Oriented Programming (Undergraduate Core): Summer 2018 and Fall 2017.

Selected Honors:

Best Paper Runner-up in INFORMS Annual Meeting E-Business Cluster	2018
Betty and Gifford Johnson Graduate Scholarship, The University of Texas at Dallas	2017
Graduate Scholarship, The University of Texas at Dallas	2013 - 2018
Wallace H. Coulter Fellowship, Georgia Institute of Technology	2010 - 2011
National Graduate Scholarship, Shanghai Jiao Tong University, China	2010 - 2012
Unified National Graduate Entrance Examination Waiver, China	2010
Excellent Student Award, Hebei Province, China	2006