Essays on Individuals' Decision-Making and Engagement in Digital Platforms

Jiaying Deng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:
Yong Tan, Chair
Ming Fan
Stephanie Lee

Program Authorized to Offer Degree:
Business Administration

University of Washington

Abstract

Essays on Individuals' Decision-Making and Engagement in Digital Platforms

Jiaying Deng

Chair of the Supervisory Committee:
Michael G. Foster Endowed Professor Yong Tan
Information Systems and Operations Management

Technology has brought about a lot of business innovations, such as fintech, sharing economy and entertainment analytics, which provides great opportunities of examining these innovations from different perspectives. My dissertation is motivated by the desire to gain a better understanding of the transformational effect of the digital innovations, with a specific focus on the individuals' decision-making behavior. First, I focus on a mobile gaming platform. To ensure the success of mobile games, game publishers need to keep their players engaged and, at the same time, seek monetization strategies that do not interrupt players' engagement. Reward ads, a relatively new and increasingly popular in-app advertising monetization model in which players can voluntarily watch an ad in exchange for a reward within the game, can help publishers generate revenue without breaking the flow of the game. In this essay, I investigate how competence-related intrinsic motivation factors including players' feeling of competence (measured by (a) perceived difficulty and (b) perceived achievement), (c) the fluctuation of perceived difficulty, and (d) reward ads (which can enhance players' intrinsic motivation and interact with players' feeling of competence) affect players' dynamic engagement state evolution. I find that players' engagement transition is a complex process - as players interact with the mobile game over time and progress through the game, players may be motivated to different intrinsic motivation factors. Therefore, intrinsic motivation factors should be treated as a dynamic process because one factor may be more or less effective as

players move to different engagement levels. Second, I study a social trading platform which is a recent innovation in the field of fintech. Social trading is a novel form of investing that allows retail investors to observe the trading behavior of other investors and to communicate through the news feed. Moreover, social trading allow inexperience investors (followers) to automatically follow the trades of experts (leaders) in real time. In this essay, I examine the determinants of the link formation and link dissolution processes, with an emphasis on the social communications among investors. The results show that the determinants of link formation and dissolution are asymmetric, and different types of social communications, such as posts and comments, have different implications for link formation and dissolution. In addition, the results show that financial performance and demographic characteristics are also important determinants of link formation. However, once a link is formed, followers mainly focus on financial performance in addition to social communication but not on demographic characteristics. The findings have important implications for both investors and social trading platforms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest gratitude to my advisor, Professor Yong Tan, for his continuous support and guidance through my Ph.D. study. He is the best role model and mentor. I feel so fortunate to have a supervisor who cared so much about my growth and encouraged me in critical times throughout the journey. I hope to emulate his research enthusiasm, methodological rigor, and humility towards work to my academic career forward.

Also, I would like to extend my sincere gratitude to my committee members, Professor Shi Chen, Professor Ming Fan, Professor Stephanie Lee, and Professor Jing Tao. Their insightful comments and constructive feedback have helped me a lot in improving my research study.

My special thanks also go to Shawna Reimers, Beau Kirkeby, Nuzulita Budhiari, Jaime Banaag, Jessica Aceves, and other staff at the Foster School of Business for their generous support during my study. They took good care of all the Ph.D. students. Whenever I need any help with my teaching, coursework, or graduation schedule, they were always there, available and supportive.

My gratitude also extends to my peers and my academic families: Yuchen Liu, Zhijin Zhou, Tongxin Zhou, Zixuan Meng, Jinyang Zheng, Xue Tan, Yi-Chun Ho and Lin Hao. I am so grateful for their support, encouragement, and their willingness to share my ups and downs during the Ph.D. years.

Last but not the least, I would like to thank my families and friends who were there at the beginning of the journey. I could not have achieved this far without their support and encouragement.

Chapter 1

INTRODUCTION

With the rapid development of technology, a lot of business innovations have been brought about, which have created enormous social and economic values for individuals, firms and society. Those innovations not only change individuals' behavior, but also enable researchers to observe and analyze their behavior. For example, traditionally, product or service suppliers were usually firms, but sharing economy enabled individuals to become the suppliers. In the past, customers were burdened with a relatively high cost in financial services, but fintech improved the efficiency, reduced the cost and made the financial services more accessible to the public. In addition, technology has also enabled companies to observe individuals' online activities and detect their preference, which in turn provides important implications to firms' business strategies and management. Understanding individual users' decision-making and engagement is of focal importance to the platforms because the business model of most platforms relies on individuals' active engagement and participation. This dissertation provides a comprehensive investigation of individuals' decision-making and engagement in two contexts.

In Chapter 3, I study players' engagement in a mobile gaming app. I try to answer a critical question for game publishers, that is what factors can engage players in the game. Engagement refers to a mental state that reflects players' desire to play the game. Players generate a sense of achievement by successfully proceeding to the next level. In general, the enjoyment generated from solving puzzles or overcoming challenges makes players more engaged in the game, but it is not so clear regarding how to decide the difficulty level in the game sequence to engage players. If the game is too easy, players may feel bored and if it is too difficult, gamers may feel overwhelmed. Therefore, one of the research questions is to examine how the average and the fluctuation of perceived difficulty affect players' engagement state. Secondly, players' engagement is also an important factor to consider in deciding the mobile games' monetization strategy. While most publishers rely on advertisement revenue, traditional advertisement can annoyingly interrupt the flow of the game, discourage players from playing the game, and lower players' engagement. Reward ads can provide a solution to this dilemma. Instead of forcing players to watch ads, reward ads are opt-in, which means that players voluntarily choose whether or not to watch reward ads in exchange for incentives or rewards within the game, such as virtual coins or other premium features. Reward ads enable players who cannot progress the game organically to continue playing the game, increase their session time, and potentially improve their engagement levels. Despite the popularity and innovativeness of reward ads, the empirical literature on reward ads and their relationship with players' engagement has been limited. This essay fills the gap. I adopt a hidden Markov model to describe players' latent engagement transitions and examine the effect of the aforementioned determinants. Through the empirical estimation, I find that as players interact with the mobile game over time and progress through the game, players are motivated by different intrinsic motivation factors. The findings enhance the understanding of players' engagement and motivations to play mobile games and provide design guidance for mobile game publishers.

In Chapter 4, I study a social trading platform and examine the determinants of the social network evolution on the platform. Compared to other online trading platforms, the uniqueness of social trading is threefold. First, information is more transparent on social trading platforms because every investor can observe the transaction history of other investors. Second, it has a social component where investors can communicate through the news feed. Most importantly, social trading allows copy trading. That is to say, as an investor on the platform, she can directly follow other investors' positions in real time by forming a link to their account. Social trading platforms also require a new perspective on considerations of the evolution of social networks. In the social networks on social trading platforms, the information flow among users is directly tied to cash flows because of the copy trading feature.

As a result, due to the monetary incentive, compared to other traditional social networks such as Facebook or Twitter, link dissolutions in social trading networks happen more frequently. Since the social network is very important and it is very different from traditional social networks, I simultaneously study the link formation and link dissolution process through a dynamic network analysis. I place a particular emphasis on social network features and study the impact of social communication on link formation and dissolution, while at the same time, controlling financial performance, demographic information and the network structure features. I find that the effects of different types of social communication are asymmetric for link formation and dissolution. The findings provide important implications for both investors and social trading platforms.

Chapter 2
LITERATURE REVIEW

I investigate the individuals' decision-making process and engagement in two different contexts. In this section, I review the related literature, and illustrate how my work extends the understanding of the previous studies. Section 2.1 first introduces the conceptual research framework and provides theoretical support on factors affecting players' engagement transition. Then, I review the recent empirical studies on players' engagement in online games and online advertisement, and discuss how my study differs from and extends previous work. Section 2.2 shows the related studies on social trading platforms. I first summarize the fast-growing literature on social trading, which generally can be divided into three groups. Then, I review the related literature that discusses the determinants of tie formation and dissolution in other related settings.

2.1 Players' Engagement in Online Mobile Games

2.1.1 Motivation theory and players' engagement

Following the huge commercial success of online games, the prior literature examines various economic aspects of online games, including pricing strategy (Liu 2010), network effect (Wu et al. 2013), demand estimation (Ishihara and Ching 2019), and business models (Guo et al. 2019a). An important fundamental research question for online games is why people enjoy playing the games. Successful games can create a state of *flow* where players are fully engaged in the activity and lose their self-consciousness and sense of time (Csikszentmihalyi and Csikzentmihaly 1990). It is important to identify factors that affect players' engagement as the more engaged the players are, the more revenue the game publisher can generate. Therefore, I build on the motivation theory literature to identify and examine the determinants of players' engagement with online games (White 1959, Ryan and Deci 2000a, Ryan et al. 2006).

Self-determination theory (SDT) is a widely accepted motivation theory that has been applied in multiple settings including education, organization, and online games (Deci et al. 1985). According to SDT, individuals are driven by two different types of motivation when conducting an activity: intrinsic motivation and extrinsic motivation (Ryan and Deci 2000a). Intrinsic motivation refers to doing something because it is inherently enjoyable whereas extrinsic motivation refers to doing something for a separable outcome such as obtaining an externally imposed reward or avoiding something negative (Hennessey et al. 2015).[1] By building upon SDT, Hsu and Lu (2004) further propose another type of motivation: social motivation, which captures users' needs to interact with others in online communities. In this study, I focus on and examine how the intrinsic motivation affects players' engagement for the following reasons. First, compared with other activities such as passing a test or finishing a work project, playing games is more intrinsically motivated due to the voluntary nature of playing the game, and most players engage in the gaming activity "for its own sake" and do not earn external game rewards or approval (Ryan et al. 2006, Malone 1981). In fact, this is especially true for mobile puzzle games, which is a popular genre of mobile games that focuses on puzzle solving.[2] It is documented that puzzle game players focus more on the intrinsic motivation factors because one of the most predominant reasons that players play puzzles is to feel accomplished for completing a challenging task.[3] In other words, puzzle players enjoy the feeling of competence, which provides the basis of maintaining intrinsic motivation (Ryan and Deci 2000a). Furthermore, the examination of players' intrinsic motivation can offer

[1] Ryan and Deci (2000b) provide a detailed discussion about different types of intrinsic and extrinsic motivation.

[2] According to Statista, the total revenue of puzzle games segment reached 15.844 billion US dollars in 2020.

[3] https://www.facebook.com/fbgaminghome/marketers/build-great-games/game-design-hub/genre-great-games-report

valuable insights and guidance to game publishers on how to design games that consumers enjoy.

Therefore, I build upon the prior literature on intrinsic motivation theory to examine how competence-related intrinsic motivation factors including players' feeling of competence (measured by (a) perceived difficulty and (b) perceived achievement), (c) the fluctuation of perceived difficulty, and (d) reward ads (which can enhance intrinsic motivation and interact with players' feeling of competence) affect players' engagement. The following subsection provides more details on the intrinsic motivation factors I examine, and Figure 2.1 summarizes the conceptual research framework.

Feeling of competence. According to White (1959), which formulates the intrinsic motivation theory, individuals are attracted to participate in activities that they can explore and affect the environment. When such attempts create positive feedback, individuals receive intrinsic rewards, such as feelings of competence and pleasure, and are motivated to continue the effectance efforts. Deci et al. (1985) find that the sense of competence is an important factor that supports individuals' intrinsic motivation. In the context of video games, Ryan et al. (2006) conduct laboratory experiments to explain video game players' motivation, and find that perceived in-game competence is among the most important satisfactions provided by games. Competence includes two concepts: a feeling of achievement and a need for a challenge (Deci et al. 1985). A feeling of achievement (or the positive feedback) can enhance the intrinsic pleasure whereas appropriate challenge (difficulty) levels can make players feel neither bored nor overwhelmed (Csikszentmihalyi and Csikzentmihaly 1990, Malone 1981). Therefore, I extend the SDT to examine the effects of players' *perceived achievement* and *perceived difficulty* on their engagement using real world data.

Fluctuation of perceived difficulty. Importantly, I also examine how the *fluctuation* of perceived difficulty, in addition to players' perceived competence (i.e., perceived achievement and perceived difficulty), is associated with players' game engagement as players play multiple rounds of games. In the prior literature, the competence related factors are generally examined in static laboratory settings and the dynamic fluctuations in gamers' perceptions has been mostly ignored. Nevertheless, I posit that dynamic fluctuations in gamers' perceptions should not be dismissed as they also affect players' intrinsic motivation via following mechanisms. First, fluctuations in difficulty sequence can meet people's variety seeking demand (Sevilla et al. 2019). People get bored of doing the same thing such as eating the same foods or listening to the same music. In the context of games, playing the same level of difficulty several times can bring about boredom (Chanel et al. 2008). Therefore, variations (i.e., fluctuation) can have a positive effect on players' engagement and help player feel more excited and stimulated over time (Berger et al. 2021, Rolls et al. 1981). In other words, variations in difficulty can provide stimulation and counteract the satiation effects from repetition (Sevilla et al. 2019). Second, the fluctuation can keep players uncertain about the future difficulty level and stimulate a feeling of suspense. Suspense, which is an effective tool that attracts individuals' attention and encourages individuals' engagement, is widely applied in the entertainment industry (Ely et al. 2015). For example, Lomas et al. (2017) find that the suspense of a close sports game (i.e., the final scores of the winning and losing team are very close) significantly improves the audience's continuation of intrinsic motivation. Third, according to the literature on the achievement motivation (Sagie et al. 1996, Atkinson 1974, Grote and James 1991), among different aspects of achievements, difficulty and quantity of achievements are important aspects of every achievement system which can be employed to enhance performance (Groening and Binnewies 2019). On one hand, when considering the difficulty of achievements, the amount of achievement generated from solving one difficult puzzle is greater than that from solving one easy puzzle because a difficult puzzle is more intellectually challenging (Groening and Binnewies 2019). At the same time, when considering the quantity of achievements, more progression through puzzle levels can satisfy the gamers' demand to make progress and accumulate multiple achievements. The fluctuation of perceived difficulty can help account for players' need for both difficulty of achievements and quantity of achievements.

Reward ads. When designing a mobile game, game publishers can not only directly affect players' intrinsic motivations by affecting players' feeling of competence and the fluctuation of perceived difficulty, but also affect players' intrinsic motivations via reward ads. Reward ads, a relatively new and increasingly popular in-app advertising monetization model in which players can voluntarily watch an ad in exchange for a reward within the game, can help publishers generate revenue without breaking the flow of the game. Therefore, I additionally extend intrinsic motivation theory to examine how the reward ad can be employed to help keep players engaged.

Reward ads can affect players' engagement and intrinsic motivations in the following ways. First, players feel accomplished when completing a challenging task, but when the puzzle becomes too difficult to solve, players may experience a sense of anxiety. In such a case, they may get stuck in a puzzle and may not be able to move forward for a long time, which can result in a game-quit decision and eventually decrease players' engagement. Reward ads can be helpful in such situations because the rewards that players receive after watching the reward ads allow gamers to buy premium hints that can lower the level of puzzle difficulty, help them proceed to the next round, and provide them an opportunity to explore more about the game (Sheng et al. 2020). Second, according to SDT, autonomy, which refers to flexibility when conducting an activity, is another factor that enhances intrinsic motivation (Ryan et al. 2006). Compared to other traditional ad formats, reward ads offer players a sense of control because players can independently decide whether and when to watch the reward ad. Reward ads are non-disruptive and non-intrusive ads that do not interfere with players' autonomy. Third, reward ads can also trigger a reciprocity effect (Rubin 1975) due to the value-exchange experience between players and publishers,[4] i.e., players get compensation for watching ads and become less annoyed, which can result in a favorable impression on players (Selle 2020).

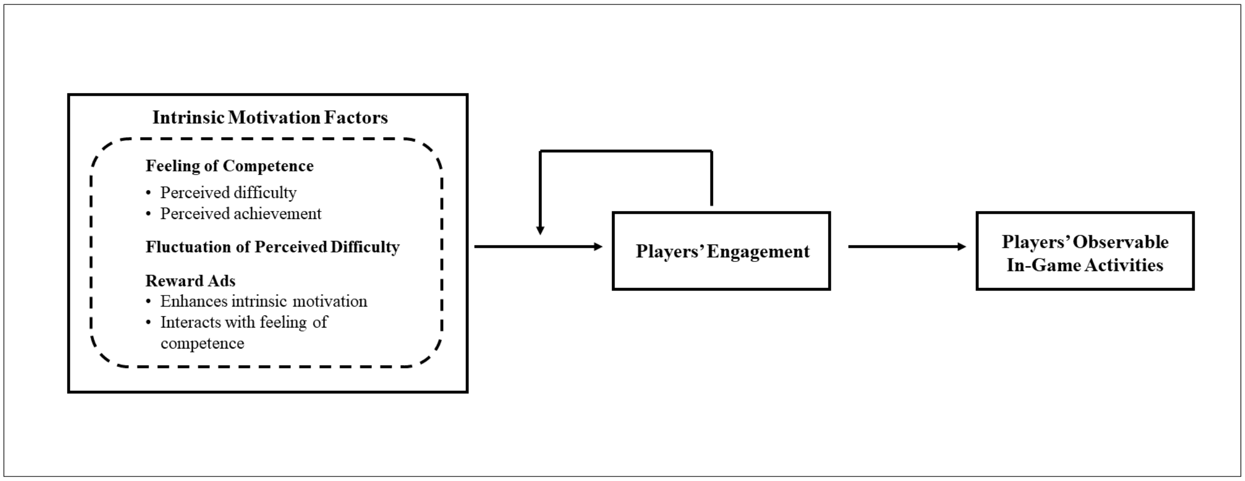
There has not yet been empirical literature that examines the extent to which reward ads lead to a positive attitude and higher engagement levels. Moreover, the relationship between reward ads watching behavior and players' latent engagement is unclear. It could

[4]Reciprocity refers to rewarding friendly actions; people are usually much nicer and more cooperative than predicted by the self-interest model.

be possible that players in a low engagement level are more likely to watch reward ads to get premium content because they generally have a low intention to make in-game purchases. It is also possible that highly engaged players are more likely to watch reward ads due to their immersion in the game. Therefore, I explicitly examine how historical reward ads affect engagement transitions, and how players' latent engagement governs their current reward ads watching behavior. Moreover, reward ads may interact with players' feeling of competence (i.e., perceived difficulty and perceived achievement) when affecting players' engagement. For example, reward ads may be more effective when the perceived difficulty of a puzzle is higher because reward ads can enable players to stay longer by providing premium hints when players cannot make progress organically. Relatedly, reward ads may be more effective and help keep players engaged when perceived achievements are higher if reward ads allow for an opportunity to explore more about the game (i.e., successfully solve more puzzles) by providing shortcuts to obtain premium hints. Therefore, I additionally examine the interaction effect between reward ads and players' feeling of competence when engaging players.

Dynamics of intrinsic motivation and engagement. In examining players' engagement with mobile games, it is important to note that players' motivations and engagement exhibit a dynamic relationship. As players interact with the mobile game over time and progress through the game, players may be motivated by different intrinsic motivation factors. Therefore, intrinsic motivation factors should be treated as a *dynamic* process because one factor may be more or less effective as players move to different engagement levels. In other words, even though the intrinsic motivation factors affect players' engagement, their effectiveness may vary depending on players' engagement levels (Yan and Tan 2014, Singh et al. 2011). Therefore, I place emphasis on players' dynamic playing behavior, and I adopt the HMM model to capture and examine the dynamics in players' engagement transitions and in-game activities. Figure 2.1 presents the conceptual research framework, in which I investigate how the intrinsic motivation factors (i.e., feeling of competence, fluctuation of perceived difficulty, and reward ads) impact players' latent engagement across different states. Then, engage-

Figure 2.1: Conceptual Research Framework of Players' Engagement



ment levels ultimately affect players' observed in-game activities. At the same time, the effectiveness of these intrinsic motivation factors is moderated by the player's engagement state.

2.1.2 Players' engagement in online games

There has been growing empirical literature that investigates why gamers enjoy playing games and how to create better game designs (Chang et al. 2008, Liu et al. 2013, Hamari et al. 2015, Huang et al. 2019, Gu et al. 2022). For example, Liu et al. (2013) conduct an experimental study to examine how competition in games affects players' engagement and emotions, and Gu et al. (2022) conducts a randomized field experiments to examine how crowdsourcing affects players' engagement. The prior literature mainly employs surveys and laboratory experiments. However, the simulated environments and study designs may limit the generalizability of findings (Epstein and Harackiewicz 1992). I use granular real-world game playing data, and, to the best of my knowledge, there is only one other paper that uses real-world observational data to examine players' engagement (Huang et al. 2019). Huang et al. (2019) develop a two-stage model where the first stage captures the evolution of players' latent engagement state transitions and, in the second stage, they propose a matching algorithm to maximize players' game-play rounds. My paper differs from Huang et al. (2019) in the following aspects: (1) Their research context is a multi-player video game whereas my research context is a single-player mobile game. Mobile games usually attract more casual gamers and are not as complex as video games. In addition, mobile games generally face more fierce commercial competition due to their low development cost and the large pool of mobile app markets. Although there are a few studies in video games, mobile games have been understudied in the literature. (2) In their paper, players compete with other players and get a ranking in each round. However, my setting does not have a ranking system or social network among players, which helps me to focus on and investigate intrinsic motivation factors while excluding social motivations, such as peer effects. (3) Their state-dependent outcome is the game-play rounds, and their ultimate goal is to develop a matching algorithm to maximize game-play rounds, whereas in this study, the state-dependent outcome includes player's reward ads volume and I investigate how players' competence related intrinsic motivation factors including reward ads watching activity affect players' engagement transitions.

2.1.3 Online advertisement

For many mobile apps, advertisements are an important source of revenue, especially given the low in-app purchase rates. As online advertisements have become increasingly important to firms, a significant amount of literature has been dedicated to online advertisements (Liu et al. 2012, Anderson and Coate 2005, Teixeira et al. 2014, Zhang and Katona 2012), especially on the topics about advertisement effectiveness and content design (Lee et al. 2018b, Bleier and Eisenbeiss 2015, Todri et al. 2020). However, although a lot of efforts have been made to improve advertisements' effectiveness, the tension between advertisements and users is still acute, especially for mobile apps. The frequency, timing, and location of displayed ads are the three most mentioned aspects of ads-related complaints in the mobile app ecosystem (Gui et al. 2017). Reward ads are an innovative ad format that can provide a solution to tensions between publishers and game players by providing a choice for players to watch ads in exchange for rewards in the game. The prior theoretical literature examines the economics of reward ads and investigates when a platform should provide the option of reward ads (Sheng et al. 2020, Guo et al. 2019b). Because reward ads are relatively new compared to traditional ads, there has been very limited empirical literature on reward ads. To the best of my knowledge, Chiong et al. (2017), which, by comparing users' conversion rate (i.e., whether users install the advertised app or not) between reward ads and non-reward ads, find that rewarding a user to watch ads has an overall positive effect on ads conversion rate, is the only other empirical literature on reward ads. However, there has not yet been empirical literature that examines factors that affect reward ads watching behavior, and the relationship between reward ads watching behavior and players' engagement is unclear. This paper bridges the important gap in the literature by examining how reward ads, a relatively new and increasingly popular in-app advertising monetization model, dynamically affect mobile game players' engagement. To the best of my knowledge, this paper is the first one to use individual-level tapstream data to investigate reward ads watching behavior.

2.2 Social Trading Platforms

While the literature has studied social networks in detail (Kane et al. 2014), it is unclear how insights from other social networks can be applied to social trading platforms. The social networks on social trading platforms differ from traditional social networks. Platforms typically do not allow for social relations akin to friendships on Facebook or follower relations on Twitter. Instead, the relations between users in such networks are directly tied to cash flows. The platform provides a service that allows "copy trading" for its customers—duplicating investment strategies from other investors with one's own money without approving each individual transaction. Investors (i.e., followers) who make use of the social features of the platform and form a copy trading link with other investors can delegate their trading; at the same time, leaders (or signal providers) can earn additional income, receiving compensation from the platform for contributing to its business model. As a result, compared to other traditional social networks such as Facebook or Twitter, link dissolution occurs more frequently. While prior literature has mainly focused on link formation, link dissolution has received little attention in the literature—partially because link dissolution is a relatively rare event in many traditional social networks.

2.2.1 Social trading

The study contributes to the fast-growing literature on social trading, which, generally, can be divided into three groups. The first group addresses the institutional aspects of social trading (see, e.g., Doering et al. 2015). A key feature of social trading platforms is a high level of information transparency. Investors can observe the trading behavior of their peers at the trade level and in real time. Considering that some investor trades may contain valuable information, making these trades available in real time potentially undercuts the platforms' payoff potential. To resolve this issue, Yang et al. (2021) propose a personalized trade-level information release policy that allows platforms to optimize their revenues.

A large second group of studies examines how the information transparency that allows investors to observe other investors in real time may affect their trading behavior and the performance implications of social trading. Gemayel and Preda (2017, 2018a) label the state of permanent reciprocal observation and scrutiny that are typical of social trading platforms a "scopic regime". Trading in a scopic regime alters investors' behavioral biases such as the disposition effect (Heimer 2016, Pelster and Hofmann 2018).[5] Focusing on the copy trading feature, the experimental study by Apesteguia et al. (2020) suggests that having the option to directly follow other investors significantly increases the risk-taking behavior of investors. This increased risk taking does not, However, yield superior investment returns (e.g., Pan et al. 2012). Several studies have found that on average, traders on social trading platforms do not outperform the market in the long term (Dorfleitner et al. 2018, Oehler et al. 2016).

[5]The disposition effect is an anomaly discovered in behavioral finance. It is related to the tendency of investors to sell assets that have increased in value while holding assets that have dropped in value.

Only a few investors are able to earn significant short-term excess returns (Dorfleitner et al. 2018, Oehler et al. 2016).

Most closely related to my paper, the last group of papers studies network relationships on social trading platforms. Ammann and Schaub (2021) find that social traders are more likely to have their investment strategies duplicated within three weeks of making (positive) posts on trading platforms. These posts do not, However, seem to contain valuable information, as they do not have any predictive power over future performance. While Ammann and Schaub (2021) focus on fund flows in their analysis, I, in contrast, focus on individual activity-based relationships between traders, i.e., the social ties within the network. Additionally, focusing on fund flows, Röder and Walter (2019) document a positive flow-performance relationship in social trading portfolios, which is limited to the top past performers (i.e., investment flows chase past performance). To optimally exploit the copy trading function, Lee and Ma (2015) develop a system called "W2F (whom to copy)" that enables users of social trading platforms to "discover expert traders" who consistently realize high risk-adjusted performance.

2.2.2 Link formation and dissolution in (social) networks

The evolution of leader-follower networks, or in other words, when and why an investor (follower) follows or unfollows another investor (leader) on a social trading platform, is one of the fundamental questions in social trading. Prior research has mostly focused on the trading behaviors of leaders in this context but has largely ignored followers' decisions about whom to follow or unfollow. In this study, I aim to shed light on this question by focusing on the factors that drive the following and unfollowing decisions of followers on social trading platforms. In this subsection, I review the related literature that discusses the determinants of tie formation and dissolution in other (related) settings and discuss social trading-specific features that may influence tie formation and dissolution.

Trust. It is widely acknowledged that a good relationship between managers and investors is beneficial. Trust is important for managing relationships and regulating their quality (Kaiser and Berger 2021). Trust is relevant when a person (the trustor) has specific expectations of another person (the trustee) and is vulnerable to whether the trustee fulfills those expectations, regardless of control (Mayer et al. 1995). Various factors can influence the existence of trust in a relationship. As noted by Mayer et al. (1995), trustworthiness must be established before any factor can lead to trust. Trustworthiness refers to the trustee and relates to specific characteristics of the trustee, for example, benevolence, integrity, or ability. The most common factor that can establish trustworthiness and ultimate trust is communication (Kaiser and Berger 2021). Communication and timely feedback increase trust; other factors that establish trust are reputation, quality, and partner fit (Kaiser and Berger 2021).

Recent developments in fintech have led to new challenges in managing relationships, particularly because investors cannot establish a personal one-on-one trust relationship but must instead seek to build a relationship with a more or less anonymous mass (Kaiser and Berger 2021). As noted by Wohlgemuth et al. (2016), trust among online community members plays an even more important role in the online trading context, where investors can automatically, without further evaluation, duplicate the investment strategies of their peers.

Social communication. Social communication can help build trust and reduce information asymmetries (Xu and Chau 2018). Duarte et al. (2012) show that borrowers who appear to be more trustworthy based on their pictures have a higher probability of having their loans funded. Xu and Chau (2018) find that both credit grades and lender-borrower communication affect funding outcomes on peer-to-peer lending platforms. In particular, social communication can be a tool that particularly allows listers with poor hard information (e.g., credit grades) to improve their chances of being funded (Xu and Chau 2018).

Social communication can be regarded as soft information. Soft information refers to qualitative information such as media press, communication texts, or market commentary, whereas hard information refers to quantitative information such as stock returns or credit ratings. Even though soft information is usually qualitative, it can be "hardened" using information technologies such as text mining and converted into a quantitative measure. Thus, the main difference between hard and soft information is that the former can be objectively verified and is independent of context, while the quantification of the latter makes use of various degrees of freedom. An immediate consequence of the nature of soft information is that its assessment (e.g., how trustworthy another agent is or other informational cues are) depends on each agent's personal standards. When evaluating information from others, the quality of this information is important for building trust (Xu and Chau 2018). I hypothesize that social communication is important for investors to establish trust on social trading platforms. Thus, I incorporate social communication variables, including a proxy for the quality of social communication, and explore how they affect link formation and dissolution.

Financial performance. As noted above, hard information is also an important determinant of whether investors decide to fund a project, sell a stock, or make a loan (Liberti and Petersen 2019). Social trading platforms are comparable to mutual and hedge funds, as they allow for some form of delegated portfolio management (Doering et al. 2015). Both mutual and hedge funds have received considerable attention in the financial literature, with a particular focus on the determinants of their performance (e.g., Agarwal et al. 2009, Grinblatt et al. 2020) and the relationship between fund performance and (net) fund flows (Sirri and Tufano 1998, Goetzmann et al. 2003). As investors can infer the skills, at least to some degree, of mutual and hedge fund managers from their past performance, (net) fund flows should be explained by past performance (Barber et al. 2016). This stream of the literature also documents not only that mean performance is important but also that the volatility of performance is negatively related to fund flows (Sirri and Tufano 1998, Huang et al. 2007). Importantly, investors seem to determine their inflows and outflows differently (Ivković and Weisbenner 2009). The importance of returns has also been documented in nonprofessional settings, such as online crowdfunding markets (Lin and Viswanathan 2016). Based on this stream of the literature, I examine how the formation and dissolution of leader-follower links are affected by financial performance.

Demographic factors and homophily. A large stream of the social networks literature documents that similarity (homophily) breeds connections (McPherson et al. 2001). Homophily describes the preference of people to favor others who are similar to them rather than those who are dissimilar to them. In financial markets, so-called home bias is a prominent example of a preference for similarity (Coval and Moskowitz 1999, Lin and Viswanathan 2016). In the social trading context, potential cultural differences or language barriers in online communications may also contribute to homophily. In addition, the profile picture (image) and biography on an investor's public profile page can affect how he/she is perceived by others (Duarte et al. 2012).

Dissolution of ties. Even though link dissolution happens less frequently in traditional social networks such as Facebook or Twitter, the dissolution of ties in interfirm networks or in financial markets is a rather frequent event. Studying interfirm networks, Greve et al. (2010) argue that the dissolution of ties may happen in particular when embeddedness is low. Polidoro et al. (2011) further study the importance of embeddedness for tie dissolution and argue that network centrality, i.e., positional embeddedness, does not promote stability, but having common partners, i.e., structural embeddedness, does. Translated to social trading, where common partners do not play a meaningful role and where there are almost no costs to dissolving a tie and engaging in a new relationship (aside from transaction costs that are charged via the spread), these findings suggest that ties could be dissolved rather quickly. Baker et al. (1998) argue that unsatisfactory performance is a main driver of tie dissolution—an argument that can be translated to social trading at face value.

Shafi et al. (2020) study the dissolution of ties in early entrepreneurial finance and argue that a tie discontinuation can have important ripple effects on other ties. In particular, once Ill-established investors cut their ties with a start-up, smaller investors may follow suit. Thus, discontinuation may have important repercussions for start-ups (Shafi et al. 2020). In social trading, despite the high level of transparency, investors face some degree of uncertainty when deciding to follow other investors. As commonly stated on delegated investment opportunities, "past performance is no guarantee of future results". Consequently, investors may not be satisfied with the outcomes of a given tie and decide to discontinue the relationship, or in the words of Shafi et al. (2020): "The decision to withdraw financial support may be primarily related to a venture's underperformance." Based on this stream of literature, I hypothesize that financial performance, or hard information, is a key factor that determines link dissolution. In addition, the role of soft information in link dissolution is not clear. I thus investigate how hard information and soft information affect link dissolution.

Chapter 3

PLAYER GAME-PLAY AND REWARD ADS WATCHING BEHAVIOR IN ONLINE MOBILE GAMES

3.1 Introduction

With the rapid development of mobile technology and consumers' increasing demand for portability, mobile gaming has become a prominent component of not only the video game market, but also of the entire entertainment industry. The value of the global mobile gaming market was $131.2 billion U.S. dollars in 2021 and is projected to reach $173.4 billion U.S. dollars in 2026 (Statista 2022).[1] However, high attrition rates and low engagement is a notorious problem faced by mobile game publishers. Mobile game apps, on average, only retain 25% of users after a single day and around 2.3% of users after a month.[2] Moreover, when monetizing their games, mobile game publishers face significant challenges in ensuring the profitability of mobile games (e.g., advertisement revenue) without losing the players' engagement and negatively affecting players' game experience (Liumila et al. 2019; Statista 2019). Although more than 90% mobile game apps rely on advertisement revenue (Business Insider 2017), traditional advertisement is obstructive. Therefore, to ensure the success of mobile games, it is important for game publishers to keep their players engaged and, at the same time, seek monetization strategies that do not interrupt players' engagement.

Players' engagement refers to an unobservable mental state that reflects players' motivation to play the game. Successful mobile games can create a state of *flow*, which refers to a mental state where players are fully engaged in an activity and lose their self-consciousness and sense of time (Csikszentmihalyi and Csikzentmihaly 1990). Higher en-

[1] `https://www.statista.com/statistics/292512/mobile-contents-market-value-worldwide/`

[2] `https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/`

gagement with games can increase revenue from existing players and additionally attract new players through word-of-mouth or network effects. Players' engagement is also an important factor to consider in deciding the mobile games' monetization strategy. Delivering advertisement content for third parties via in-app advertising creates a dilemma for game publishers; on the one hand, in-game advertisements generate revenue, but, on the other hand, the advertisement can annoyingly interrupt the flow of the game, discourage players from playing the game, and lower players' engagement. Reward ads can provide a solution to this dilemma. Instead of forcing players to watch ads, reward ad are opt-in, which means that players voluntarily choose whether or not to watch reward ads in exchange for incentives or rewards within the game, such as virtual coins, puzzle hints, an extra life in a game, or other premium features. Reward ads are described as a win-win-win for players, game publishers, and advertisers; gamers are happy to be in control of their ad experiences, game publishers can increase revenue while keeping users engaged, and advertisers can enjoy a higher return on ad spend (Google 2020; Nuruljihad 2020). In fact, reward video ads are 53% more likely to be described positively than non-choice-based ads, and mobile games with reward ads are more likely to receive 5-star ratings.[3] Despite the popularity and innovativeness of reward ads, the empirical literature on reward ads and their relationship with players' engagement has been limited.

Playing online games is generally regarded as an intrinsically motivated activity because players volunteer to play the game and the instant feedback that players receive after completing a game provides inherent enjoyment to players (Malone 1981, Deci et al. 1985). Therefore, to design mobile games that induce high engagement levels, mobile game publishers need to understand how players' intrinsic motivation factors affect players' engagement (Liu et al. 2013). I extend the self-determination theory and intrinsic motivation theory to identify and examine intrinsic motivation factors that affect players' engagement. First, I examine how players' feeling of competence, which is measured by (a) perceived difficulty

[3]https://www.facebook.com/audiencenetwork/news-and-insights/2cv-mobile-games-research-paper

(i.e., the difficulty level the players perceive about the game) and (b) perceived achievement (i.e., the positive feedback coming from the success in the game), affect players' engagement states (Ryan et al. 2006). In addition, I examine the effect of (c) the *fluctuation* of perceived difficulty on players' engagement evolution, which has not been empirically studied in the previous empirical literature. Although most prior literature mainly considers intrinsic motivation as a static process, I posit that it is important to examine *dynamic* fluctuation of players' perceptions. If a mobile game is either consistently difficult or easy for players, the players may be more likely to quit the game because they can easily predict the difficulty levels of the next few rounds and forecast whether they can succeed or not. Instead, fluctuation in players' perceived difficulty levels (e.g., "easy-difficult-easy" fluctuations) can keep them in suspense and in being uncertain about the next game round's difficulty level, which can help improve players' engagement levels (Chanel et al. 2008). Lastly, I again extend the intrinsic motivation literature to examine how (d) reward ads can enhance players' intrinsic motivation and interact with players' feeling of competence to affect players' engagement. From the perspective of intrinsic motivation, reward ads enable players who cannot progress the game organically to continue playing the game, increase their session time, and improve their engagement levels. For example, gamers who are stuck in the game can continue playing the game by watching reward ads and receiving hints in return. Furthermore, as reward ads can potentially help lower the puzzle difficulty and interact with players' feeling of achievement, I additionally examine the interaction effects between reward ads and players' feeling of competence. To the best of my knowledge, this is the first paper to examine mobile game players' various intrinsic motivation factors using real-world data. Specifically, I address following research questions: (1) How do intrinsic motivation factors, including *perceived difficulty*, *perceived achievement*, *fluctuation of perceived difficulty*, and *reward ads*, affect players' engagement state transitions? (2) What factors affect players' observable in-game activities and how are they related to players' corresponding engagement levels?

I build a Hidden Markov Model (HMM) to examine how intrinsic motivation factors affect players' engagement levels, and to examine their heterogeneous effects on players' dynamic engagement transitions. The HMM framework provides a flexible way to account for possible changes in players' engagement levels and observed activities (outcomes) by explicitly modeling the structural dynamics. The HMM assumes there are two processes where one process is unobserved (hidden) and stochastically determines the observed outcome process. In this context, players' engagement, which refers to their immersion in the game, is a latent unobserved mental state, and playing behaviors (i.e., game-play decision and reward ads watching behavior) are observed activities, which are closely related to and determined by players' latent engagement states. I set reward ad as one of the observed outcomes because it has direct effects on game publishers' revenue. I calibrate the HMM using a longitudinal data set that is obtained from a large mobile game company in the United States. The mobile game is a single-player puzzle-solving game, and reward ads are an important revenue source for the company.

The findings can be summarized as follows. First, watching reward ads can help players to transition to a higher engagement level, and the magnitude of the positive effect diminishes as players transition to higher engagement states. Second, a higher fluctuation of perceived difficulty increases the probability that players move to a higher state. However, the second-order effect is negative, which means that the function is concave and the marginal effect diminishes. Third, I find that the perceived difficulty is helpful for players who are in a low or high engagement state, but for players in a medium engagement state, a larger perceived difficulty results in a lower engagement state. At the same time, a feeling of achievement has a positive effect on players across three states and is most effective for players in a medium engagement state. Furthermore, I study the interaction effect between reward ads and players' feeling of competence, and I find that reward ads is more helpful in engaging players when perceived difficulty or perceived achievement is higher. Overall, I find that player's engagement transition is a complex process, and players in high, medium, or low engagement states respond differently to different factors. Finally, when examining factors that affect reward ads watching behavior, I find that players who are in a low engagement state are most likely to watch reward ads.

The essay makes several important theoretical contributions, especially to the extant literature on the self-determination and motivation theory. First, the findings enhance and deepen our understanding of players' engagement and its dynamic evolution process. I contribute to the literature in motivation theory by using real-world mobile game data and providing empirical evidence on how theory-based intrinsic motivation factors affect players' engagement. I track players' perception of intrinsic motivation factors at a granular level and examine their dynamic impact on players' engagement transitions by using detailed tap-stream data. Second, I extend the motivation theory by examining the effect of fluctuations in gamers' perceptions on their engagement. Specifically, I examine the effect of the fluctuation of perceived difficulty on players' engagement evolution, which has not been empirically studied in previous literature. Finally, I employ intrinsic motivation theory to examine the effects of reward ads on players' engagement. Reward ad is an innovative advertisement format that helps mitigate the nuisance cost of advertisement and, at the same time, generate revenue for game publishers. To the best of my knowledge, this is the first essay to link reward ads to intrinsic motivation theory and provide empirical evidence on how reward ad affects players' engagement levels. Finally, distinguishing intrinsic motivation from extrinsic or social motivation through observational data is often challenging due to omitted variable concerns. Fortunately, the single-player setting has no social interactions (e.g., connecting with friends through other social media channels or having a ranking system among players), which alleviates the concern of confounding the effect of social motivation with that of intrinsic motivation, helps to identify the effect of those intrinsically related motivation factors, and enriches the motivation theory.

3.2 Context and Model

3.2.1 Model setting

I examine individual-level game-playing data from an anonymous mobile game app company in the United States. The company has over 100 million downloads across its portfolio of games. The game that I focus on is a single-player puzzle solving game that has over 200 thousand monthly active users. Puzzle games are ranked as the top three most popular mobile game genres (Udonis 2020), and the game in my setting was nominated for the "Best Puzzle Game" at The Independent Game Developers' Association Awards in 2018.[4] In each puzzle, players need to spell the right word based on hints from two pictures, and they earn a fixed number of virtual coins after successfully solving a puzzle. Players can use virtual coins to buy premium hints, such as revealing one letter, revealing one word in the picture, or skipping the current puzzle. Players obtain coins mainly in two ways: by solving the puzzle or by watching the reward ads.[5] The reward ad is a 30-second video advertisement, which is administrated by a third party, and the reward ad usually advertises other mobile apps. Players receive a fixed number of virtual coins after watching a reward ad. Watching a reward ad is a completely voluntary action. In other words, a player needs to click on a reward ad button to watch reward ads and receive virtual coins as a reward.[6] Players can choose to watch as many reward ads as they need to earn enough coins. Watching reward ads can be especially useful when the current puzzle is too difficult to solve and when players have a low coin balance.

The research setting is ideal for examining intrinsic motivation factors. First, as previously mentioned, the puzzle-solving game I examine has no social interactions through

---

[4]`https://tiga.org/news/tiga-announces-games-industry-awards-2018-finalists`

[5]Although players can make in-game purchases to get virtual coins, it is rare in my data and the majority of players obtain virtual coins through solving puzzles and watching reward ads.

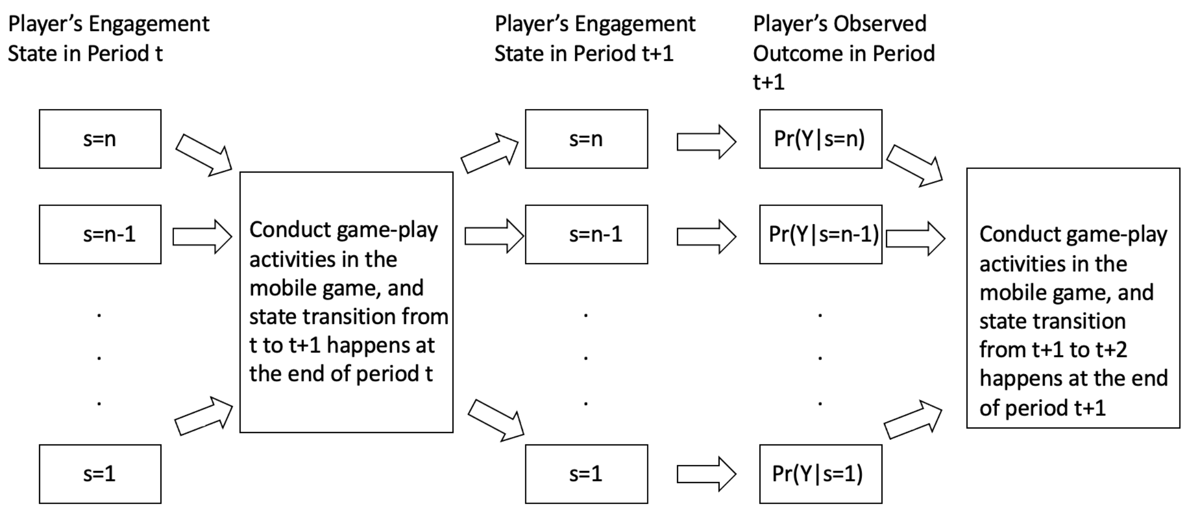
[6]I note that a reward ad is different from pop-up ads that are automatically shown that players are coerced to watch.

other social media channels or ranking system among players, which alleviates the concern of confounding the effect of social motivation with that of intrinsic motivation. In fact, compared with players in other game genres, players in puzzle games are less interested in social engagement and are more motivated by intrinsic challenges that the puzzle provides. Finally, the puzzle sequence is pre-determined in the game design phase and there is no dynamic puzzle difficulty adjustment based on players' skills or playing patterns. In other words, all players face the same puzzle sequence regardless of players' expertise in the game, which helps us to identify how variations in intrinsic motivation factors affect engagement evolution.

3.2.2 Hidden Markov model (HMM)

I use a Hidden Markov Model (HMM) that identifies a player's engagement as a latent state, and the transition across different engagement stages is a Markovian process. The HMM that I adopt is widely used in the marketing and information systems fields (Zhang et al. 2019, Netzer et al. 2008, Singh et al. 2011, Lee et al. 2018a). An HMM is a model of stochastic process that is not directly observed but can be observed through another set of stochastic processes that produce a set of observations (MacDonald and Zucchini 1997). The HMM is suitable for my context because players' engagement level is not directly observed, but is rather a reflection of motivation which leads to the observed activities (e.g., game-play decisions, reward ads watching). Moreover, the HMM allows us to segment players according to their engagement levels and investigate the varying impacts of intrinsic related factors on engagement evolution. In the HMM, game-play decision and reward ads volume are set as the observed outcomes that are stochastically determined by hidden engagement states. Although the in-app purchase can also directly contribute to the game publisher's revenue, the rate of purchase is very low (less than 1 percent) in my setting and does not provide meaningful insights. Therefore, in this paper, I focus on playing decision and reward ads volume. Players first decide whether to play the game or not in each period, and then conditional on playing the game, they decide the number of reward ads to watch. Specifically,

Figure 3.1: Hidden Markov Model of Player's Engagement State Transition



I denote $d_{it} = 1$ if a player i played the game on day t and solved at least one puzzle on day t. When $d_{it} = 1$, I observe and examine the number of reward ads that the player watched. If a player i did not solve any puzzle on day t, I set $d_{it} = 0$ and I do not observe the reward ads watching behavior.[7] Figure 3.1 presents the framework of the HMM. The figure illustrates how players transition between different levels of engagement through game-play and reward ads watching activities and how their observed outcomes depend on the latent engagement states. Similar to other HMMs, the proposed HMM in my paper comprises three elements: the initial state distribution, π; the state-transition matrix, Q; and the observed outcome distribution, P.

[7]Note that it could be possible that a player watched reward ads without successfully solving a single puzzle in period t. In such a case, the number of reward ads watched in day t would not be zero whereas the number of puzzles solved would be zero (i.e., $d_{it} = 0$). However, upon examining the data, it only accounts for around 1.5% of the total records and I believe excluding them will not affect the findings.

3.2.3 State transition matrix

I model the transitions between states as a Markov process and assume there are n states, with 1 denoting the lowest state and n denoting the highest one. In the main HMM model, I relax the assumption of a random walk and allow players to drop from the highest to the lowest (or rise from the lowest to the highest) levels directly to capture the possibility that they suddenly lose or build interest in the game. To be specific, the state transition probability is defined as $q_{it}(j,m) = p(s_{it+1} = m | s_{it} = j), 1 \leqslant j, m \leqslant n$; and $\sum_{m=1}^{n} q_{it}(j,m) = 1, 0 \leqslant q_{it}(j,m) \leqslant 1$.

Following the literature on HMM, the state transition matrix is defined as an ordered logit model:

$$
\begin{aligned}
q_{it}(j,n) &= 1 - \frac{exp(\mu(j,n) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,n) - \alpha_j * X_{it} - \eta_i)};\\
q_{it}(j,n-1) &= \frac{exp(\mu(j,n) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,n) - \alpha_j * X_{it} - \eta_i)} - \frac{exp(\mu(j,n-1) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,n-1) - \alpha_j * X_{it} - \eta_i)};\\
&\ldots\\
q_{it}(j,j) &= \frac{exp(\mu(j,j+1) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,j+1) - \alpha_j * X_{it} - \eta_i)} - \frac{exp(\mu(j,j-1) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,j-1) - \alpha_j * X_{it} - \eta_i)};\\
&\ldots\\
q_{it}(j,2) &= \frac{exp(\mu(j,2) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,2) - \alpha_j * X_{it} - \eta_i)} - \frac{exp(\mu(j,1) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,1) - \alpha_j * X_{it} - \eta_i)};\\
q_{it}(j,1) &= \frac{exp(\mu(j,1) - \alpha_j * X_{it} - \eta_i)}{1 + exp(\mu(j,1) - \alpha_j * X_{it} - \eta_i)};
\end{aligned}
\tag{3.1}
$$

where $\mu(j,m)$ is the threshold for the transition from state j to state m, and $\mu(j,n) \geqslant \mu(j,n-1) \geqslant ... \geqslant \mu(j,j+1) \geqslant \mu(j,j-1) \geqslant ... \geqslant \mu(j,1)$. X_{it} is a vector of covariates that affect state transition, α_j is a vector of corresponding coefficients, and η_i is the player specific random effect to capture the individual unobserved heterogeneity that may affect state transition. Note that in the main HMM model, I assume that there is no state transition if a player i does not play in period t (i.e.,$d_{it} = 0$). In the robustness check presented in Section 3.5.2, I relax this assumption and assume that a different set of variables affects state transitions when $d_{it} = 0$; the results remain consistent.

3.2.4 State-dependent outcome

Given the player's state, the observed outcomes are assumed to be conditionally independent. In the model, state-dependent outcomes are observed players' playing decisions and reward ads watching volume. Players first decide whether to play the game or not in each period, and then conditional on playing the game, they decide on how many reward ads to watch. I denote the decision to play as a logit model (i.e.,$Pr(d_{it}=1)=exp(\beta_j)/(1+exp(\beta_j)))$ where β_j is a state-dependent constant, and the number of reward ads as a negative binomial distribution. That is:

$$Pr(Y_{it}|S_{it}=j)=\frac{\Gamma(\theta_j^2+Y_{it})}{\Gamma(Y_{it}+1)\Gamma(\theta_j^2)}(\frac{\theta_j^2}{\theta_j^2+\lambda_j})^{\theta_j^2}(\frac{\lambda_j}{\theta_j^2+\lambda_j})^{Y_{it}} \tag{3.2}$$

where Y_{it} is the number of reward ads for player i in period t, and θ_j is the over-dispersion rate of Y_{it}. In addition, $\lambda_j=exp(\gamma_j Z_{it}+\zeta_i)$, where Z_{it} is a vector of covariates that affect the observed outcome, γ_j is a vector of corresponding coefficients, and ζ_i is the player specific random effect to capture the individual unobserved heterogeneity that may affect the reward ads volume. I allow the player-specific random effects η_i and ζ_i to be correlated by modeling the conditional distribution of the two random effects.

I can write the likelihood function as follows:

$$\begin{gathered}L(Y(i)|\zeta,\eta)=\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}...\sum_{s_{T_i}=1}^{n}Pr(s_{i1}=s_1)\prod_{t=2}^{T_i}Pr(s_{it}=s_t|s_{it-1}=s_{t-1})\times\\ \prod_{t=1}^{T_i}(Pr(d_{it}=1|s_{it}=s_t)\times Pr(Y_{it}|s_{it}=s_t))^{d_{it}=1}\times(1-Pr(d_{it}=1|s_{it}=s_t))^{d_{it}=0}\end{gathered} \tag{3.3}$$

where $s_t \in 1,2,...,n$ is the state that a player can possibly reside in period t. T_i is the length of player i's outcome sequence. Note that I model the process from the time that players start to play the game until their last observation in the time window. Thus, different players can have different outcome length, and T_i varies across players. Finally, the likelihood of player i can be obtained by integrating over η and ζ.

$$L(Y(i))=\int_\zeta\int_\eta L(Y(i)|\zeta,\eta)dG(\eta|\zeta)dH(\zeta) \tag{3.4}$$

3.3 Data and Variable Construction

Mobile games are an important component of the mobile app market ecosystem, and factors that affect users' engagement in mobile games may differ from factors that affect users' engagement in other non-game apps, including shopping, traveling, news, and social networking apps.[8] Mobile games are mainly used for entertainment purposes and players have repeated interactions with mobile games. In fact, an important characteristic of mobile games is that players get immediate feedback on their actions and players can feel a sense of achievement by successfully playing the game. Therefore, compared with other non-game apps, intrinsic motivation factors can have stronger effects on mobile game app users' engagement.

To examine players' engagement evolution and reward ads watching behaviors, I obtain players' tapstream data of a puzzle game from a U.S. based mobile game app company. I randomly select 1,232 players who joined the mobile game app on or after January 1, 2018. In Appendix Table A.3, I show that the results are robust to employing alternative user samples. The data include players' complete game-play tapstream from January 1 to February 28, 2018. I aggregate in-game activities, including game-play decisions and reward ad watching volume, to a daily level, and time period t in the HMM refers to date t. The aggregation approach is appropriate in this setting and provides multiple advantages. First, players can choose to solve a puzzle without watching any reward ads, which can result in many zeros for puzzle-level observations. This can adversely affect the identification of the HMM parameters (Huang et al. 2019). Second, the average time to solve a single puzzle is around 80 seconds, and state transitions are unlikely to happen so frequently. Table 3.1 provides a summary of variable descriptions and statistics. As shown in the model in Section 3.2.2, the HMM has two important elements: state transition matrix and state-dependent outcome. Therefore, two sets of variables are constructed: variables affecting state-transition and variables affecting state-dependent outcomes. In the remaining section, I describe those variables in detail.

---

[8] `https://medium.com/tapjoy/tapjoy-research-why-do-people-play-mobile-games-930182e872c2`

3.3.1 Variables affecting state-transition

By extending the literature on intrinsic motivation theory, I first examine how players' feeling of competence (i.e., *perceived difficulty* (Malone 1981) and *perceived achievement* (Deci et al. 1985)) affect engagement state transitions. In a single-player game context, puzzle-solving time is a natural way to measure how difficult the player feels about the puzzle. I measure the *perceived difficulty* by using each player's average puzzle-solving time in period t. Positive feedback from the game can induce a sense of achievement, which in turn increases players' engagement (Ryan et al. 2006). In this research setting, the number of successfully solved puzzles in each period t can be a measurement of players' achievement in the game. When playing the puzzle, players can make soft currency purchases, which refers to using virtual coins to buy premium hints (e.g., revealing one letter, revealing one word in the picture, or skipping the current puzzle) to help solve the puzzle. As soft currency purchases help lower the puzzle's difficulty level, players may attribute puzzle completion after making soft currency purchases less to their own ability. Therefore, *perceived achievement* is constructed as the total number of successfully solved puzzles divided by the number of soft currency purchases in each period. However, if players do not make any soft currency purchases in a given period, achievement goes to infinity. To account for this situation, I invert achievement and adopt *perceived achievement*$^{-1}$ in my estimation model.

Second, I examine round-to-round shifts in perceived difficulty, which has not been studied in prior empirical literature. *Fluctuation of perceived difficultly*, which is measured as the standard deviation of puzzle-solving time across all puzzles solved in period t. I also include its quadratic term to examine whether there is a nonlinear relationship between the fluctuation of perceived difficulty and the propensity of moving to or staying at a higher engagement state. To alleviate the potential correlation between the perceived difficulty and the fluctuation of perceived difficulty, I standardize the fluctuation of perceived difficulty by dividing it by the perceived difficulty.

Third, to examine how reward ads affect players' intrinsic motivations and help keep players engaged, I construct the variable *reward ads*, which measures the number of reward ads watched in period t. I examine the effects of reward ads on state evolution from period t to period $t+1$. In addition, I include the interaction terms between *reward ads* and players' feeling of competence (i.e., *perceived difficulty* and *perceived achievement*$^{-1}$), to examine how reward ads and players' competence interact and affect the players' engagement transitions. I examine the interaction effects as reward ads can help lower the perceived difficulty of puzzles and provide players with an opportunity to further explore the game. Finally, similar to most mobile apps, there can be a satiation or addiction effect among players as they spend more time on the mobile game (Kwon et al. 2016, Han et al. 2015). Therefore, I construct *cumulative playing time* until period t to examine its effect on engagement state transitions. I note that all variables that affect state transitions are constructed at the individual level by using the detailed tapstream data.

3.3.2 Variables affecting observed outcomes

As previously discussed in the model section, the state-dependent observed outcomes are players' playing decisions and watched reward ad volume. In each period, a player first decides whether to play the game, and then decides on how many reward ads to watch conditional on playing the game. A state-dependent constant β_j is used to model the playing decision, and I adopt a normalization requirement that β_j is ordered, which ensures identification of the states. This indicates that, without any stimulus, players in a higher engagement state are more likely to play the game than players in a lower engagement state.

As I additionally aim to identify factors that affect players' reward ad volume, I introduce other variables that can directly affect reward ad watching behaviors. As a direct benefit of watching reward ads, players collect virtual coins as a reward. Therefore, coin-related factors should affect players' reward ad watching behavior. More specifically, I examine variables *coin balance*, *soft currency purchase*, and *in-app purchases*. *Coin balance* measures the coin balance on a player's account at the beginning of each period. *Soft currency purchase* measures the number of virtual coins that a player spends to purchase premium hints (e.g., revealing one letter, revealing one word in the picture, or skipping the current puzzle) in each period. Players can buy virtual coins using real money, and *in-app purchases* measures the number of in-app purchases in each period. All of the aforementioned three variables reflect the extent to which players are in need of virtual coins, which are used to help reduce the difficulty of a puzzle. In addition, I control for the average difficulty of a puzzle by including a variable *puzzle difficulty*, which is computed at the population level. The variable accounts for how one puzzle may be generally more difficult than another puzzle. Specifically, to measure the difficulty of puzzle k, I use the average puzzle-solving time of puzzle k among all players who played that puzzle. The underlying assumption is that, if all players take a longer time to solve puzzle k, puzzle k would be more difficult than other puzzles.

Furthermore, to capture how previous game-playing behaviors relate to reward ad watching behaviors, which can help provide strategic guidance to game publishers aiming to target ads, I construct three variables that describe players' game-play patterns: *frequency*, *latest intensity*, *historical intensity*, and *irregularity*. The variable *frequency* is constructed as the number of game-playing active days (i.e., a day is considered active if at least one puzzle is solved on that day) until time t divided by the total days elapsed since the player joined the platform. To capture players' game-playing intensity, I include *latest intensity*, which captures the intensity in the most recent active period, and *historical intensity*, which captures the intensity before the most recent active period. Specifically, I compute the number of puzzles solved in most recent active period to measure *latest intensity*, and the total number of puzzles solved in days before the most recent active period divided by the number of active days before the most recent active period to construct the variable *historical intensity*. The variable *irregularity* is defined as the standard deviation of time intervals among active days until time t. For example, if the time interval among three active days is Δt_1 and Δt_2, the standard deviation is be calculated as $sd = \sqrt{\frac{\sum_{i=1}^{n}(\Delta t_i - \overline{\Delta}t)^2}{n-1}}$ where $n = 2$ in this example, and $\overline{\Delta}t$ is the average of Δt_1 and Δt_2. The higher the standard deviation, the more irregular playing pattern the player has. Finally, I include two indicator variables *weekend* (i.e., a variable that is equal to one if period t is Saturday or Sunday, and zero otherwise) and *work*

Table 3.1: Data Description and Statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Variables Affecting Observed Outcome					
Coin balance	The coin balance at time t	345.50	430.73	0	9129
Soft currency purchase	The total number of coins spent at time t	203.98	226.39	0	2835
In-app purchases	The total number of in-app purchases at time t	0.01	0.10	0	3
Puzzle difficulty	The average puzzle difficulty a gamer played at time t	81.20	35.47	24.26	523.24
Frequency	The number of active days till time t divided by the duration since join the platform	0.47	0.29	0.02	1
Latest intensity	The number of played puzzles in most recent active period	11.53	15.61	0	157
Historical intensity	The total number of played puzzles divided by the active days before the most recent active period	12.82	14.65	0	131
Irregularity	The standard deviation of historical intervals before t	1.27	2.26	0	28.28
Weekend	A dummy variable: whether period t in on Saturday or Sunday	0.28	0.45	0	1
Work hour	A dummy variable: whether playing time is between 9am to 5pm	0.26	0.44	0	1
Variables Affecting State Transition					
Fluctuation of perceived difficulty	The standard deviation of puzzle solving time at time t	93.30	82.86	0	898.73
Perceived difficulty	The average puzzle solving time in seconds at time t	99.76	75.62	12	872.50
Reward ads	The number reward ads at time t	1.77	6.56	0	123
Perceived achievement^{-1}	The total soft currency purchases divided by the number of solved puzzles at time t	26.95	30.95	0	318.33
Cumulative playing time	The cumulative playing time in minutes until time t	117.96	161.76	0.5	1737.35

hour (i.e., a variable that is equal to one if the playing time is between 9 am and 5 pm, and zero otherwise), to control for the time availability of players.

3.4 *Estimation Results and Findings*

In this section, I discuss the model selection and present the estimation results.

3.4.1 Model selection

In the HMM, the number of hidden states is unknown, and the model needs to find the number of hidden states that best fits the data. I use maximum likelihood estimation to estimate the HMM parameters. Table 3.2 shows that the model with three hidden states outperforms all other specifications because it has a smaller Bayesian information criterion (BIC) and Akaike information criterion (AIC). The three states represent low, medium, and high engagement states. I start with a latent class model to estimate the initial distribution for players' hidden engagement states. To avoid the label switching, I add constraints on the intercept in the logit probability of playing, such that $\beta_L < \beta_M < \beta_H$, to ensure that players in higher engagement states have a higher intrinsic propensity to play the game. I follow

Table 3.2: Selection of the Number of States

Number of states	Log-likelihood	BIC	AIC	Number of variables
1	-23275.71	46643.93	46577.42	13
2	-22520.22	45353.56	45128.44	44
3	**-20258.46**	**41007.95**	**40654.92**	**69**
4	-20319.84	41322.85	40831.68	96

Heckman and Singer (1984) to use a non-parametric approach to estimate the random effect, ζ and η. The approximation process for the underlying unknown probability distribution is evaluated by a finite set of supporters associated with their corresponding probability mass functions. I relate two normalizing constants C_ζ and C_η with the supporters and set the boundary on each random effect variable to be between zero and one. The optimal number of supporters is chosen by model fit, and Appendix A.1 presents the corresponding Log-likelihood value with different combination of supporter numbers. In my study, the optimal number for ζ and η is three and two, respectively.

3.4.2 Dynamic engagement state transitions

Table 3.3 reports the estimated parameters of the three-state HMM. In the table, I label the three engagement states as low, medium, and high. The estimates of thresholds in Table 3.3 represents the minimum effort that the players need to make to transition to other states. As expected, the threshold estimates for moving to a higher engagement state are positive whereas threshold estimates for moving to lower states are negative.

I first examine the effects of intrinsic motivation factors on state transition probabilities (lower panel in Table 3.3). First, I find that the *fluctuation of perceived difficulty* increase the probability of a transition to a higher engagement level for all three states ($9.6593, p < 0.01$,

Table 3.3: Estimated Parameters for the Three-State HMM

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	-1.6504	(1.1192)	-14.6448***	(1.3257)	-1.3962	(1.0618)
Soft currency purchase	3.4089***	(1.0739)	7.3529***	(1.0908)	2.4288**	(1.0817)
In-app purchases	-1.0441	(1.0488)	-9.0633***	(1.0796)	-0.2833	(1.0371)
Puzzle difficulty	5.6099***	(1.0154)	1.8805*	(1.0339)	10.3035***	(1.0514)
Frequency	0.0262	(1.0871)	6.0133***	(1.2164)	-2.2807**	(0.9848)
Latest intensity	-1.2109	(2.7700)	-0.1544	(1.1888)	0.6926	(1.1955)
Historical intensity	0.4203	(1.2689)	0.0742	(1.9343)	2.1005**	(1.0482)
Irregularity	-5.2184***	(1.0885)	-2.8372***	(1.0065)	-9.5794***	(1.0042)
Weekend	-0.0472	(1.0988)	-0.0796	(1.0352)	-0.0063	(0.9958)
Work hour	-0.1165	(1.1965)	-0.5938	(1.0086)	0.0775	(1.0347)
Constant	-2.1035**	(1.0560)	-7.7183***	(1.7072)	0.9280	(1.0326)
θ Dispersion rate	0.5544	(0.3850)	-0.8459	(1.0026)	-1.2760	(0.8003)
Intercept in the logit probability of playing	-1.5138*	(0.8915)	1.6205***	(0.4246)	1.6375***	(1.1560)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	9.6593***	(1.0960)	10.4182***	(1.0136)	9.6559***	(1.0592)
Fluctuation of perceived difficulty^2	-9.3569***	(1.0001)	-6.1543***	(1.0320)	-8.8212***	(1.0261)
Reward ads	11.3517***	(1.1167)	0.8976	(2.5936)	9.1034***	(1.3819)
Perceived difficulty	3.7346*	(2.0593)	-4.5620***	(1.0097)	3.0644**	(1.3821)
Perceived achievement^{-1}	-4.9319***	(1.0288)	-22.3065***	(1.8600)	-10.1723***	(1.0050)
Cumulative playing time	14.0328***	(1.0537)	4.8206***	(1.6756)	1.3152	(1.0034)
Perceived difficulty * Reward ads	5.4776***	(1.0147)	1.7745*	(1.0584)	4.5635***	(1.0058)
Perceived achievement^{-1} * Reward ads	-4.1796***	(1.0008)	-6.0095***	(1.0058)	-2.7056***	(1.0012)
Thresholds						
State 1			2.5435*	(1.3933)	4.0772	(1.2850)
State 2	-10.9055***	(1.0465)			3.3480***	(1.0139)
State 3	-10.3669***	(0.9169)	-9.7490	(1.0617)		
Unobserved Heterogeneity						
$C_\zeta = 6.4005, C_\eta = -10.1136$	$\zeta_1 = 0$		$\zeta_2 = 0.2568$		$\zeta_3 = 1$	
Probability(ζ)	0.6145		0.3845		0.0001	
Conditional Distribution: Probability($\eta\|\zeta$)						
$\eta_1 = 0$	0.0019		0.4676		0.9213	
$\eta_2 = 1$	0.9981		0.5324		0.0787	
Log Likelihood			-20196.57			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

In the model part, I assume the transition threshold $\mu(j,n) \geqslant \mu(j,n-1) \geqslant ... \geqslant \mu(j,j+1) \geqslant \mu(j,j-1) \geqslant ... \geqslant \mu(j,1)$. Therefore, $\mu(3,2) = -9.7490 = -10.3669 + exp(-0.4815)$, and the standard error in parenthesis (1.0617) is for -0.4815 which is not significant. The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

$10.4182, p < 0.01$, and $9.6559, p < 0.01$, for low, medium, and high states, respectively). The results indicate that variation in perceived difficulty meets players' demand for variations and counteracts satiation from repetition, which highlights the importance of providing stimulation and taking into consideration players' dynamic perceptions when designing a mobile game. As the quadratic term is statistically significant and negative for all states ($-9.3569, p < 0.01$, $-6.1543, p < 0.01$, and $-8.8212, p < 0.01$, for low, medium, and high states, respectively), there is a diminishing marginal effect of *fluctuation of perceived difficulty*. In other words, a moderate level of difficulty fluctuations, or suspense, in the game can help players stay engaged in the game. The results show the diminishing marginal effect of *fluctuation of perceived difficulty* is least prominent for players in a medium engagement state, compared to players in the other two states. This highlights the heterogeneous effect of the *fluctuation of perceived difficulty* on engaging players in different engagement levels.

Second, the results show that *reward ads* can increase the propensity of moving to or staying a higher engagement state for players in a either low or high engagement state ($11.3517, p < 0.01$, and $9.1034, p < 0.01$, respectively). The magnitude of the coefficient is largest for players in a low engagement state. A possible explanation for the smaller coefficient for players in a high engagement state can be that reward ads can be regarded as an easy shortcut that helps players who are stuck in the game make progress. Players in high engagement states may place greater value on competence-related intrinsic motivation factors, such as whether the player is making a meaningful achievement in the game, and thereby place less value on reward ads.

I find that higher *perceived difficulty* has positive effects on the propensity of moving to, or staying at, a higher engagement state when players are in a low or high engagement state ($3.7346, p < 0.1$ and $3.0644, p < 0.05$ respectively) whereas it has a negative effect when players are in a medium engagement state ($-4.5620, p < 0.01$). I further examine whether the mismatch between perceived difficulty and player's skill affects the engagement transitions, and the results are reported in Appendix Table A.2. The effect of *perceived achievement*$^{-1}$ is significantly negative for all three states ($-4.9319, p < 0.01$, $-22.3065, p < 0.01$, and

-10.1723, $p < 0.01$, for low, medium, and high states, respectively), which indicate that the feeling of achievement increases the probability that players move to a higher engagement level for all states. The magnitude of the coefficient is largest for players in a medium engagement state. By combining the previous two findings, I can infer that for players in a medium engagement state, the need for achievement discourages them from being engaged with a puzzle with a higher difficulty level. Next, the effect of *cumulative playing time* is significantly positive for players in low and medium engagement states (14.0328, $p < 0.01$ and 4.8206, $p < 0.01$, respectively). This indicates that an addiction effect (i.e., the more time players spend on the game, the more engaged they become) is present for players in low and medium engagement states, but the effect diminishes as players transit to higher engagement states. Finally, I examine the interaction effect between reward ads and players' feeling of competence, which is captured by *perceived difficulty* and *perceived achievement*. For all engagement states, I find that reward ads are more helpful in engaging players when the perceived difficulty is higher. A possible explanation is that players enjoy a moderate level of challenge and the flexibility and shortcut that reward ads provide can be a useful moderator to lower the perceived puzzle difficulty and help players make progress to next puzzle levels. In addition, I find that reward ads are more effective in engaging players when the feeling of achievement is higher. The effects of reward ads are stronger for players with a high feeling of achievement as reward ads provide players who have a high feeling of achievement with an additional opportunity to explore the game further (i.e., successfully solve more puzzles) and help keep such players engaged.

3.4.3 State-dependent outcomes

I now examine factors that affect the state-dependent outcomes (top panel in Table 3.3). Players' decisions to play the game is modeled as a logit model, and the state-specific intercepts in the logit probability of playing the game are -1.5138 ($p < 0.1$), 1.6205 ($p < 0.01$), and 1.6375 ($p < 0.01$), for low, medium, and high states, respectively, indicating that players in higher engagement states are more likely to play the game. It can also be noted that the gap in the intercept (i.e., the gap in the probabilities of playing the game) between the low and medium engagement states is much larger than that between the medium and high engagement states. The estimates suggest that players in a low engagement state have a significantly lower probability of playing the game, and highlights the importance of moving players, especially those in a low engagement state, to a higher engagement state to encourage game participation.

When examining the number of reward ads a player will watch if the player decides to play the game in a given period, the estimates of the constant term in the negative binomial model indicate that players in a low engagement are more likely to watch reward ads compared to players in a medium engagement level. Next, as the primary motivation for watching reward ads is to collect virtual coins, I examine the extent to which virtual coins related factors affect reward ad volume. I find that that *coin balance* at the beginning of each period has statistically significant and negative effects on reward ad volume for players who are in a medium engagement state (-14.6448, $p < 0.01$, respectively). I additionally note that consumption of virtual coins via *soft currency purchases* increases reward ad volume for all three states (3.4089, $p < 0.01$, 7.3529, $p < 0.01$, and 2.4288, $p < 0.05$, for low, medium, and high states, respectively). Additionally, although the *in-app purchase* rate is very low in my setting, I control the *in-app purchase*, and find that it significantly reduces the ads volume for players in a medium engagement level ($-9.0633, p < 0.01$). Furthermore, I find that players tend to watch more reward ads when the *puzzle difficulty* increases (5.6099, $p < 0.01$, 1.8805, $p < 0.1$, and 10.3035, $p < 0.01$ for low, medium, and high states, respectively). I additionally control for the effects of players' historical game-play on players' reward ad watching behavior patterns by controlling for *frequency*, *latest intensity*, *historical intensity*, and *irregularity*. I find that historical game-play patterns have different effects on reward ad volume for players in different states. For example, *frequency* is positively associated with reward ads watching behavior for players in a medium engagement state (6.0133, $p < 0.01$), the *historical intensity* is positively associated with ads volume for players in a high engagement state (2.1005, $p < 0.1$), and players tend to watch fewer ads when the playing pattern is more irregular for all states (-5.2184, $p < 0.01$, -2.8372, $p < 0.01$ and -9.5794, $p < 0.01$, for low, medium, and high states, respectively). The estimates can help game publishers, who are aiming to optimize reward ad volume, employ different advertising strategies based on players' historical game patterns.

3.4.4 Posterior and counterfactual analysis

Based on the estimation results of parameters in the HMM, I recover players' posterior probability of being classified into a state through a forward-backward algorithm (Yu and Kobayashi 2003). Then, I make a random draw based on the posterior probabilities to decide which state the player is classified into. I find that the average proportion of players belongs to a low, medium, and high engagement state is 0.77, 0.13, and 0.10, respectively. In fact, the hidden state distribution is consistent with the industry observations that many mobile games suffer from the low-engagement problem and highlights the need to improve players' engagement.[9]

Next, I evaluate different game design policies that publishers can adopt to improve players' engagement through simulation experiments. I consider three different game design policies and examine their effects on players' engagement transitions. First, I double the value of *fluctuation of perceived difficulty* to simulate a game design where the perceived difficulty becomes more volatile. Second, I double the value of *perceived difficulty* to simulate a game design where the puzzle becomes more difficult. Finally, I double the value of *perceived achievement* (i.e., halve the value of *perceived achievement*$^{-1}$). Table 3.4a, 3.4b, 3.4c present the changes in the transition probabilities in each scenario respectively. For example, in Table 3.4a, when *perceived difficulty fluctuation* is doubled, the probability that a player in a low engagement state moving to a medium engagement state increases by 1.19% and moving to a high engagement state increases by 0.51%. Similarly, in Table 3.4c, when *perceived*

[9] The posterior distribution also reflects the model specification as I assume that if a player doesn't play on date t, the latent engagement remains the same. Since a player is more likely to be in a low engagement state when she is not playing, I may observe this high proportion of low engagement state.

achievement is doubled, the probability that a player in a medium engagement level moving to a high engagement increases by 0.12% and staying in the current medium engagement state increases by 5.89%. The results indicate that doubling the fluctuation of perceived difficulty or doubling the perceived achievement increases the probability of moving to, or staying at, a higher engagement state for players in all three states. Second, doubling perceived difficulty decreases the probability of players in a medium engagement state moving to, or staying at, a higher engagement state, but increases the probability of engaging for players in the other two states, which confirms my findings that for players in a medium engagement state, the need for achievement discourages them from being engaged with a puzzle with a higher difficulty level. The counterfactual analysis highlights the importance of examining the heterogeneous effects of intrinsic motivation related factors on players in different engagement states. It is important for game publishers to strategically employ players' intrinsic motivation factors to improve players' engagements.

3.5 Robustness Checks

I conduct a set of robustness checks to ensure the robustness of my results. Overall, I find that the results are robust to alternative model structures and variable measurements.

3.5.1 Alternative model: zero-inflated negative binomial model

In the HMM, a player first decides whether to play the game, and then decides on how many reward ads to watch conditional on playing the game. The player can choose to play a game without watching reward ads, which may result in a zero-inflation concern of the observed reward ad outcome. Therefore, I use a zero-inflated negative binomial model to account for zero observations. To be specific, the distribution of the number of reward ads volume can be written as:

$$Pr(Y_{it}|S_{it}=j) = \begin{cases} \pi_j + (1-\pi_j)g(0|S_{it}=j) & if Y_{it}=0; \\ (1-\pi_j)g(Y_{it}|S_{it}=j) & if Y_{it}>0 \end{cases}$$

Table 3.4: Changes in Transition Probability

	Low	Medium	High
Low	-0.0170	0.0119	0.0051
Medium	-0.0636	0.0497	0.0139
High	-0.0831	-0.0279	0.1110

(a) Doubling *Perceived Difficulty Fluctuation*

	Low	Medium	High
Low	-0.0077	0.0057	0.0020
Medium	0.0912	-0.0898	-0.0014
High	-0.0574	-0.0078	0.0652

(b) Doubling *Perceived Difficulty*

	Low	Medium	High
Low	-0.0008	0.0006	0.0002
Medium	-0.0601	0.0589	0.0012
High	-0.0282	0.0022	0.0260

(c) Doubling *Perceived Achievement*

Note: For each counterfactual simulation, each cell in the table shows the change in probability of moving from the state labeled in the first column to the state labeled in the first row.

where π_j is the zero inflation rate that is state dependent (i.e., $\pi_j = exp(\kappa_j)/(1 + exp(\kappa_j)))$ and $g(.)$ is the probability mass function of negative binomial distribution. The estimation results are reported in Table 3.5. The estimations of π_j across three states are very close to zero, which indicates that zero inflation is not a concern in my data, and the negative binomial model employed in the original HMM is enough to explain the data variation. Furthermore, Table 3.5 shows that all the main results remain robust to employing the zero-inflated negative binomial model.

3.5.2 Alternative state transition assumption

In the main HMM, I assume that players stay at the same engagement level and there is no state transition when players are not playing the game (i.e.,$d_{it} = 0$). In this robustness check, I allow state-transition to take place in periods when players have no game-play activities (i.e.,$d_{it} = 0$). Specifically, I construct a variable *event interval* (i.e., days elapsed since the last game playing activity) that affects the state transition when $d_{it} = 0$. I include a state-dependent constant ω_j to determine the intrinsic (baseline) transition propensity for state j when $d_{it} = 0$. The estimation results are shown in Table 3.6. The estimates of state-dependent constant ω_j reveal that players in higher engagement states generally have a higher intrinsic propensity of moving toward higher engagement levels. In addition, the longer the interval since the last playing activity, the less likely it is that players will move to or stay in a higher engagement state. The magnitude of this statistically significant and negative effect is largest for players in a low engagement state and smallest for players in a high engagement state. At the same time, Table 3.6 shows that all the main results remain robust to allowing state transitions in a no-game-play period.

3.5.3 Alternative variable measurements

I further investigate whether the findings are sensitive to different measurements of variables. In the main HMM, I use the standard deviation of puzzle-solving time as a proxy for the *fluctuation of perceived difficulty*. In this robustness check, I use an alternative measurement

Table 3.5: Alternative Model: Zero-Inflated Negative Binomial Model

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−1.7351*	(1.0216)	−14.3312***	(1.0589)	−1.4768	(1.1963)
Soft currency purchase	3.4662***	(1.0034)	7.2915***	(1.0327)	2.4231**	(1.0078)
In-app purchases	−1.0936	(1.0170)	−9.5923***	(1.0007)	−0.2982	(1.0501)
Puzzle difficulty	4.5166***	(1.0055)	1.6836*	(1.0083)	11.2432***	(1.1297)
Frequency	0.2394	(1.0938)	5.4956***	(1.0075)	−1.7814	(1.1370)
Latest intensity	−1.5075	(2.9467)	−0.2871	(1.1632)	0.7532	(1.6241)
Historical intensity	0.5819	(1.1512)	0.2955	(1.0038)	2.0737**	(1.0315)
Irregularity	−4.9061***	(1.0042)	−0.8034	(1.0001)	−8.7463***	(1.0016)
Weekend	0.0053	(1.1423)	−0.0556	(1.0124)	−0.0043	(0.9964)
Work hour	−0.1314	(1.2600)	−0.6108	(1.0286)	0.0878	(0.9999)
Constant	−2.1905*	(1.1532)	−7.2395***	(1.0005)	0.3729	(0.9403)
θ Dispersion rate	0.5579	(0.6387)	−0.8579	(0.9042)	−1.2495	(0.8621)
κ zero inflation intercept	−8.0004***	(1.0001)	−9.0011***	(1.0001)	−10.0029***	(1.0001)
Intercept in the logit probability of playing	−1.5038***	(0.5263)	1.5460*	(0.6265)	1.5640***	(1.0006)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	9.5995***	(1.0065)	9.8478***	(1.0007)	10.0225***	(1.0048)
Fluctuation of perceived difficulty^2	−10.6610***	(1.0001)	−4.6544***	(1.0022)	−6.5956***	(1.0023)
Reward ads	9.2226***	(1.0545)	2.7658**	(1.3198)	7.8879***	(1.1062)
Perceived difficulty	3.3978***	(1.0052)	−4.6087***	(1.0247)	3.8894***	(1.0212)
Perceived achievement^{-1}	−4.1157***	(1.0029)	−22.4429***	(1.0444)	−10.9603***	(1.0014)
Cumulative playing time	13.7127***	(1.0926)	5.6722***	(1.0101)	1.4845	(1.0023)
Perceived difficulty * Reward ads	3.5442***	(1.0001)	3.4789***	(1.0078)	5.8037***	(1.0002)
Perceived achievement^{-1} * Reward ads	−3.6545***	(1.0001)	−4.9857***	(1.0007)	−3.8709***	(1.0002)
Thresholds						
State 1			2.9288***	(1.0406)	2.9324***	(1.0371)
State 2	−10.5664***	(0.8861)			16.9560***	(1.1001)
State 3	−10.0981***	(0.9465)	−9.3702	(0.9263)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20234.50			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.6: Alternative Assumption: Allowing State-Transition when $d_{it} = 0$

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−2.8389	(2.934 1)	−14.9611***	(1.700 3)	−1.1625	(0.933 6)
Soft currency purchase	3.7215	(2.369 2)	7.5491***	(1.056 1)	2.5456**	(1.091 3)
In-app purchases	−0.9728	(1.167 2)	−9.8862***	(1.008 8)	−0.2015	(1.362 4)
Puzzle difficulty	4.9752***	(1.136 2)	1.4811	(1.004 9)	10.8136***	(1.303 3)
Frequency	0.2228	(1.057 9)	5.7054***	(1.769 1)	−2.1131*	(1.144 2)
Latest intensity	−1.8953	(1.200 2)	−0.0695	(1.217 3)	0.7637	(1.481 6)
Historical intensity	0.7341	(1.297 2)	0.1396	(1.087 6)	2.1304	(1.380 6)
Irregularity	−4.9083***	(1.073 1)	−0.7408	(1.014 9)	−8.3440***	(1.071 1)
Weekend	−0.0794	(1.156 5)	−0.1115	(1.458 8)	0.0040	(0.963 3)
Work hour	−0.1099	(1.447 1)	−0.6530	(1.620 7)	0.0442	(1.084 1)
Constant	−2.0876**	(1.045 4)	−7.5083***	(1.862 2)	0.5058	(1.055 7)
θ Dispersion rate	0.6300	(0.904 7)	−0.8201	(1.009 5)	−1.3530	(1.193 7)
Intercept in the logit probability of playing	−1.6331***	(0.527 7)	1.8782***	(0.296 1)	1.8912***	(1.213 7)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	10.1250***	(1.091 4)	9.3230***	(1.120 1)	9.6309***	(1.015 1)
Fluctuation of perceived difficulty^2	−10.6392***	(1.000 1)	−5.1356***	(1.039 8)	−6.9744***	(1.013 5)
Reward ads	9.2044***	(1.007 9)	−1.8779	(1.631 8)	8.7521***	(1.040 9)
Perceived difficulty	2.7954**	(1.106 6)	−4.1449***	(1.067 4)	1.1435	(1.537 1)
Perceived achievement^{-1}	−4.5717***	(1.064 3)	−24.3264***	(1.391 1)	−11.0875***	(1.128 1)
Cumulative playing time	15.3048***	(1.326 2)	3.9697**	(1.579 3)	1.7585	(1.224 0)
Perceived difficulty * Reward ads	3.3009***	(1.008 7)	2.7799***	(1.016 3)	5.3263***	(1.034 0)
Perceived achievement^{-1} * Reward ads	−3.7544***	(1.001 6)	−5.1887***	(1.001 1)	−3.9485***	(1.000 3)
No-play state dependent constant ω_j	2.3168**	(1.164 1)	7.2048***	(1.002 6)	8.2083***	(1.003 2)
No-play event interval	−13.5528***	(2.833 0)	−2.9972***	(1.000 1)	−1.9976**	(1.000 1)
Thresholds						
State 1			3.0693***	(0.990 9)	4.4341	(1.282 6)
State 2	−11.2965***	(0.968 5)			18.0772***	(1.000 2)
State 3	−10.9199***	(1.042 9)	−10.3656	(1.561 1)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20156.19			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

that more granularly measures how the puzzle difficulty changes over time. To be specific, in each period t, for $k \in 2, 3, ..., K$ where K is the total number of played puzzles at time t, I calculate the absolute difference between the average puzzle-solving time of all puzzles before puzzle k and the puzzle-solving time at puzzle k. The gap Δ_k measures the perceived difficulty change in the current puzzle k compared with puzzles before k. Then, I take an average over all Δ_k for $k \in 2, 3, ..., K$ as a proxy for the overall *fluctuation of perceived difficulty* in period t. The estimation results are shown in Table 3.7. The estimates of parameters in this HMM model are similar to those in the original HMM model, and all the main results remain robust and consistent when employing an alternative measurement of the fluctuation of perceived difficulty.

3.6 Discussion and Conclusion

Mobile games have become an important component in the entertainment industry. Similar to other mobile apps, users' (players') engagement and activity levels are key performance indicators, and improving engagement levels is strongly related to a mobile game's success in the market. At the same time, game publishers also need to monetize the app. Many mobile games adopt a freemium business model, and due to low in-game purchase rates, in-app advertising has been a significant revenue source for many mobile games. In-app advertising creates a dilemma for game publishers, as in-game advertisements generate revenue, but the advertisement can interrupt the flow of the game and harm players' in-app experience. Reward ads can offer a solution that balances publishers' needs for revenue and players' experiences in the game.

In this paper, I use a HMM to examine how competence-related intrinsic motivation factors affect players' dynamic engagement transitions. First, I extend the literature on SDT to find that players' feeling of competence, including perceived difficulty and perceived achievement, has significant effects on players' latent engagement evolution, and the effect is heterogeneous depending on the gamer's corresponding engagement level. Second, it is important to consider the effect of the dynamic fluctuation of perceived difficulty on play-

Table 3.7: Alternative Variable Measurements

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−1.6535	(1.2447)	−14.5826***	(1.5652)	−1.4344	(0.9894)
Soft currency purchase	3.4589***	(1.0189)	7.2360***	(1.3074)	2.4452**	(1.0513)
In-app purchases	−0.8660	(1.6123)	−9.6317***	(1.0325)	−0.3050	(1.0747)
Puzzle difficulty	4.6273***	(1.2415)	1.8630	(1.4027)	10.4750**	(4.6295)
Frequency	0.2053	(1.0280)	5.4602***	(0.9967)	−1.8421**	(0.9299)
Latest intensity	−1.9478	(2.5042)	0.2390	(2.7547)	0.7258	(1.7197)
Historical intensity	0.1399	(2.2527)	0.2708	(1.1216)	2.2021	(1.9469)
Irregularity	−5.0990***	(1.5478)	−0.8045	(1.0002)	−8.8352***	(1.1584)
Weekend	−0.0120	(1.0130)	−0.0470	(1.0336)	−0.0026	(1.0446)
Work hour	−0.1168	(1.6490)	−0.6004	(1.0780)	0.0808	(1.0400)
Constant	−2.0980**	(1.0175)	−7.2739***	(1.0756)	0.4659	(0.9058)
θ Dispersion rate	0.5467	(0.3478)	−0.8624	(1.0013)	−1.2806	(0.9055)
Intercept in the logit probability of playing	−1.4983**	(0.6562)	1.5116*	(0.6638)	1.5286***	(1.1162)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	9.6262***	(1.0622)	9.2228***	(2.9285)	9.8917***	(1.3085)
Fluctuation of perceived difficulty^2	−10.6745***	(1.0015)	−4.6913***	(1.0483)	−6.6579***	(1.0905)
Reward ads	9.6554***	(2.8693)	1.7133	(6.8724)	8.7272*	(4.8536)
Perceived difficulty	3.4112***	(1.0108)	−4.1857*	(2.5173)	3.9483***	(1.0287)
Perceived achievement^{-1}	−4.0676***	(1.0673)	−22.9254***	(3.0294)	−11.1098***	(1.2774)
Cumulative playing time	14.1735***	(3.1623)	5.6837***	(1.0209)	1.5737	(1.0784)
Perceived difficulty * Reward ads	3.5496***	(1.0001)	3.3190**	(1.4320)	5.8065***	(1.0012)
Perceived achievement^{-1} * Reward ads	−3.6504***	(1.0002)	−5.0361***	(1.0493)	−3.8818***	(1.0029)
Thresholds						
State 1			2.9313***	(1.0548)	2.9365**	(2.4949)
State 2	−10.6971***	(1.1036)			16.8033***	(1.0002)
State 3	−10.2710***	(1.5657)	−9.5393	(0.9878)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20249.87			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

ers' engagement evolution because a higher fluctuation of perceived difficulty increases the probability that a player moves to a higher engagement state. Third, I extend the intrinsic motivation theory to find that reward ads can not only serve as a monetization tool but also can help keep players engaged. Furthermore, I examine the interaction effect between reward ads and players' feeling of competence, and I find that reward ad is more helpful in engaging players when players' perceived difficulty or perceived achievement is higher. Finally, after identifying players' latent engagement, I further show that players who are in a low engagement state are more likely to watch reward ads, and players across all three states consume more ads while in need of coins or when the puzzle becomes difficult.

The findings also offer various important managerial and practical implications. Players' engagement is of focal importance to game publishers. I show the heterogeneous effects of different intrinsic motivation factors on players' engagement evolution. Such knowledge can help game publishers build personalized engagement-improving strategies based on players' corresponding engagement levels. Specifically, I find that players in high, medium, and low engagement states respond differently to different motivation factors and the findings can provide guidance to game publishers in the game-design phase. For example, I find that a feeling of achievement has the largest effects on players in a medium engagement state, while a higher perceived difficulty discourages players in a medium state from engaging in games. Therefore, game publishers can consider lowering the perceived difficulty to generate a higher sense of achievement for players who are identified to be in a medium engagement level. In addition, I find that the perceived difficulty fluctuation is positively associated with players' engagement transitions. Therefore, mobile game app developers can consider detecting and dynamically changing players' perceived puzzle difficulty to meet players' needs for diversity and fluctuation. Furthermore, I investigate the relationship between reward ads and players' engagement state. Reward ad, as a relatively new ad format, can potentially create a win-win-win situation for players, game publishers, and ad providers. I find that players who are in a low engagement state are more likely to watch reward ads. Game publishers can leverage such findings to encourage more ad views among players in a low engagement state. Finally, the findings can help improve game designs, which in turn, would benefit players and third-party ad providers as Ill. Players can have an improved in-game experience if publishers take players' heterogeneity into consideration. For advertisers, the findings on reward ads provide important practical implications regarding advertisement targeting strategies and advertisement budget allocation.

There are various ways this paper's findings can be further extended by future research. First, although this paper does not examine in-game purchases as an outcome because the in-app purchase rates are low in my context, future studies can further explore the interplay between game-play decisions, reward ad watching behaviors, and in-game purchase activities. Second, my research context is a single-player mobile game, and I examine the intrinsic motivation factors. Future research can examine different game contexts, such as multi-player mobile games, and investigate the interaction between intrinsic, extrinsic, and social motivation factors. It may be possible that different players may be driven by different types of motivations in games that incorporate players' social and extrinsic motivations. Therefore, I admit that the generalizability of my findings may be limited to single-player mobile games where players are mainly driven by intrinsic motivation factors. Therefore, future research can complement the findings by examining motivation factors in different game contexts. Third, future research can conduct field experiments to examine possible strategies that can increase ad impressions or increase players' engagement levels. For example, the amount of reward that players received from watching reward ads remained fixed in my setting, but a field experiment that changes rewards from reward ad watching can deepen our understanding of reward ads and their relationship with players' engagement.

Chapter 4

NETWORK EVOLUTION IN SOCIAL TRADING PLATFORMS

4.1 Introduction

Online engagement among individual investors has grown significantly in recent years. The recent Reddit hype (Pedersen 2022) vividly shows that social media plays an important role in financial markets and in transmitting relevant information to potential investors. Social trading platforms incorporate elements from the worlds of social media *and* online trading and have recently garnered tremendous attention in both research and practice (Ammann and Schaub 2021, Apesteguia et al. 2020, Yang et al. 2021). Social trading is a novel form of investing that allows retail investors to observe the trading behavior of other investors and to automatically follow their investment strategies through so-called “copy trading” or “mirror trading” (Apesteguia et al. 2020). An autocopy service (mirror trading) enables novice investors (followers) to link their trading accounts to those of expert investors (leaders) and thereby delegate their trading activities (Doering et al. 2015). Experienced investors are able to earn additional income by sharing their trading knowledge with a large group of followers. Large social trading platforms, such as eToro, Zulutrade, and FX Junction, have gained popularity, as evidenced by the growing pool of investors on such platforms. For example, in 2019, eToro was operating in 140 countries with over 10 million users.[1]

Social trading platforms offer several unique features. First, social trading platforms offer a very transparent information flow, as (potential) followers are able to see the details of the transactions completed by other investors and track their gains and losses in real time. Second, these platforms allow for straightforward and transparent communication

[1] `https://www.coindesk.com/company/etoro`, last accessed Jan. 8, 2021.

among investors. Investors can share their opinions, publish posts, and leave comments in a news feed that is publicly available to all users. Third, different from mutual fund managers, most participants in social trading platforms are individual traders who lack institutional endorsement.

Social trading platforms also require a new perspective on considerations of the evolution of social networks, as their network structure follows a different dynamic from that of traditional social networks such as Facebook and Twitter, which have been studied extensively (Li et al. 2017, Kim et al. 2018). In the social networks on social trading platforms, the information flow among users is directly tied to cash flows because of the copy trading feature. Individual investors may become leaders who share their trading strategies or may become followers who copy the trading strategies of leaders. Platforms typically share some of their revenue with leaders. As a result of this monetary incentive, in contrast with other traditional social networks such as Facebook or Twitter, link dissolutions in social trading networks are more frequent. A link connecting two individuals on a traditional social network is commonly characterized by stability and longevity. The link connecting two individuals on a social trading network is short-lived and volatile (Pelster 2017). Thus, not only the process of link formation but also that of link dissolution is crucial. Considering the increasing spread of these networks and their economic implications, an extensive understanding of the evolution of these networks is important. However, the evolution of networks with a frequent dissolution of links has not yet been studied in detail. A large number of studies have focused on the preformation process, i.e., how a social network is formed, but none have analyzed the postformation process. My study fills this void.

I study how the directed leader-follower networks on the largest social trading platform evolve over time. Building on the theory of soft information and hard information (Liberti and Petersen 2019), I investigate the determinants of link formation and link dissolution. Social trading platforms provide a transparent environment that releases two types of information with which potential followers can evaluate leaders: their *trading activities* (financial performance, i.e., hard information), and their *social activities* (social communication, i.e., soft information). The combination of these data is typically difficult to obtain. For example, in traditional mutual funds, researchers can observe the financial performance of a mutual fund manager but typically lack soft communication information. While some mutual fund managers may have social media channels on, for example, YouTube (e.g., Cathie Wood), this is not the case for all managers. In addition, the mutual fund industry allows for private communication between managers and investors that is unobservable to other investors, which may affect investment decisions. In contrast, social trading platforms do not allow for private communication between leaders and followers.[2] The access to various information on social trading platforms allows us to examine the role of soft and hard information in this innovative form of delegated investment management.

I place a particular emphasis on social network features and study the impact of social communication on link formation and dissolution. While prior studies document that investors chase past financial performance, the role of social communication is not clear. Financial performance signals a trader's trading ability. The platform summarizes such information in a highly transparent manner, and it does not allow users to modify or manipulate the data, making them trustworthy. Social communication provides an additional channel for leaders to convince potential followers of their superior investment skill and thus to follow their investment strategies. Given that the primary goal of investors is to make money, followers may focus mainly on financial performance, which provides an objective measure of investment skill, instead of non-monetary soft information—in particular, since, in contrast to the objective features of financial performance, the textual information in communication is more complex to interpret and evaluate. It requires more time for followers to read through text messages and filter out the irrelevant information. The limited attention of followers may make social communication less effective. In addition, it is also questionable how reliable social communication is, given that individually disclosed information is not screened by the platform. Hence, social communication may not be as trustworthy as

[2]Except in rare cases where leaders and followers may know each other in real life.

financial performance, and the role of social communication is unclear.

In this study, I build a separable temporal exponential random graph model (STERGM) to disentangle the reasons why a follower follows *and* the reasons why she or he unfollows a leader in the social trading context. To capture unobserved heterogeneity and address potential endogeneity concerns, I incorporate Chamberlain correlated random effects (Chamberlain 1980) into the STERGM. I find that financial performance, social communication, and demographic characteristics are important determinants of link formation. However, once a link is formed, followers mainly focus on financial performance and social communication (instead of demographic characteristics) to decide whom to unfollow. I also find that the impact of these factors is asymmetric in the link formation and dissolution processes. Different types of social communication, such as posts and comments, have different implications for link formation and dissolution. Both the quality and the quantity of a leader's posts increase the follower's probability of forming a new link and of maintaining an existing link. Followers rely on "peer reviews": Leaders who receive more positive comments are more likely to attract new followers and keep existing ones. Followers are less likely to form new links or sustain existing links if the leaders receive more negative comments. Moreover, the impacts of negative and positive comments are asymmetric; negative comments have a larger impact than positive comments in both link formation and link dissolution. Overall, I find that social communication plays an important role in leaders' ability to convince potential followers to follow their trading strategies and existing followers to sustain their links.

My work makes several contributions to the extant literature. First, this study is the first to model leader-follower network evolution on social trading platforms. Different from those in traditional social networks, relations between investors on social trading platforms involve a monetary dimension, and therefore, social trading features frequent link formation *and* dissolution as investors adjust their investment strategies. My findings enrich the literature on the determinants of social networks by providing empirical evidence of the evolution process of an innovative network structure. Second, the study contributes to the literature on financial advice. Recent developments in fintech have made it easier and more convenient for investors to share their trading knowledge and turn to other investors for advice. With the increasing importance of social interactions on financial markets, the results contribute to this stream of literature by showing that social communication, especially from leaders, can generate economic impacts (i.e., leaders can attract or maintain more followers to earn higher compensation). Third, the study contributes to the literature on individual investor behavior. While financial performance is an important signal of traders' trading skills, I find that followers also rely on communication when evaluating peer traders. Fourth, I contribute to the literature on hard and soft information. I find evidence that both hard information (i.e., financial performance) and soft information (i.e., social communication) play important roles in the link formation and dissolution processes on social trading platforms. Finally, from a methodological perspective, I incorporate Chamberlain random effects into the STERGM to alleviate concerns about confounding effects from individual-level unobserved heterogeneity in the network analysis.

My paper has important managerial implications. While social trading has some features that are comparable to mutual funds in the sense of "delegated portfolio management" (Doering et al. 2015)[3], the extreme flexibility of followers in dissolving links and thereby terminating their relationship instantaneously brings about large income uncertainty for the leader. Thus, for a leader, a thorough understanding of network evolution and its determinants is crucial. In this context, social communication can mitigate information asymmetries and help to build trust. Thus, the results on the impact of social communication can provide important guidance for leaders on when and how to communicate with followers. Second, the findings provide implications for the providers of social trading platforms. For their business

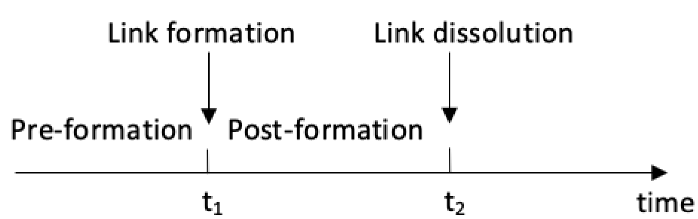
[3]In particular, investors who invest in mutual funds entrust their money to a third party who then makes specific investment decisions for them. This is the same in copy trading on social trading platforms: Investors entrust their money to a leader, and the leader makes specific investment decisions for them. Due to these similarities, the Markets in Financial Instruments Directive (MiFID) characterizes social trading as portfolio management (see, `https://financefeeds.com/mifid-ii-entering-age-completely-self-directed-traders-final-nail-goes-copy-trading-coffin/`).

model to work, platforms need to ensure that both leaders and followers are satisfied with the services provided and thus need to provide a positive investment experience. Third, as most recently demonstrated by the GameStop frenzy, vocal leaders on social media may exert a significant influence on financial markets (see, e.g., Pedersen 2022). Thus, a better understanding of the evolution of social networks with an investment focus is important for regulators.

4.2 Model

Due to the nature of social trading networks, particularly the frequent link formation and dissolution in such networks, it is important to study both link formation and dissolution. Figure 4.1 illustrates a typical link formation and dissolution process on social trading platforms. A link between a follower and a leader is formed at t_1 and dissolved at t_2. Following a leader is equivalent to (automatically) copying the trading strategy of the leader. A link is formed when a follower follows a leader; the link is dissolved when the follower stops following the leader. The network is constructed through leader-follower links. I use the word "follower" for consistency with prior literature (Ammann and Schaub 2021, Yang et al. 2021).

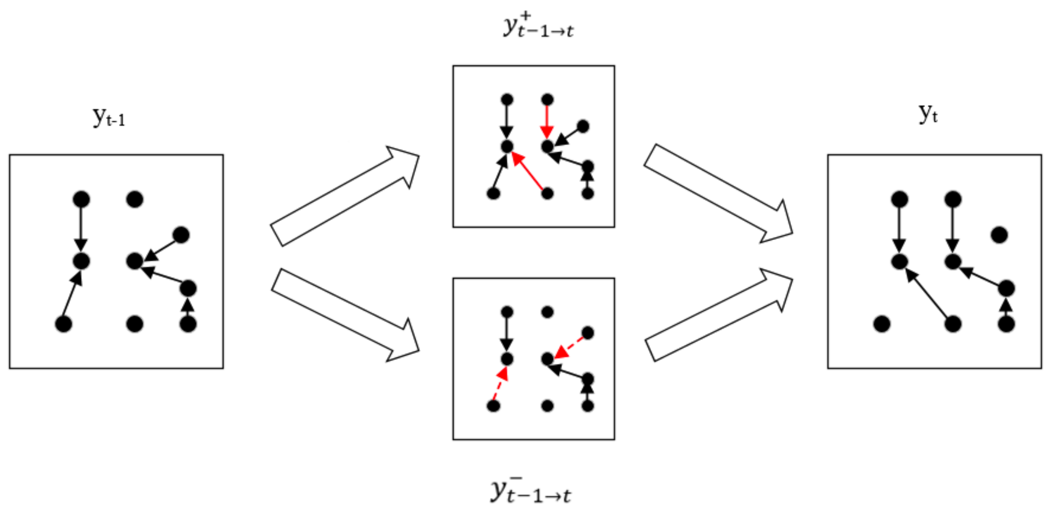
Figure 4.1: An Illustration of Link Formation and Dissolution



4.2.1 Separable temporal exponential random graph model

I use extensions of the exponential random graph model (ERGM) (Snijders et al. 2006, Robins et al. 2007) to model the network evolution. ERGMs represent a general class of models based on exponential family theory that can be used to specify the probability distribution underlying a set of random graphs or networks (Robins et al. 2007, Snijders et al. 2006) and are widely used for network analyses in the field of information systems (Yan et al. 2015, Hwang et al. 2022). The aim of the ERGM is to identify the factors that affect link formation in a network by comparing the probability of the realized network structure with all alternative network configurations. However, the conventional ERGM neither accounts for the intertemporal dependence in longitudinally observed networks nor models the link dissolution process. In this study, I adopt the STERGM (Krivitsky and Handcock 2014), an extension of the ERGM, to model the network dynamics that allows us to capture both intertemporal dependence and the link dissolution process. I consider dynamic leader-follower networks with a total of T time periods. At time t, suppose that there are N_t nodes, and let Y_t be an $N_t \times N_t$ adjacency matrix for a random network. $y_{ijt} = 1$ indicates a link between nodes i and j at time t, and $y_{ijt} = 0$ indicates that there is no link between these nodes at time t. I define $\mathcal{Y}_t$ as the set of all possible networks among the nodes and y_t as a realized network for $y_t \in \mathcal{Y}_t$ at time t. Figure 4.2 provides an illustration that visualizes directed network changes from time $t-1$ to t. I show the realized network at times $t-1$ and t, denoted as y_{t-1} and y_t, respectively. I define two networks to track the evolution of the network from time $t-1$ to t: the *formation network* y^+ and the *dissolution network* y^-. $y^+_{t-1\to t}$ is defined as network y_{t-1} plus the links established from time $t-1$ to t. Similarly, $y^-_{t-1\to t}$ is defined as network y_{t-1} minus the links dissolved from time $t-1$ to t. In my illustration, two new links are added (denoted by red solid arrows), and two existing links are removed (denoted by red dashed arrows). Thus, I are able to track the network evolution in terms of links from time $t-1$ to t. Although I observe only networks y_{t-1} and y_t, I can recover $y^+_{t-1\to t}$ and $y^-_{t-1\to t}$ since $y^+_{t-1\to t} = y_{t-1} \cup y_t$ and $y^-_{t-1\to t} = y_{t-1} \cap y_t$. Appendix B.1 gives a detailed description of how I track the network evolution and construct y^+ and y^- in each period.

Figure 4.2: A Visualization of Network Changes from Time $t-1$ to t



Mathematically, the formation process is modeled as

$$\mathrm{P}(Y^+_{t-1\to t} = y^+_{t-1\to t}|Y_{t-1} = y_{t-1};\theta^+) = \frac{e^{(\theta^{+\prime} g^+(y^+_{t-1\to t}, X_{t-1}))}}{\kappa(\theta^+, X_{t-1}, \mathcal{Y}^+(y_{t-1}))}, \tag{4.1}$$

and the dissolution process is modeled as

$$\mathrm{P}(Y^-_{t-1\to t} = y^-_{t-1\to t}|Y_{t-1} = y_{t-1};\theta^-) = \frac{e^{(\theta^{-\prime} g^-(y^-_{t-1\to t}, X_{t-1}))}}{\kappa(\theta^-, X_{t-1}, \mathcal{Y}^-(y_{t-1}))}, \tag{4.2}$$

where $g^+(y^+_{t-1\to t}, X_{t-1})$ $(g^-(y^-_{t-1\to t}, X_{t-1}))$ is the vector of model covariates for formation network $y^+_{t-1\to t}$ (dissolution network $y^-_{t-1\to t}$), and θ^+ (θ^-) is the vector of coefficients for network $y^+_{t-1\to t}$ $(y^-_{t-1\to t})$. The denominators in Equations (4.1) and (4.2) are normalizing factors that represent the sum of the numerator over all possible networks to ensure that the probability of observing the realized formation (dissolution) network is between 0 and 1. Mathematically, the factor is defined as follows:

$$\kappa(\theta^+, X_{t-1}, \mathcal{Y}^+(y_{t-1})) = \sum_{z^+ \in \mathcal{Y}^+(y_{t-1})} e^{(\theta^{+\prime} g^+(z^+, X_{t-1}))} \tag{4.3}$$

and

$$\kappa(\theta^-, X_{t-1}, \mathcal{Y}^-(y_{t-1})) = \sum_{z^- \in \mathcal{Y}^-(y_{t-1})} e^{(\theta^{-\prime} g^-(z^-, X_{t-1}))}, \tag{4.4}$$

where z^+ (z^-) denotes a possible formation (dissolution) network from time $t-1$ to t.

4.2.2 Identification

A dynamic network analysis of thousands of nodes requires significant computing resources and is computationally intractable (Yan et al. 2015). Thus, I adopt a degenerate statistical model to estimate the coefficients in the link formation and dissolution processes, similar to maximum pseudolikelihood estimation (Strauss and Ikeda 1990).

A common issue in network analysis is endogeneity. First, I use lagged independent variables (i.e., from period $t-1$) to mitigate potential reverse causality. Second, in this context, the information provided on the platform is rich and highly transparent. I observe the information that is observed by followers on the platform, which may affect followers' link formation and dissolution decisions. I have access to the complete transactions history and social communications of each trader, and rich demographic information; the platform does not allow for a private chat channel. I construct various covariates, including follower characteristics, leader characteristics, homophily, and network structure, as elaborated in a later section. However, some determinants that explain link formation and dissolution may still be unobserved, at least to researchers. For example, when followers make their decisions, their investment goals on the platform and their intrinsic trust in others might affect their link formation and dissolution. Hence, to mitigate the concern of omitted variables, I control for follower-specific unobservables (η_i) in the link formation and dissolution model. Specifically, for the link formation process[4], I define

$$y_{ijt} = \begin{cases} 1 & y^*_{ijt} > 0, \text{and} \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

[4]The link dissolution process is defined in the same fashion. In the dissolution process, y_{ijt} is equal to 1 if follower i dissolves the link from leader j in period t.

y_{ijt} is a binary variable that is equal to 1 if follower i forms a link with leader j from period $t-1$ to t, and y^*_{ijt} is the corresponding latent utility. The utility of follower i from forming a link to leader j at time t is defined as follows:

$$y^*_{ijt} = \boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij} + \boldsymbol{\tau} C_i + \eta_i + \epsilon_{ijt}, \tag{4.6}$$

where X_{it-1} is a vector of follower i's time-variant covariates at period $t-1$, W_{jt-1} is a vector of leader j's time-variant covariates at period $t-1$, V_{ij} is a set of dummies indicating whether follower i and leader j share the same demographics (homophily), C_i is a set of follower-specific time-invariant observable controls, and η_i is the follower-specific fixed effect. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\tau}$ are the corresponding vectors of coefficients to be estimated.

A conventional approach to estimating fixed effects is to treat η_i as parameters and use maximum likelihood estimation. However, such estimation is inconsistent when the number of nodes is large and the number of time periods is finite, which is the incidental parameter problem (Neyman and Scott 1948). To correct for the incidental parameter problem, Chamberlain (1980) proposes a correlated random effects model (Wooldridge 2010). I control for follower-specific Chamberlain correlated random effects in the model.[5] Chamberlain (1980) allows for follower-specific unobservables to be correlated with the independent variables. Specifically, I implement Chamberlain correlated random effects following Mundlak (1978). η_i is defined as follows:

$$\eta_i = \psi + \boldsymbol{\xi} \bar{X}_i + a_i, \tag{4.7}$$

where a_i follows a normal distribution with mean zero and variance σ^2_a, ψ is a constant, and $\bar{X}_i$ is the time average of the follower's time-variant observables. $\bar{X}_i = (\Gamma_i)^{-1} \sum_{t=1}^{\Gamma_i} X_{it}$, where Γ_i equals the number of periods that follower i exists on the platform times the number of leaders that follower i can potentially follow in each period.

[5] I also estimate link formation and dissolution using an alternative estimation approach, the conditional logit estimator. The estimation results remain consistent with the results from the main model. Please refer to Appendix B.7 for more details.

In Equation (4.7), a_i is not assumed to depend on X_i, and the model allows for dependence between η_i and X_i by adding $\bar{X}_i$ to the equation. From Equation (4.7), I see that η_i follows a conditional normal distribution, that is, $\eta_i|X_i \sim Normal(\psi + \bar{X}_i\xi, \sigma_a^2)$. Thus, unlike the conventional fixed effects model, the coefficients on the follower-specific time-invariant controls C_i in Equation (4.6) can be identified.

Although no regulatory changes were made to social trading in general during the sample period and the regulation remains intact, link formation and dissolution might be affected by some other events that are time dependent—for example, some policy change on the platform. To mitigate this concern, I include time fixed effects, which allows us to control for time-specific peculiarities. The estimation results remain consistent with the results from the main model. Please refer to Appendix B.2 for more details.

Finally, despite the various leader characteristics included in the model, there might still be some unobservables that affect link formation and dissolution. For example, it is possible for leaders on the platform to advertise themselves via other social media platforms. To mitigate omitted variables on the leader's side, I also include leader-specific Chamberlain correlated random effects in the link formation model. The estimation results remain consistent with the results from the main model. Please refer to Appendix B.3 for more details.

4.3 Data and Variables

In this section, I first introduce the data and then describe how I construct the variables used in the analysis.

4.3.1 Data

I obtained the data from eToro, the largest social trading platform. Similar to other online trading brokerage services, the platform allows its customers to trade stocks and contracts-for-differences (CFDs) on indices, commodities, currency pairs, and crypto assets. For each trade, the broker charges transaction fees as a portion of the bid-ask spread. In addition, the platform incorporates various features that are typical of social media. Specifically, the platform contains a news feed in which investors can disclose their trading activities (*open book trading*) and publish posts. Here, investors can conveniently discuss their trading strategies, like and comment on others' trading activities, and automatically copy the trades of other investors. Investors who have their trades copied receive monetary compensation from the brokerage service in relation to their number of followers, their assets under management, and their investment performance, similar to professional fund managers. Each investor has a public profile page, which shows detailed and transparent information on his/her past trading activities, including financial performance, social activities (e.g., posts, comments, likes, and replies), and number of followers.

The data cover the complete social activity and trading activity histories of all investors in 2016 and 2017. Social activity histories include all posts, comments, replies, and likes together with the exact timestamp of each activity. Trading activity histories include detailed information on each trade. In addition, the data include the dynamics in the leader-follower networks (i.e., when a follower follows or unfollows a leader). In other words, for each link in the network, I know the exact timestamp of the formation and dissolution of the link between the follower and the leader. Finally, the data include each investor's nationality, age, gender, use of a profile image, publication of a biography, trading experience before joining the platform, wealth, income, and desired risk level upon registration.

In my analysis, I consider each investor to be a node in the network. If an investor (follower) follows or autocopies another investor (leader), then this relation is modeled as a directed link between the follower and leader. I set $y_{ijt} = 1$ if a link exists between nodes i and j in period t. I use the data from 2016 to proxy for historical trading performance (e.g., average profit and standard deviation of profit) to guarantee a long-term horizon for followers to evaluate leaders. I examine link formation and dissolution using the leader-follower network in 2017. I define each period at the monthly level. I first illustrate how I sample investors in period 1. I select all leaders who have at least 5 followers (to alleviate the sparsity of the network and exclude some casual investors) in period 1 and stay on the platform in two successive months (i.e., period 1 and period 2), ending with 462 leaders.[6] I then obtain information on all the followers of these leaders, ending with 13,533 unique followers who exist during these two successive periods. As the large number of nodes can cause computational intractability issues in network analyses, I randomly sample 600 followers out of 13,533, resulting in 1,057 unique investors (because some investors may be both followers and leaders, the total number of investors is less than the summation of leaders and followers). In period 2, as some investors sampled from period 1 may exit the platform and some new investors may join the platform, I first keep those investors who remain on the platform over two successive months (i.e., period 2 and period 3). Then, using the number of investors in period 1 as an anchor, I add investors by randomly sampling from the new investors who join the platform in period 2 and stay on the platform during both period 2 and period 3. I repeat this procedure across all the periods. Table 4.1 summarizes the network statistics: the number of nodes, the number of links, and the density of the network over time.

From Table 4.1, we see that the number of unique nodes varies slightly across different periods because the number of traders who are both a follower and a leader is different in each period. Although the number of nodes does not vary much across periods, the node set does change since new nodes can join and existing nodes can exit. It is a common practice to sample a smaller set of nodes to achieve computational feasibility when estimating network analysis models (Yan et al. 2015, Lee et al. 2016). For example, Lee et al. (2016) study strategic network formation in a location-based social network. Their network analyses were conducted on three city-level subsamples consisting of 336, 129, and 146 users. Yan et al. (2015) examine the driving forces behind patients' social network formation and evolution using a subsample consisting of 1,322 individuals. Although the number of nodes is approximately 1,057 in the network analysis, the leader-follower relations between the nodes are described by a 1,057×1,057-dimensional matrix in each period, and there are 12 periods

[6] Existence over two successive months is the minimum requirement because the formation network (dissolution network) is constructed by tracking the links added (removed) between two successive periods.

Table 4.1: Network Dynamics

Period	Nodes	Links	Density
1	1057	1595	0.0014
2	1053	1588	0.0014
3	1053	1658	0.0015
4	1053	1703	0.0015
5	1053	2025	0.0018
6	1054	2012	0.0018
7	1055	1923	0.0017
8	1055	1857	0.0017
9	1056	1832	0.0016
10	1056	1809	0.0016
11	1057	1742	0.0016
12	1057	1250	0.0011

in total. The total number of observations in the network formation model is 11,000,219. However, I acknowledge that this current sample size is still small compared to the large user base on the platform. To further address concerns about the external validity of the findings, I take another random sample and re-estimate the model; the main results are generally consistent. Please refer to Appendix B.4 for more details.

4.3.2 Variables

Based on the theoretical background presented in Section 2.2, I consider four groups of variables that may affect the dynamics of follower-leader networks. I next describe in detail how I construct these variables.

Social communication

An important feature of social trading platforms is that investors are able to conveniently interact with other investors. For example, eToro allows its users to publish posts, comment on posts, and distribute likes. Investors can, for example, publish posts to broadcast their recent achievements, explain their trading strategies, share their financial advice, or simply communicate with others about recent events. Other investors can comment on these posts to voice their opinions, to request additional information, or to ask for clarification regarding comments.[7] As investors sometimes choose to reply to a comment that is made on their original post by leaving another comment, I label these types of comments "replies" to distinguish them from the original comments. All social interactions are shown on the platform news feed and in the investor's public profile, similar to the widely adopted news feeds used on typical social media platforms such as Facebook or Twitter. eToro does not provide its users with the ability to chat privately. Consequently, the news feed is the only way that users can communicate with each other, and all social activities on the platform are public. Examples of different types of social communication (posts, comments, and replies) are presented in Appendix B.5.

I measure investors' social activities using the following variables. For each investor, I use the total number of posts over period t to measure his/her post writing intensity (*post quantity*). Following the literature (Khern-am-nuai et al. 2018, Cao et al. 2011), I use the number of likes that a post receives to measure its quality. Then, I take the average over all posts in period t as a proxy for an investor's *post quality.* Considering prior evidence that sentiments in user-generated content play an important role in agents' decision-making processes (Xu and Chau 2018), I do not simply examine the effect of the number of comments but instead focus on the sentiments expressed in those comments. Due to the international customer base of the platform, comments are posted in different languages. Thus, to conduct a sentiment analysis, I first use a Google cloud translation application programming interface

[7] I do not observe automated comments in the data and are confident that investors manually post these comments. In addition, I do not work under the assumption that all users read all comments. In fact, given the substantial number of comments and the limited attention of investors (Hirshleifer and Teoh 2003), I believe this to be very unlikely. However, if comments are not being read, then the impact on relationships should be zero and statistically insignificant, which would be reflected in the estimation results.

(API) to translate all comments into English.[8] I then remove stop words, perform word stemming, and use a lexicon-based content analysis to perform the sentiment analysis. I implement the Valence Aware Dictionary and sEntiment Reasoner (VADER), specifically attuned to sentiments expressed on social media (Hutto and Gilbert 2014). VADER has been recently applied in finance and trading and performs as Ill as individual human raters at matching ground truth (Hutto and Gilbert 2014). The package enables us to label each comment with positive and negative sentiment scores by calculating the percentage of the text that falls in each category. Then, I average the positive and negative sentiment scores for all comments on an investor's post in period t and obtain the variables *comment positive* and *comment negative*, respectively.[9] Figures 4.3a and 4.3b show the distribution of negative and positive comment scores, where a higher negative score indicates that a comment contains a larger percentage of negative words. Finally, I count the replies provided by investors to their received comments in period t using the variable *reply*.

Financial performance

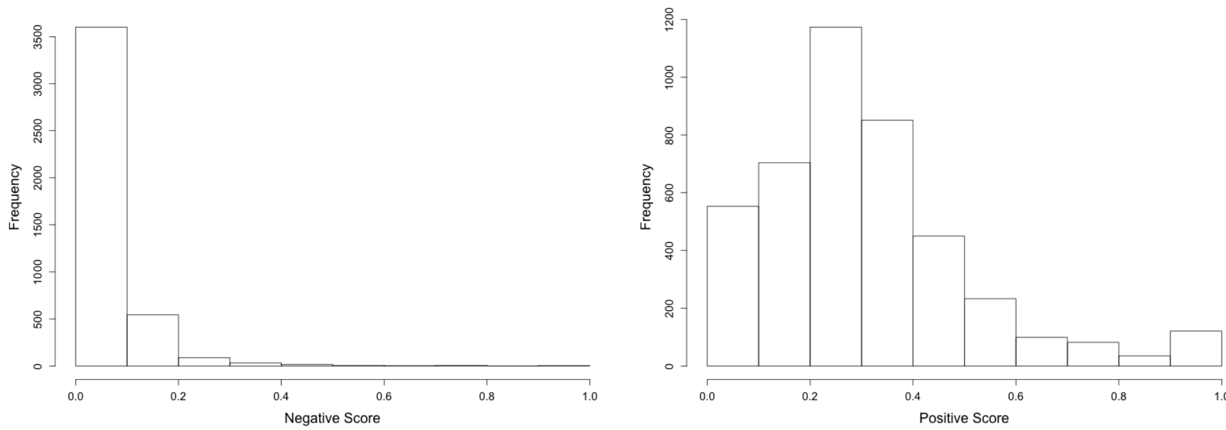
To measure an investor's financial performance, I first calculate their average daily profit over period t. However, statistics from only one period (month) may not reflect the investor's overall performance. Thus, I use their historical average profit until period t to measure their *average profit*. Similarly, I first calculate the standard deviation of daily returns over period t and then take the average of the historical returns until period t as a proxy for investors' volatility (in line with, e.g., Sirri and Tufano 1998, Huang et al. 2007).

I construct the following variables to proxy for investors' trading strategies. *Holding time* measures the duration from opening to closing a particular position, reflecting the extent to which a trader prefers "day trading" versus a buy-and-hold strategy. I account for investors'

[8]This procedure is consistent with the practices of the platform, which provides a "translate" icon for each post and comment that allows users to view all posts and comments in English.

[9]As Tirunillai and Tellis (2012) found that the effect of negative and positive user-generated content is asymmetric, I include both positive and negative comments in the model.

Figure 4.3: Distribution of Comment Sentiment Scores



(a) **(b)**

portfolio features using the Herfindahl-Hirschman index (*HHI*), a measure of diversification based on the sum of squared portfolio weights (Dorn et al. 2008). A smaller HHI indicates a better diversified portfolio. I include a measure for investors' preferences for investing in lottery-type stocks following Kumar (2009), based on the observation that retail investors are attracted to lottery stocks (i.e., stocks with positively skewed returns) and that this attraction to lottery stocks can increase as a result of social interactions even if investors do not have inherent preferences for skewness (Han et al. 2022). In this vein, Bali et al. (2021) show that social interactions aggravate the lottery anomaly. I define the variable *lottery preference* as the fraction of trades that a given investor executes in lottery-type stocks to all trades by the investor.

Demographics

I include some variables to control for homophily based on demographic characteristics. First, I use a dummy variable that takes a value of one if investors come from the same country and zero otherwise (*Nationality*). In a similar fashion, I control for homophily along the investor gender dimension (*Gender*). I also construct a dummy variable to indicate whether investors are in the same age range (*Age*). Social investors may also generate trust by having a detailed profile page that includes their image and/or a biography (Wohlgemuth et al. 2016). Consequently, I include two dummy variables for leaders, *Image* and *Bio*, to denote whether a profile picture[10] or a biography are provided on an investor's profile page. Finally, I incorporate investor characteristics upon registration, including trading experience in years before joining the platform (*Experience*), total wealth in dollars (*wealth*), annual income in dollars (*Income*), and the reported desired risk level (*Risk*), to capture observed heterogeneity.

Network structure

In addition to node characteristics and dyadic covariates, the network structure may affect network evolution through reciprocity and transitivity (Wasserman et al. 1994, Holland and Leinhardt 1971). In the data, the number of mutual links (i.e., $i \to j$ and $j \to i$) is zero, and thus, there is no reciprocity effect, which is intuitive, as it is unlikely that leaders will follow their followers' trading strategies in a social trading network. Therefore, I do not consider reciprocity in my study. However, I incorporate a triadic term to capture potential transitivity. When links $i \to j$ and $j \to k$ exist, the likelihood that a new link $i \to k$ will be formed increases. While triadic effects represent the local hierarchy within the network, I also incorporate the global hierarchy among all nodes within the network—node-level in-degree-related popularity and out-degree-related activity (Hunter et al. 2008). It is possible

[10]The dataset does not contain more detailed information on the picture, for example, whether it is a symbolic image or a photo showing a real person. Due to the anonymous nature of the data, I are not able to collect this information and merge it with the dataset. I study a random sample of the profile pictures of eToro users to analyze how many fantasy pictures, on average, are used. The analysis of slightly over 500 randomly selected profile pages shows that approximately 80% of them contain photos showing a real person (that is not a Ill-known celebrity).

Table 4.2: Data Description and Statistics

Variable	Description	Mean	Std. Dev.	Min.	Max.
Social communication					
Post quantity	The number of posts made by investors	23.77	146.39	0	5406
Post quality	The number of likes that an investor's posts receive	4.28	9.59	0	236
Comment positive	The percentage of positive words in the comments that an investor receives	0.31	0.21	0	1
Comment negative	The percentage of negative words in the comments that an investor receives	0.06	0.08	0	1
Reply	The number of replies to comments	20.57	62.49	1	1551
Financial performance					
Average profit	The average profit	0.0008	0.04	-0.55	0.97
Std. dev. profit	The standard deviation of profit	0.05	0.11	0	3.11
Holding time	The duration of a particular position from opening to closing	14.00	38.23	0	749.21
HHI	Herfindahl-Hirschman index of portfolio diversification	0.19	0.31	0	1
Lottery preference	Fraction of trades in lottery-type stocks	0.02	0.06	0	1
Demographics					
Gender	Dummy =1 if both investors are females or both are males and 0 otherwise	0.83	0.38	0	1
Age	Dummy =1 if both investors are in the same age range and 0 otherwise	0.29	0.45	0	1
Nationality	Dummy =1 if both investors are from the same country and 0 otherwise	0.07	0.26	0	1
Image	Dummy =1 if the investor uploads a profile picture and 0 otherwise	0.48	0.50	0	1
Bio	Dummy =1 if the investor uploads a biography to his/her profile and 0 otherwise	0.22	0.41	0	1
Experience	Trading experience before joining the platform at the time of registration	1.53	1.07	0	3
wealth	Reported wealth at the time of registration	105,541	251,475.2	10,000	2,000,000
Income	Reported annual income at the time of registration	113,644	209,495.7	10,000	2,000,000
Risk	Reported risk at the time of registration	24.35	15.41	3	48
Network structure					
In-degree popularity	The number of incoming ties in an investor's social network	1.70	10.68	0	371
Out-degree activity	The number of outgoing ties initiated by an investor in his/her social network	1.70	2.18	0	43
Transitivity	The number of triadic closures for each node	0.21	1.42	0	42

that the effects of popularity and activity are different for a leader and a follower. Therefore, I distinguish between leader and follower nodes. For a leader node, I measure how in-degree-related popularity and out-degree-related activity affect the propensity of others forming a link with that leader. A positive in-degree-related popularity effect indicates that others find it attractive to follow a leader with more in-ties (in-links). Similarly, a positive effect from out-degree-related activity means that others find it attractive to follow a leader with more out-ties (out-links). For follower nodes, I measure the propensity of the follower to follow someone else (form a link). Specifically, a positive estimate for in-degree-related popularity (out-degree-related activity) implies that followers with more in-ties (out-ties) are more likely to follow leaders. For all the above variables, I provide brief definitions and summary statistics in Table 4.2.

4.4 Results

I apply the STERGM with Chamberlain correlated random effects to investigate link formation and dissolution in social trading. I study the determinants presented in Section 4.3.2

and estimate the coefficients that best fit the model using a maximum likelihood procedure. I provide several additional robustness checks in Section 4.5 and in the Appendix B.

Table 4.3 summarizes the main estimation results separately for link formation and dissolution. A positive coefficient in the formation model indicates an increased probability of forming a new link, whereas a positive coefficient in the dissolution model indicates a positive effect on link duration (i.e., the link is less likely to dissolve). I distinguish between the variables for a leader's account and those for a follower's account since these actors play different roles in the network evolution. Model 1 is the baseline model without social communication. Model 2 is the full model that includes all explanatory variables. Hereafter, I discuss the main results based on Model 2, which includes social communication, financial performance, demographic characteristics, and network structure. Overall, I find evidence that hard information (financial performance) and soft information (social communication and demographic characteristics) play different roles in the link formation and dissolution processes. I discuss the findings for each group of variables in detail.

4.4.1 Social communication

Table 4.3 shows positive coefficients for a leader's post quantity for both link formation (coefficient of 0.3162, $p < 0.01$) and link dissolution (coefficient of 0.1163, $p < 0.01$). Similarly, the coefficients for a leader's post quality during link formation (coefficient of 0.5097, $p < 0.01$) and link dissolution (coefficient of 0.0439, $p < 0.05$) are also positive, indicating that the propensity to form a link and maintaining existing links increases as leaders publish a larger number of posts of higher quality.[11] This observation is consistent with the notion that high-quality posts provide useful information and increase the transparency of a leader's investment strategy, which in turn increases trust in the leader, attracts more incoming links,

[11]Also of particular interest is the interplay between post quality and post quantity. A robustness check including a quantity-quality interaction term indicates that post quality has a positive moderating effect on post quantity in the link formation process. The interaction term is not significant in the link dissolution process. Please refer to Appendix B.6 for more details.

Table 4.3: Estimation Results

	Model 1				Model 2			
Variable	Formation		Dissolution		Formation		Dissolution	
Social communication								
Leader's post quantity					0.3162***	(0.026 8)	0.1163***	(0.028 3)
Leader's post quality					0.5097***	(0.025 0)	0.0439**	(0.020 6)
Leader's number of replies					0.0238	(0.027 7)	−0.0105	(0.031 0)
Leader's comment received positive score					1.1122***	(0.139 7)	0.5754***	(0.145 8)
Leader's comment received negative score					−4.7688***	(0.752 5)	−2.1181***	(0.471 5)
Follower's post quantity					0.0655	(0.047 4)	−0.5311***	(0.053 6)
Follower's post quality					−0.3348***	(0.079 7)	−0.0044	(0.074 9)
Financial performance								
Leader's average profit	0.0536***	(0.011 4)	0.0383**	(0.017 5)	0.0903***	(0.013 6)	0.0707***	(0.018 3)
Leader's std. dev. profit	−0.3113	(0.476 0)	−1.3945*	(0.730 6)	−2.1610***	(0.556 7)	−2.6485***	(0.761 5)
Follower's average profit	0.0031	(0.011 6)	0.0416***	(0.012 4)	0.0012	(0.011 4)	0.0440***	(0.012 0)
Follower's std. dev. profit	−0.8744*	(0.514 4)	−8.8636***	(0.582 7)	−0.6665	(0.508 6)	−8.1553***	(0.558 2)
Leader's average holding time	0.1740***	(0.041 6)	−0.0106	(0.064 4)	0.3465***	(0.044 6)	0.0652	(0.066 1)
Leader's lottery preference	0.5191	(0.322 0)	1.1677***	(0.393 7)	0.2019	(0.361 9)	1.0743***	(0.394 8)
Leader's HHI	−0.8337***	(0.091 3)	−0.0956	(0.091 7)	−0.6094***	(0.090 5)	−0.1435	(0.092 9)
Demographics								
Nationality	0.8856***	(0.069 1)	0.3435***	(0.079 0)	0.7456***	(0.070 4)	0.3408***	(0.079 7)
Age	0.1028**	(0.050 8)	0.0456	(0.053 9)	0.1221**	(0.050 5)	0.0519	(0.054 3)
Homophily (male)	0.9723***	(0.135 8)	−0.1302	(0.115 2)	0.9862***	(0.135 9)	−0.1510	(0.116 4)
Homophily (female)	−0.8507**	(0.416 4)	0.2365	(0.314 5)	−0.8602**	(0.416 6)	0.2906	(0.317 7)
Image	2.9703***	(0.417 1)	0.6320	(0.475 9)	2.4784***	(0.423 8)	0.7280	(0.465 2)
Bio	3.6225***	(0.177 8)	−0.1228	(0.187 0)	2.9204***	(0.186 9)	−0.1499	(0.188 6)
Experience	−0.0511	(0.037 2)	0.1489***	(0.042 2)	−0.0444	(0.037 1)	0.1460***	(0.040 4)
Wealth	0.0217	(0.032 8)	0.0795**	(0.036 4)	0.0243	(0.032 4)	0.0618*	(0.034 7)
Income	−0.0027	(0.039 6)	0.0205	(0.042 9)	−0.0090	(0.039 3)	0.0239	(0.040 9)
Risk	−0.0922*	(0.049 2)	−0.0624	(0.055 3)	−0.0839*	(0.048 8)	−0.0397	(0.052 8)
Network structure								
Leader's popularity	0.0157***	(0.000 3)	0.0021***	(0.000 3)	0.0082***	(0.000 4)	0.0014***	(0.000 4)
Leader's activity	−0.1376***	(0.017 2)	0.0312***	(0.011 7)	−0.0353***	(0.012 3)	0.0380***	(0.011 9)
Follower's popularity	−0.0726***	(0.013 0)	−0.0201***	(0.006 9)	−0.0546***	(0.012 7)	−0.0089	(0.006 5)
Follower's activity	0.0009	(0.007 0)	−0.0838***	(0.008 8)	0.0026	(0.007 1)	−0.0769***	(0.008 8)
Transitivity	0.0938***	(0.026 7)	−0.0057	(0.066 6)	0.0820***	(0.025 6)	−0.0218	(0.065 6)
Constant	−19.2557***	(5.212 6)	0.8493	(0.659 7)	−15.8578***	(5.663 5)	2.0938***	(0.644 6)
Log Likelihood	-14,418.19		-9,345.29		-13,620.36		-9,152.26	
Observations	11,000,219		19,744		11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of the posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

and helps maintain existing links. Overall, this notion is consistent with the literature that argues that communication can increase trustworthiness and trust (Kaiser and Berger 2021).

The coefficients on positive comments for link formation and dissolution are significantly positive (coefficients of 1.1122 with $p < 0.01$ and 0.5754 with $p < 0.01$, respectively), indicating a higher probability that a follower will form or maintain a link if the leader receives more positive comments. Similarly, negative and significant estimates for negative comments indicate that negative comments are associated with a lower probability of link formation or a shorter link duration (coefficients of -4.7688 with $p < 0.01$ and -2.1181 with $p < 0.01$, respectively). The different magnitudes of the coefficients further indicate that the impacts of negative and positive comments are asymmetric and that negative comments are particularly relevant in online contexts, consistent with previous evidence from social media (Xu and Chau 2018).[12]

Comparing the estimates for the effect of social communication in link formation and dissolution, I find that social communication plays an important role in both processes. Thus, communication on social trading platforms seems to have an economic impact on leaders given that the links in the leader-follower network are tied to cash flow and directly affect the compensation that a leader receives from the platform. Communication, a type of soft information, not only helps leaders attract new links but also helps to maintain existing links. The coefficients indicate that the effect of communication is stronger for link formation than for link dissolution, which may be explained by the fact that posts can affect link formation through an additional channel (i.e., the attention channel) that is less relevant for dissolution (see, e.g., Barber and Odean 2008, for a similar argument in the financial markets context).

Regarding the social activities of followers, I find that the probability of establishing new links decreases as the quality of posts increases (coefficient of -0.3348, $p < 0.1$). Intuitively,

[12]The t-test of the difference in coefficients between negative comments and positive comments is 7.85 in the link formation model and 5.57 in the link dissolution model, indicating that they are significantly different.

followers with higher-quality posts may have greater financial knowledge and expertise and consequently may be more likely to trade by themselves instead of following others.

4.4.2 Financial performance

Turning to financial performance, I find that a leader's average profit tends to attract followers for both link formation and link duration (coefficients of 0.0903 with $p < 0.01$ and 0.0707 with $p < 0.01$, respectively). In addition, greater volatility in the leader's financial performance is negatively associated with link formation and link duration (coefficients of -2.1610 with $p < 0.01$ and -2.6485 with $p < 0.01$, respectively). Overall, these observations are in line with prominent findings from the mutual fund flow literature that investors chase past performance (Barber et al. 2016) and with previous evidence on social trading (Doering et al. 2015).

Next, I focus on the performance of followers. I find that they are insensitive to their past profit on the platform when choosing a new link, as the coefficients on a follower's average profit and the standard deviation of profit are insignificant in the link formation model (coefficients of 0.0012, $p = 0.91$ and -0.6665, $p = 0.19$, respectively). However, once a link is formed, followers tend to be more likely to maintain the link if they have a higher average profit (coefficient of 0.0440, $p < 0.01$) and a lower volatility level (coefficient of -8.1553, $p < 0.01$).

Finally, I turn to the trading strategies of leaders and find that followers prefer leaders who tend to follow diversified buy-and-hold strategies and are more likely to establish links with those leaders. Once links are established, investing in lottery-like stocks is associated with a longer link duration.

4.4.3 Demographics and network structure

With respect to demographic characteristics, I find—most notably—that followers tend to establish and maintain links with leaders who have the same nationality (coefficients of 0.7456, $p < 0.01$ and 0.3408, $p < 0.01$), which is consistent with studies on peer-to-peer credit markets (Lin and Viswanathan 2016). These effects may be driven by language barriers or cultural differences. I also find evidence in support of age homophily (coefficient of 0.1221, $p < 0.05$) and gender homophily among male investors (coefficient of 0.9862, $p < 0.01$) during link formation. However, female followers are more likely to form a link with male leaders (coefficient of -0.8602, $p < 0.05$). In line with previous findings from the literature (Wohlgemuth et al. 2016), I observe that the presence of a picture on a leader's profile page and of a biographical description significantly increases the likelihood that followers will form a new link (coefficients of 2.4784, $p < 0.01$ and 2.9204, $p < 0.01$, respectively). The disclosure of a profile picture or a biography may increase the perceived trustworthiness of the leader and therefore the likelihood of new links. Next, I consider a follower's experience, wealth, and risk preference. The results show that followers with a higher risk score are less likely to form a new link (coefficient of -0.0839, $p < 0.1$), whereas followers with a higher experience level or wealth level tend to maintain their existing links (coefficients of 0.1460, $p < 0.01$ and 0.0618, $p < 0.1$, respectively).

Comparing the link formation and dissolution processes, I find strong differences in the impact of the demographic characteristics of leaders. This observation is intuitive in the sense that once followers have considered the demographic characteristics and established a link, there is no need to consider them again, as demographics remain stable over time. Thus, other factors, such as financial performance and social communication, become more relevant.

Finally, I briefly discuss the variables that capture the network structure. A leader's popularity (in-ties) increases his/her propensity to attract followers who form new links and to maintain existing links (coefficients of 0.0082, $p < 0.01$ and 0.0014, $p < 0.01$, respectively), indicating preferential attachment. In contrast, the coefficient on a leader's activity (out-ties) is significantly negative (coefficient of -0.0353, $p < 0.01$), indicating that leaders who follow other investors are less attractive to potential followers. Followers with higher popularity are less likely to form new links (coefficient of -0.0546, $p < 0.01$), whereas those with higher levels of activity are less likely to maintain existing links (coefficient of -0.0769, $p < 0.01$). I also find a significant transitivity effect (coefficient of 0.0820, $p < 0.01$), indicating that the presence of a link from i to j and from j to k increases the likelihood of a direct link between i and k being formed.

4.4.4 Heterogeneous effects across follower age

Prior studies have found that individuals who are younger in age are more likely to blog, visit social network sites, and rely on social media in their decision-making than those who are older in age (Chou et al. 2009). Thus, in this subsection, I further scrutinize the implications related to age. Considering that social trading is a novel way to participate in financial markets that may particularly attract younger individuals, it is natural to ask whether the findings hold across all age groups. I thus examine whether the impact of the social communication and financial performance variables varies across investors by age group.

I split the dataset into two subsamples based on followers' age ranges. The first group includes followers between 18 and 44 years of age, and the second group includes followers who are older than 44 years. I again apply the STERGM with Chamberlain correlated random effects and summarize the estimation results in Table 4.4. Younger followers are rather sensitive to positive and negative comments in the link dissolution process (coefficients of 0.6684, $p < 0.01$ and -2.3591, $p < 0.01$, respectively), whereas the effects of comments are not significant for older followers (coefficients of 0.3637, $p = 0.19$ and -1.5742, $p = 0.11$, respectively). Interestingly, I observe that a leader's post quantity increases the probability of younger followers maintaining existing links (coefficient of 0.1302, $p < 0.01$), while the post quality becomes insignificant. At the same time, However, for older followers, it is a leader's post quality (coefficient of 0.1076, $p < 0.01$) rather than post quantity that increases the probability of maintaining existing links.

4.5 Robustness Tests

In this section, I present a series of robustness checks.

Table 4.4: Estimation Results: Heterogeneous Effects by Follower Age

	Group 1				Group 2			
Variable	Formation		Dissolution		Formation		Dissolution	
Social communication								
Leader's post quantity	0.2975***	(0.0317)	0.1302***	(0.0334)	0.3624***	(0.0501)	0.0770	(0.0537)
Leader's post quality	0.5126***	(0.0296)	0.0217	(0.0241)	0.5041***	(0.0471)	0.1076***	(0.0407)
Leader's number of replies	0.0517	(0.0326)	−0.0220	(0.0364)	−0.0400	(0.0531)	0.0222	(0.0596)
Leader's comment received positive score	1.1623***	(0.1662)	0.6684***	(0.1721)	0.9892***	(0.2590)	0.3637	(0.2780)
Leader's comment received negative score	−4.2946***	(0.8679)	−2.3591***	(0.5416)	−6.1927***	(1.5113)	−1.5742	(0.9766)
Follower's post quantity	0.1476***	(0.0528)	−0.5766***	(0.0599)	−0.1739	(0.1121)	−0.3404***	(0.1270)
Follower's post quality	−0.4948***	(0.0893)	0.0136	(0.0801)	0.2838*	(0.1720)	−0.1798	(0.2439)
Financial performance								
Leader's average profit	0.0914***	(0.0166)	0.0741***	(0.0222)	0.0885***	(0.0238)	0.0603*	(0.0338)
Leader's std. dev. profit	−2.4100***	(0.6752)	−2.9491***	(0.9208)	−1.6267*	(0.9773)	−1.7755	(1.4176)
Follower's average profit	−0.0021	(0.0124)	0.0348***	(0.0132)	0.0662**	(0.0275)	0.0880***	(0.0301)
Follower's std. dev. profit	−0.1756	(0.5337)	−7.8365***	(0.6058)	−5.0396**	(1.9718)	−10.4074***	(1.4808)
Leader's average holding time	0.3505***	(0.0532)	0.0917	(0.0762)	0.3371***	(0.0828)	−0.0311	(0.1390)
Leader's lottery preference	0.1726	(0.4300)	1.4006***	(0.4611)	0.3151	(0.6711)	0.1110	(0.7660)
Leader's HHI	−0.6326***	(0.1070)	−0.0990	(0.1082)	−0.5813***	(0.1696)	−0.2657	(0.1834)
Demographics	Yes				Yes			
Network structure	Yes				Yes			
Log Likelihood	-9,795.51		-6,779.52		-3,762.49		-2,327.82	
Observations	8,015,758		14,476		1,658,708		5,268	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

4.5.1 Two-stage selection model

Since the total number of leaders on the platform is large and followers have limited attention, it is possible that some leaders are more visible than others, which might affect whether potential followers follow any particular leader. Thus, in this robustness check, I develop a two-stage selection model that attempts to accurately model the link formation process. In the first stage, to account for different exposure to leaders' profiles among followers, I model the probability of followers being aware of leader j as

$$Pr(A_{jt} = 1) = \frac{e^{\gamma z_{jt}}}{1 + e^{\gamma z_{jt}}}, \tag{4.8}$$

where z_{jt} denotes the number of followers of leader j in period t. I choose the number of followers in the first step because, first, the finance literature provides substantial evidence that herding is a relevant behavioral trait in financial markets (Devenow and Welch 1996). In the context of social trading, Gemayel and Preda (2018b) show that the scopic regime can increase the extent of herding. Given that the number of followers already reflects financial performance to some degree (i.e., investors with poor performance are less likely to have many followers), I argue that investors will start their filtering decision based on fewer criteria in order to simplify their decision process as much as possible. Second, I refer to the concept of preferential attachment in social networks from the IS literature, and leaders with large follower bases should be more likely to attract additional followers (Neyman and Scott 1948).

In the second stage, followers decide whether to form a link with the leader based on the hard information and soft information that they see on the leader's profile page. The second stage is identical to the main model; the probability of follower i following leader j during period t is modeled as

$$Pr(y_{ijt}=1)=\frac{e^{(\boldsymbol{\alpha}X_{it-1}+\boldsymbol{\beta}W_{jt-1}+\boldsymbol{\lambda}V_{ij}+\boldsymbol{\tau}C_i+\eta_i)}}{1+e^{(\boldsymbol{\alpha}X_{it-1}+\boldsymbol{\beta}W_{jt-1}+\boldsymbol{\lambda}V_{ij}+\boldsymbol{\tau}C_i+\eta_i)}}. \tag{4.9}$$

Taking step 1 and step 2 together, I derive the overall likelihood as follows:

$$\begin{aligned} L_{ijt} = y_{ijt} \times (Pr(A_{jt}=1) \times Pr(y_{ijt}=1|A_{jt}=1)) \\ +(1-y_{ijt}) \times (Pr(A_{jt}=1) \times Pr(y_{ijt}=0|A_{jt}=1)+Pr(A_{jt}=0)) \end{aligned} \tag{4.10}$$

The overall log-likelihood value is further written as follows:

$$TLL(\gamma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\tau})=\sum_{i=1}^{I_t}\sum_{t=1}^{T_i}\sum_{j=1}^{J_{it}}\ln(L_{ijt}) \tag{4.11}$$

where I_t is the number of followers in period t, T_i is the number of periods that follower i exists on the platform, and J_{it} is the number of leaders that follower i can potentially follow in period t. I estimate the proposed two-stage selection model by maximizing the overall log-likelihood value. For computational tractability, following Heckman and Singer

(1984), I apply a non-parametric approach to estimate the follower's random effects after controlling for the time average of the follower's time-variant observables. The estimation results are reported in Table 4.5. I find that potential followers are more likely to be aware of a leader if that leader has a larger number of followers of his/her account (coefficient of 0.3483, $p < 0.01$). The results from the second stage are generally consistent with the findings from the main model.

4.5.2 Alternative sentiment dictionary

In the main analysis, I use the widely applied VADER sentiment dictionary to calculate the sentiment scores for comments. It is possible that different sentiment dictionaries generate different sentiment scores, which might affect the estimated effects of positive and negative scores for a leader's comments on link formation and dissolution. Therefore, in this robustness check, I first apply an alternative sentiment dictionary to calculate the sentiment scores for comments. In particular, I adopt the Harvard General Inquirer[13] dictionary, another widely adopted dictionary for extracting sentiment from social media, to perform the sentiment analysis (Ammann and Schaub 2021). After that, I re-estimate the STERGM with Chamberlain correlated random effects. The estimation results are presented in Table 4.6. I find that the results are generally consistent with the findings in the main model.

4.5.3 Alternative measure of financial risk

The volatility of performance is a symmetric measure of risk that takes into account both positive and negative deviations from the mean. Investors may, However, be mostly concerned with extreme negative profit outcomes, i.e., with large losses. Consequently, I consider the maximum drawdown (MDD) as an alternative risk measure that accounts for large losses. The MDD measures the monthly maximum observed loss in a leader's daily profit. Similar to the standard deviation, the MDD is a widely used risk measure (Cvitanić and Karatzas

---

[13]`http://www.wjh.harvard.edu/~inquirer/`, last accessed Jan. 8, 2021.

Table 4.5: Estimation Results of the Two-stage Selection Model

Variable	Formation	
First stage		
Num. of followers	0.3483***	(0.0430)
Second stage		
Social communication		
Leader's post quantity	0.2297***	(0.0407)
Leader's post quality	0.2674***	(0.0392)
Leader's number of replies	0.0927**	(0.0431)
Leader's comment received positive score	1.1130***	(0.2025)
Leader's comment received negative score	−3.4844***	(0.9125)
Follower's post quantity	−0.0424	(0.0644)
Follower's post quality	−0.1392	(0.1166)
Financial performance		
Leader's average profit	0.9184***	(0.1867)
Leader's std. dev. profit	0.5326	(0.6810)
Follower's average profit	−0.0414	(0.1270)
Follower's std. dev. profit	0.3549	(0.5457)
Leader's average holding time	0.0875	(0.0748)
Leader's lottery preference	0.8159	(0.4981)
Leader's HHI	−0.9456***	(0.1388)
Demographics		
Nationality	1.1929***	(0.1285)
Age	0.1773**	(0.0759)
Homophily (male)	0.9895***	(0.1729)
Homophily (female)	−1.1487**	(0.5223)
Image	−1.4969***	(0.5574)
Bio	1.4299***	(0.2641)
Experience	−0.5724	(0.3572)
Wealth	−0.4306	(0.3099)
Income	0.0968	(0.3723)
Risk	−1.2409***	(0.4730)
Network structure		
Leader's popularity	4.8850***	(0.4789)
Leader's activity	−3.0419**	(1.5117)
Follower's popularity	−17.6188***	(2.0538)
Follower's activity	30.8006***	(2.4906)
Transitivity	0.9729***	(0.0922)
Constant	17.6665***	(4.6990)
Log Likelihood	-13,426.99	
Observations	11,000,219	

Notes: The number of posts, the quality of posts (the number of likes) and the number of replies are log-transformed. Average profit and std. dev. profit are scaled by a factor of 10. Average holding time is scaled by a factor of 1/100. Popularity and activity are scaled by a factor of 1/100. wealth, income and risk are log-transformed and then scaled by a factor of 1/10. Experience is scaled by a factor of 1/10.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 4.6: Estimation Results Using an Alternative Sentiment Dictionary

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3177***	(0.026 2)	0.1213***	(0.028 2)
Leader's post quality	0.4695***	(0.024 3)	0.0356*	(0.020 8)
Leader's number of replies	−0.0151	(0.027 3)	−0.0227	(0.031 1)
Leader's comment received positive score	1.4295***	(0.161 7)	0.7911***	(0.183 4)
Leader's comment received negative score	−0.7144	(0.443 6)	−1.0227**	(0.414 3)
Follower's post quantity	0.0648	(0.047 4)	−0.5332***	(0.053 5)
Follower's post quality	−0.3354***	(0.079 7)	−0.0036	(0.074 8)
Financial performance				
Leader's average profit	0.0913***	(0.013 5)	0.0706***	(0.018 3)
Leader's std. dev. profit	−2.0776***	(0.548 5)	−2.6433***	(0.762 2)
Follower's average profit	0.0002	(0.011 5)	0.0425***	(0.011 9)
Follower's std. dev. profit	−0.6806	(0.508 5)	−8.0848***	(0.555 8)
Leader's average holding time	0.3416***	(0.044 1)	0.0737	(0.066 1)
Leader's lottery preference	0.0585	(0.362 9)	1.0407***	(0.394 0)
Leader's HHI	−0.5866***	(0.090 4)	−0.1197	(0.092 9)
Demographics				
Nationality	0.7461***	(0.070 3)	0.3395***	(0.079 7)
Age	0.1165**	(0.050 6)	0.0494	(0.054 3)
Homophily (male)	0.9602***	(0.135 9)	−0.1622	(0.116 4)
Homophily (female)	−0.8334**	(0.416 6)	0.3173	(0.318 0)
Image	2.4534***	(0.423 7)	0.7248	(0.465 6)
Bio	2.9068***	(0.187 3)	−0.1510	(0.188 5)
Experience	−0.0450	(0.037 1)	0.1445***	(0.040 2)
Wealth	0.0220	(0.032 4)	0.0603*	(0.034 6)
Income	−0.0046	(0.039 3)	0.0291	(0.040 8)
Risk	−0.0893*	(0.048 7)	−0.0395	(0.052 6)
Network structure				
Leader's popularity	0.0086***	(0.000 4)	0.0014***	(0.018 3)
Leader's activity	−0.0347***	(0.012 3)	0.0382***	(0.762 2)
Follower's popularity	−0.0534***	(0.012 6)	−0.0085	(0.011 9)
Follower's activity	0.0024	(0.007 0)	−0.0763***	(0.555 8)
Transitivity	0.0798***	(0.025 7)	−0.0203	(0.066 1)
Constant	−14.2699***	(5.438 8)	2.1800***	(0.644 2)
Log Likelihood	-13,643.37		-9,158.25	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

1999, de Melo Mendes and Lavrado 2017). I re-estimate the STERGM with Chamberlain correlated random effects; the estimation results are presented in Table 4.7. The coefficients on key determinants are generally consistent with the main results, and the MDD shows the expected negative effect.

4.6 Conclusions

Social trading is a novel form of trading that combines online brokerage with traditional social media features. It has attracted a large number of investors and increasing attention from practitioners and academia. Social trading allows investors to seek financial advice from their peers, to observe their peers' trading strategies, and to directly follow other investors in real time. Thus, inexperienced retail investors may benefit from their peers, while experienced investors are able to provide signals and earn additional income. Due to the monetary aspects involved in these leader-follower relationships, network evolution follows a distinct pattern that differs from that of traditional social media platforms. In particular, link dissolution is an important part of social trading.

I study a dynamic social trading network using a STERGM and examine how various factors, such as social communication, financial performance, and demographics, affect the link formation and dissolution processes. I show that social communication, financial performance, and demographics have different implications for the link formation and dissolution processes. Followers consider financial performance, social communication, and demographics when deciding whom to follow (link formation process). However, once a link is formed, demographic characteristics become less important, as followers mainly focus on leaders' financial performance as Ill as social communication to decide whether to sustain the link or not (link dissolution process). Focusing on the different types of social communication, I find that the quality and quantity of a leader's posts increase the likelihood of followers forming a new link and sustaining existing links. Followers are less likely to form new links or sustain existing links with leaders who receive more negative comments. Leaders who receive more positive comments are more likely to attract new followers and to keep existing followers.

Table 4.7: Estimation Results with the MDD as the Measure of Financial Risk

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3093***	(0.026 7)	0.1001***	(0.027 9)
Leader's post quality	0.5150***	(0.025 1)	0.0450**	(0.020 7)
Leader's number of replies	0.0242	(0.027 9)	−0.0071	(0.031 0)
Leader's comment received positive score	1.1092***	(0.139 6)	0.5300***	(0.145 9)
Leader's comment received negative score	−4.7028***	(0.753 1)	−2.0796***	(0.472 0)
Follower's post quantity	0.0652	(0.047 5)	−0.5491***	(0.054 8)
Follower's post quality	−0.3370***	(0.079 8)	−0.0298	(0.075 5)
Financial performance				
Leader's average profit	0.0369***	(0.005 2)	0.0199***	(0.006 8)
Leader's MDD	0.0221	(0.021 9)	−0.0625**	(0.024 7)
Follower's average profit	−0.0019	(0.010 6)	0.0574***	(0.011 4)
Follower's MDD	−0.1071***	(0.025 4)	−0.2378***	(0.026 2)
Leader's average holding time	0.3761***	(0.043 4)	0.0397	(0.066 6)
Leader's lottery preference	0.1687	(0.365 1)	0.9932**	(0.394 7)
Leader's HHI	−0.5842***	(0.090 6)	−0.1560*	(0.092 7)
Demographics				
Nationality	0.7488***	(0.070 4)	0.3282***	(0.079 7)
Age	0.1173**	(0.050 6)	0.0470	(0.054 3)
Homophily (male)	0.9847***	(0.136 1)	−0.1252	(0.116 7)
Homophily (female)	−0.8613**	(0.416 7)	0.2514	(0.317 2)
Image	2.5500***	(0.423 6)	0.7448	(0.468 8)
Bio	2.8402***	(0.183 3)	−0.1102	(0.188 3)
Experience	−0.0453	(0.037 4)	0.1181***	(0.042 6)
Wealth	0.0333	(0.032 8)	0.0915**	(0.036 8)
Income	−0.0144	(0.039 7)	0.0176	(0.043 3)
Risk	−0.0899*	(0.049 2)	−0.1038*	(0.055 8)
Network structure				
Leader's popularity	0.0083***	(0.000 4)	0.0016***	(0.000 4)
Leader's activity	−0.0369***	(0.012 5)	0.0400***	(0.011 9)
Follower's popularity	−0.0563***	(0.012 9)	−0.0112	(0.006 9)
Follower's activity	0.0048	(0.007 0)	−0.0714***	(0.009 0)
Transitivity	0.0840***	(0.025 7)	−0.0316	(0.066 0)
Constant	1.7241	(9.281 2)	2.9661***	(0.682 8)
Log Likelihood	-13,617.23		-9,231.18	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, MDD, wealth, income, and risk are log-transformed. Average profit is scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

In addition, the impacts of negative and positive comments are asymmetric. Negative comments have a larger impact than positive comments on the link formation and dissolution processes.

The study contributes to the growing literature on social trading by first modeling the dynamics of leader-follower networks. The granular data allow us to thoroughly examine the implications of various factors on the link formation and dissolution processes. The study also contributes to a better understanding of how hard information (e.g., financial performance) and, in particular, soft information (e.g., social communication) affect leader-follower network evolution in the social trading context.

The study has practical managerial implications. I document link formation and link dissolution processes and thereby broaden and deepen our understanding of leader-follower network evolution in social trading. Social trading platforms were established in the aftermath of the global financial crisis to provide retail investors with an alternative to traditional wealth management in response to the eroding trust in financial markets following the crisis (Doering et al. 2015). Although social trading platforms provide a high level of informational transparency, investors face new challenges in building trust, in particular because most investors on the platform are individuals who lack institutional endorsements and the relationship is online with a more and less anonymous mass. In this study, I find that social communication plays an important role in leaders' ability to convince potential followers to follow their trading strategies and existing followers to sustain their links. Social communication is effective in building trust among investors on social trading platforms. Therefore, leaders have to use caution when making posts. If not used properly, those posts receiving more negative comments can backfire, which can reduce link formation and increase link dissolution. Negative comments have a larger impact on link formation and dissolution than positive comments. In addition, leaders should make high-quality posts, as high-quality posts both increase link formation and reduce link dissolution. Thus, by communicating in a balanced manner, leaders can attract new followers to follow their trading strategies and encourage existing followers to sustain their links. The results can guide leaders on when and how to communicate with followers on social trading platforms. Given the importance of social communication in the evolution of leader-follower networks with real money flow, social trading platforms should carefully regulate social communication to sustain a healthy ecosystem.

My analysis has some caveats. First, I abstract away from the potential dependence between a follower's following decisions and a leader's subsequent trading strategies. Thus, I assume that a leader's trading strategy does not change regardless of any given follower's decision. However, such dependencies do exist (Pelster and Hofmann 2018) and may affect network evolution. Future research may aim to study these coevolution effects in more detail. Second, I focus only on a follower's decision regarding whether to follow a leader, without considering the monetary amount assigned to the link. Future research can consider the amount of money allocated to a link in a weighted directed (leader-follower) network. Such adjustments may reveal additional insights regarding the role of social communication in network evolution. Finally, I incorporate Chamberlain correlated random effects in the model to address the omitted variable issue; for example, some leaders might advertise themselves on other social media platforms, which might affect link formation. Such behavior indicates a leader's general propensity to engage in advertising their trading on social media, and such general propensities are rather stable over time. I acknowledge that the Chamberlain correlated random effects capture only the time-invariant unobservables and, for instance, do not account for the possibility that the advertising activities of investors on other social platforms change over time.

Chapter 5

CONCLUDING REMARKS

The rapid development of technology has brought about a lot of business innovations. These innovations not only improve the efficiency of business practice, but also shift individuals' behavior. My dissertation is motivated by a better understanding of individuals' decision-making and engagement because it is of focal importance to firms' revenue. In this dissertation, I look into two topics related to individuals' decision-making process. One is player game-play and reward ads watching behavior in online mobile games and the other is network evolution in social trading platforms.

In Chapter 3, I aim to examine players' dynamic engagement transition in mobile games. Higher engagement with games can increase revenue from existing players and additionally attract new players through word-of-mouth or network effects. Players' engagement is also an important factor to consider in deciding the mobile games' monetization strategy. To the best of my knowledge, this is the first study to investigate players' dynamic perception in games and connect the intrinsic motivation factors with players' reward ads watching behavior. Through a hidden Markov model, I find that players' perceived difficulty and perceived achievement have significant effects on players' latent engagement evolution, and the effect is heterogeneous depending on the gamer's corresponding engagement level. In addition, I find that a higher fluctuation of perceived difficulty increases the probability that a player moves to a higher engagement state and watching reward ads can also help players to transition to a higher engagement level. The findings make several contributions. First, I contribute to the literature in motivation theory by using real-world mobile game data and providing empirical evidence on how theory-based intrinsic motivation factors affect players' engagement. Second, I extend the motivation theory by examining the effect of fluctuations in gamers' perceptions on their engagement. Finally, I employ intrinsic motivation theory to examine the effects of reward ads on players' engagement. The empirical analysis also generates useful implications for both game publishers and third-party ad providers.

In Chapter 4, I study the determinants of link formation and link dissolution in a social trading platform. Social trading is a novel form of trading that combines online brokerage with traditional social media features, which has attracted a large number of investors and increasing attention from practitioners and academia. I study a dynamic social trading network evolution using a separable temporal exponential random graph model and examine how social communication, financial performance, demographics and network structure affect followers' following and unfollowing decisions. The results show that social communication, financial performance, demographics and network structures have different implications for the link formation and dissolution processes. Followers consider financial performance, social communication, and demographics when deciding whom to follow. However, once a link is formed, demographic characteristics become less important, as followers mainly focus on leaders' financial performance as well as social communication to decide whether to maintain the link or not. The findings contribute to the growing literature on social trading by first modeling the dynamics of leader-follower networks. In addition, it also contributes the literature on financial advice by showing that social communication, especially from leaders, can generate economic impacts. Third, the study contributes to the literature on individual investor behavior. While financial performance is an important signal of traders' trading skills, I find that followers also rely on communication when evaluating peer traders. Finally, from a methodological perspective, I incorporate Chamberlain's correlated random effects into the STERGM to alleviate concerns about confounding effects from individual-level unobserved heterogeneity in the network analysis. This study also has important managerial implications. The results on the impact of social communication can provide important guidance for leaders on when and how to communicate with followers. Second, the findings provide implications for the providers of social trading platforms regarding how to regulate social communication to sustain a healthy ecosystem.

BIBLIOGRAPHY

Agarwal V, Daniel ND, Naik NY (2009) Role of managerial incentives and discretion in hedge fund performance. *The Journal of Finance* 64(5):2221–2256.

Ammann M, Schaub N (2021) Do individual investors trade on investment-related internet postings? *Management science* 67(9):5679–5702.

Anderson SP, Coate S (2005) Market provision of broadcasting: A welfare analysis. *The review of Economic studies* 72(4):947–972.

Apesteguia J, Oechssler J, Weidenholzer S (2020) Copy trading. *Management Science* 66(12):5608–5622.

Atkinson JW (1974) Strength of motivation and efficiency of performance. *Motivation and achievement* 193–218.

Baker WE, Faulkner RR, Fisher GA (1998) Hazards of the market: The continuity and dissolution of interorganizational market relationships. *American sociological review* 63(2):147–177.

Bali TG, Hirshleifer D, Peng L, Tang Y (2021) Attention, social interaction, and investor attraction to lottery stocks. Technical report, National Bureau of Economic Research.

Barber B, Huang X, Odean T (2016) Which factors matter to investors? evidence from mutual fund flows. *The Review of Financial Studies* 29(10):2600–2642.

Barber BM, Odean T (2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies* 21(2):785–818.

Berger J, Kim YD, Meyer R (2021) What makes content engaging? how emotional dynamics shape success. *Journal of Consumer Research* 48(2):235–250.

Bleier A, Eisenbeiss M (2015) Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science* 34(5):669–688.

Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems* 50(2):511–521.

Chamberlain G (1980) Analysis of covariance with qualitative data. *The Review of Economic Studies* 47(1):225–238.

Chanel G, Rebetez C, Bétrancourt M, Pun T (2008) Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, 13–17.

Chang KTT, Koh ATT, Low BYY, Onghanseng DJS, Tanoto K, Thuong TST (2008) Why i love this online game: The mmorpg stickiness factor. *ICIS 2008 Proceedings* 88.

Chiong K, Chen R, Yang S (2017) Incentivized advertising: Treatment effect and adverse selection. *arXiv preprint arXiv:1709.00197* .

Chou W, Hunt Y, Beckjord E, Moser R, Hesse B (2009) Social media use in the united states: implications for health communication. *Journal of medical Internet research* 11(4):e1249.

Coval JD, Moskowitz TJ (1999) Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance* 54(6):2045–2073.

Csikszentmihalyi M, Csikzentmihaly M (1990) *Flow: The psychology of optimal experience*, volume 1990 (Harper & Row New York).

Cvitanić J, Karatzas I (1999) On dynamic measures of risk. *Finance and Stochastics* 3(4):451–482.

de Melo Mendes BV, Lavrado RC (2017) Implementing and testing the maximum drawdown at risk. *Finance Research Letters* 22:95–100.

Deci EL, Ryan RM, et al. (1985) The general causality orientations scale: Self-determination in personality. *Journal of research in personality* 19(2):109–134.

Devenow A, Welch I (1996) Rational herding in financial economics. *European economic review* 40(3-5):603–615.

Doering P, Neumann S, Paul S (2015) A primer on social trading networks–institutional aspects and empirical evidence. *EFMA annual meetings*.

Dorfleitner G, Fischer L, Lung C, Willmertinger P, Stang N, Dietrich N (2018) To follow or not to follow–an empirical analysis of the returns of actors on social trading platforms. *The Quarterly Review of Economics and Finance* 70:160–171.

Dorn D, Huberman G, Sengmueller P (2008) Correlated trading and returns. *The Journal of Finance* 63(2):885–920.

Duarte J, Siegel S, Young L (2012) Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies* 25(8):2455–2484.

Ely J, Frankel A, Kamenica E (2015) Suspense and surprise. *Journal of Political Economy* 123(1):215–260.

Epstein JA, Harackiewicz JM (1992) Winning is not enough: The effects of competition and achievement orientation on intrinsic interest. *Personality and Social Psychology Bulletin* 18(2):128–138.

Gemayel R, Preda A (2017) Does a scopic regime produce conformism? herding behavior among trade leaders on social trading platforms. *The European Journal of Finance* 24(14):1144–1175.

Gemayel R, Preda A (2018a) Does a scopic regime erode the disposition effect? evidence from a social trading platform. *Journal of Economic Behavior & Organization* 154:175–190.

Gemayel R, Preda A (2018b) Does a scopic regime produce conformism? herding behavior among trade leaders on social trading platforms. *The European Journal of Finance* 24(14):1144–1175.

Goetzmann WN, Ingersoll Jr JE, Ross SA (2003) High-water marks and hedge fund management contracts. *The Journal of Finance* 58(4):1685–1718.

Greve HR, Baum JAC, Mitsuhashi H, Rowley TJ (2010) Built to last but falling apart: Cohesion, friction, and withdrawal from interfirm alliances. *Academy of Management Journal* 53(2):302–322.

Grinblatt M, Jostova G, Petrasek L, Philipov A (2020) Style and skill: Hedge funds, mutual funds, and momentum. *Management Science* 66(12):5505–5531.

Groening C, Binnewies C (2019) "achievement unlocked!"-the impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior* 97:151–166.

Grote GF, James LR (1991) Testing behavioral consistency and coherence with the situation-response measure of achievement motivation. *Multivariate behavioral research* 26(4):655–691.

Gu Z, Bapna R, Chan J, Gupta A (2022) Measuring the impact of crowdsourcing features on mobile app user engagement and retention: A randomized field experiment. *Management Science* 68(2):1297–1329.

Gui J, Nagappan M, Halfond WG (2017) What aspects of mobile ads do users care about? an empirical study of mobile in-app ad reviews. *arXiv preprint arXiv:1702.07681* .

Guo H, Hao L, Mukhopadhyay T, Sun D (2019a) Selling virtual currency in digital games: Implications for gameplay and social welfare. *Information Systems Research* 30(2):430–446.

Guo H, Zhao X, Hao L, Liu D (2019b) Economic analysis of reward advertising. *Production and Operations Management* 28(10):2413–2430.

Hamari J, Keronen L, Alha K (2015) Why do people play games? a review of studies on adoption and use. *2015 48th Hawaii International Conference on System Sciences*, 3559–3568 (IEEE).

Han B, Hirshleifer D, Walden J (2022) Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis* 57(1):390–412.

Han SP, Park S, Oh W (2015) Mobile app analytics: A multiple discrete-continuous choice framework. *Management Information Systems Quarterly (MISQ), Forthcoming* .

Heckman J, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society* 271–320.

Heimer RZ (2016) Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies* 29(11):3177–3209.

Hennessey B, Moran S, Altringer B, Amabile TM (2015) Extrinsic and intrinsic motivation. *Wiley encyclopedia of management* 1–4.

Hirshleifer D, Teoh SH (2003) Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36:337–386.

Holland PW, Leinhardt S (1971) Transitivity in structural models of small groups. *Comparative group studies* 2(2):107–124.

Hsu CL, Lu HP (2004) Why do people play on-line games? an extended tam with social influences and flow experience. *Information & management* 41(7):853–868.

Huang J, Wei KD, Yan H (2007) Participation costs and the sensitivity of fund flows to past performance. *The Journal of Finance* 62(3):1273–1311.

Huang Y, Jasin S, Manchanda P (2019) "level up": Leveraging skill and engagement to maximize player game-play in online video games. *Information Systems Research* 30(3):927–947.

Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008) ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software* 24(3):nihpa54860.

Hutto C, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, volume 8, 216–225.

Hwang EH, Guo X, Tan Y, Dang Y (2022) Delivering healthcare through teleconsultations: Implications for offline healthcare disparity. *Information Systems Research* forthcoming.

Ishihara M, Ching AT (2019) Dynamic demand for new and used durable goods without physical depreciation: The case of japanese video games. *Marketing Science* 38(3):392–416.

Ivković Z, Weisbenner S (2009) Individual investor mutual fund flows. *Journal of Financial Economics* 92(2):223 – 237, ISSN 0304-405X.

Kaiser M, Berger ESC (2021) Trust in the investor relationship marketing of startups: a systematic literature review and research agenda. *Management Review Quarterly* 71(2):491–517.

Kane GC, Labianca GJ, Borgatti SP (2014) What's different about social media networks? a framework and research agenda. *MIS Quarterly* 38(1):275–304.

Khern-am-nuai W, Kannan K, Ghasemkhani H (2018) Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* 29(4):871–892.

Kim H, Kankanhalli A, Lee S (2018) Examining gifting through social network services: A social exchange theory perspective. *Information Systems Research* 29(4):805–828.

Krivitsky PN, Handcock MS (2014) A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 76(1):29.

Kumar A (2009) Who gambles in the stock market? *The Journal of Finance* 64(4):1889–1933.

Kwon HE, So H, Han SP, Oh W (2016) Excessive dependence on mobile social apps: A rational addiction perspective. *Information Systems Research* 27(4):919–939.

Lee C, Ofek E, Steenburgh TJ (2018a) Personal and social usage: The origins of active customers and ways to keep them engaged. *Management Science* 64(6):2473–2495.

Lee D, Hosanagar K, Nair HS (2018b) Advertising content and consumer engagement on social media: evidence from facebook. *Management Science* 64(11):5105–5131.

Lee GM, Qiu L, Whinston AB (2016) A friend like me: Modeling network formation in a location-based social network. *Journal of Management Information Systems* 33(4):1008–1033.

Lee W, Ma Q (2015) Whom to follow on social trading services? a system to support discovering expert traders. *The Tenth International Conference on Digital Information Management (ICDIM 2015)* 188–193.

Li Z, Fang X, Bai X, Sheng O (2017) Utility-based link recommendation for online social networks. *Management Science* 63(6):1938–1952.

Liberti JM, Petersen MA (2019) Information: Hard and soft. *Review of Corporate Finance Studies* 8(1):1–41.

Lin M, Viswanathan S (2016) Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* 62(5):1393–1414.

Liu D, Kumar S, Mookerjee VS (2012) Advertising strategies in electronic retailing: A differential games approach. *Information Systems Research* 23(3-part-2):903–917.

Liu D, Li X, Santhanam R (2013) Digital games and beyond: What happens when players compete? *Mis Quarterly* 111–124.

Liu H (2010) Dynamics of pricing in the video game console market: skimming or penetration? *Journal of marketing research* 47(3):428–443.

Lomas JD, Koedinger K, Patel N, Shodhan S, Poonwala N, Forlizzi JL (2017) Is difficulty overrated? the effects of choice, novelty and suspense on intrinsic motivation in educational games. *Proceedings of the 2017 CHI conference on human factors in computing systems*, 1028–1039.

MacDonald IL, Zucchini W (1997) *Hidden Markov and other models for discrete-valued time series*, volume 110 (CRC Press).

Malone TW (1981) Toward a theory of intrinsically motivating instruction. *Cognitive science* 5(4):333–369.

Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Academy of management review* 20(3):709–734.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual review of sociology* 27(1):415–444.

Mundlak Y (1978) On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society* 69–85.

Netzer O, Lattin JM, Srinivasan V (2008) A hidden markov model of customer relationship dynamics. *Marketing science* 27(2):185–204.

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society* 1–32.

Oehler A, Horn M, Wendt S (2016) Benefits from social trading? empirical evidence for certificates on wikifolios. *International Review of Financial Analysis* 46:202–210.

Pan W, Altschuler Y, Pentland AS (2012) Decoding social influence and the wisdom of the crowd in financial trading network. IEEE, ed., *International Conference on Privacy, Security, Risk and Trust and International Conference on Social Computing*, 203–209 (Washington DC).

Pedersen LH (2022) Game on: Social networks and markets. *Journal of Financial Economics* forthcoming.

Pelster M (2017) I'll have what s/he's having: A case study of a social trading network. *Proceedings of the International Conference on Information Systems 2017* .

Pelster M, Hofmann A (2018) About the fear of reputational loss: Social trading and the disposition effect. *Journal of Banking & Finance* 94:75–88.

Polidoro F, Ahuja G, Mitchell W (2011) When the social structure overshadows competitive incentives: The effects of network embeddedness on joint venture dissolution. *Academy of Management Journal* 54(1):203–223.

Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. *Social networks* 29(2):173–191.

Röder F, Walter A (2019) What drives investment flows into social trading portfolios? *Journal of Financial Research* 42(2):383–411.

Rolls BJ, Rolls ET, Rowe EA, Sweeney K (1981) Sensory specific satiety in man. *Physiology & behavior* 27(1):137–142.

Rubin Z (1975) Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology* 11(3):233–260.

Ryan RM, Deci EL (2000a) Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* 25(1):54–67.

Ryan RM, Deci EL (2000b) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55(1):68.

Ryan RM, Rigby CS, Przybylski A (2006) The motivational pull of video games: A self-determination theory approach. *Motivation and emotion* 30(4):344–360.

Sagie A, Elizur D, Yamauchi H (1996) The structure and strength of achievement motivation: A cross-cultural comparison. *Journal of Organizational Behavior* 17(5):431–444.

Selle K (2020) Rewarded ads reward more than just app users. URL `https://applift.com/blog/rewarded-ads-reward-more-than-just-app-users`.

Sevilla J, Lu J, Kahn BE (2019) Variety seeking, satiation, and maximizing enjoyment over time. *Journal of Consumer Psychology* 29(1):89–103.

Shafi K, Mohammadi A, Johan SA (2020) Investment ties gone awry. *Academy of Management Journal* 63(1):295–327.

Sheng L, Ryan CT, Nagarajan M, Cheng Y, Tong C (2020) Incentivized actions in freemium games. *Manufacturing & Service Operations Management* .

Singh PV, Tan Y, Youn N (2011) A hidden markov model of developer learning dynamics in open source software projects. *Information Systems Research* 22(4):790–807.

Sirri ER, Tufano P (1998) Costly search and mutual fund flows. *The journal of finance* 53(5):1589–1622.

Snijders TA, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. *Sociological methodology* 36(1):99–153.

Strauss D, Ikeda M (1990) Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* 85(409):204–212.

Teixeira T, Picard R, El Kaliouby R (2014) Why, when, and how much to entertain consumers in advertisements? a web-based facial tracking field study. *Marketing Science* 33(6):809–827.

Tirunillai S, Tellis GJ (2012) Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science* 31(2):198–215.

Todri V, Ghose A, Singh PV (2020) Trade-offs in online advertising: Advertising effectiveness and annoyance dynamics across the purchase funnel. *Information Systems Research* 31(1):102–125.

Wasserman S, Faust K, et al. (1994) *Social network analysis: Methods and applications* (Cambridge University Press).

White RW (1959) Motivation reconsidered: The concept of competence. *Psychological review* 66(5):297.

Wohlgemuth V, Berger ES, Wenzel M (2016) More than just financial performance: Trusting investors in social trading. *Journal of Business Research* 69(11):4970–4974.

Wooldridge JM (2010) *Econometric analysis of cross section and panel data* (MIT press).

Wu CC, Chen YJ, Cho YJ (2013) Nested network effects in online free games with accessory selling. *Journal of Interactive Marketing* 27(3):158–171.

Xu J, Chau M (2018) Cheap talk? the impact of lender-borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems* 35(1):53–85.

Yan L, Peng J, Tan Y (2015) Network dynamics: how can we find patients like us? *Information Systems Research* 26(3):496–512.

Yan L, Tan Y (2014) Feeling blue? go online: An empirical study of social support among patients. *Information Systems Research* 25(4):690–709.

Yang M, Zheng Z, Mookerjee V (2021) How much is financial advice worth? the transparency-revenue tension in social trading. *Management Science* forthcoming.

Yu SZ, Kobayashi H (2003) An efficient forward-backward algorithm for an explicit-duration hidden markov model. *IEEE signal processing letters* 10(1):11–14.

Zhang K, Katona Z (2012) Contextual advertising. *Marketing Science* 31(6):980–994.

Zhang Y, Li B, Luo X, Wang X (2019) Personalized mobile targeting with user engagement stages: Combining a structural hidden markov model and field experiment. *Information Systems Research* 30(3):787–804.

Appendix A

A.1 Selection Number of Supporters

I use a non parametric approach to estimate the random effects ζ and η. Besides, I allow this individual specific unobserved heterogeneity to be correlated by assuming a conditional distribution. Instead of specifying a specific distribution for ζ and η, I assume a finite set of supporters associated with their corresponding probability mass functions. Then, I are able to integrate the random effects into the Log-likelihood and maximize the function. I try different combination of supporter numbers and select the one that has the largest Log-likelihood value. The results are shown in Table A.1.

Table A.1: Combination of Supporters

Supporter combination	Log-likelihood
(2,2)	-20241.73
(2,3)	-20243.34
(3,2)	**-20196.57**
(3,3)	-20227.32
(3,4)	-20218.88

A.2 The Balance between Player's Skill and Perceived Difficulty Level

Previous studies suggest that a balance between difficulty and player's skill is a key to success when participating activities, and gamers may become anxious if difficulty is way above the skill while feel bored if it is very much below the ability (Csikszentmihalyi and

Csikzentmihaly 1990). In this robustness check, I also examine how the mismatch between difficulty and player's skill affects player's engagement state transitions. However, I cannot observe player's skill through the observational data, and I infer this information from her historical activities. To be specific, I use the average puzzle solving time over all puzzles that the player i solved as a proxy for her skill in the game (denoted as a_i).[1] The basic assumption behind is that the puzzle sequence is pre-determined in the game design phase which is the same for all players. If a player is good at this game, she will use less time to solve the puzzle on average, which can be used as an indication of her corresponding puzzle solving skill. Next, I calculate the absolute difference in ability a_i and the in-period difficulty that the player i perceives (i.e.,average puzzle solving time for player i in period t), denoted as Δa_{it}, to measure the *mismatch* between difficulty of the puzzle and player's skill. The results are shown in Table A.2. The *mismatch* between difficulty and player's skill has a significantly negative effect on players with a low level of engagement, which indicates that compared with players who are in medium or high engagement state, players in a low engagement state are less tolerant with the mismatch between her skill and game difficulty level.

[1] I assume that the skill is fixed for each player during the data period because I only have two months observational data and it is unlikely for players to improve their skill drastically in such a short time.

Table A.2: HMM with Mismatch between Difficulty and Player's Skill

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−1.5955	(1.158 6)	−14.5949***	(1.063 7)	−1.7107	(1.315 7)
Soft currency purchase	3.1409***	(1.034 2)	7.3101***	(1.020 7)	2.4920**	(1.045 5)
In-app purchases	−0.9645	(1.280 7)	−9.7136***	(1.010 3)	−0.4268	(1.420 0)
Puzzle difficulty	4.9548***	(1.076 4)	2.4504**	(1.015 4)	10.5342***	(1.019 5)
Frequency	0.1528	(1.343 6)	3.7759***	(1.064 3)	−2.1454**	(0.977 1)
Latest intensity	−1.1494	(1.217 8)	−0.0724	(1.068 2)	0.8718	(1.113 4)
Historical intensity	0.3719	(1.647 5)	0.2584	(1.200 3)	1.9811*	(1.101 0)
Irregularity	−4.8276***	(1.000 9)	−0.6628	(1.001 2)	−9.4220***	(1.226 3)
Weekend	0.0672	(1.005 3)	−0.1350	(1.027 2)	−0.0707	(1.015 7)
Work hour	−0.0995	(1.057 4)	−0.5684	(1.015 1)	−0.0031	(1.002 5)
Constant	−1.9505*	(1.159 6)	−5.3029***	(1.005 0)	1.1994	(0.906 0)
θ Dispersion rate	0.5461*	(0.297 0)	−0.8364	(1.075 1)	−1.1150	(0.809 7)
Intercept in the logit probability of playing	−1.5165**	(0.708 6)	1.3948***	(0.180 6)	1.4024***	(1.173 1)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	10.3259***	(1.109 0)	10.0034***	(1.000 1)	9.4616***	(1.328 9)
Fluctuation of perceived difficulty^2	−10.6366***	(1.000 1)	−4.9396***	(1.046 1)	−6.8566***	(1.038 6)
Reward ads	10.1980***	(1.253 3)	−1.1874	(3.820 6)	11.4481***	(2.059 1)
Perceived difficulty	5.8721***	(1.806 0)	−5.9123***	(1.297 8)	6.4276***	(1.018 4)
Perceived achievement^{-1}	−3.6528***	(1.102 6)	−23.1460***	(1.145 5)	−10.0944***	(1.066 5)
Cumulative playing time	15.1250***	(1.007 6)	5.8751***	(1.011 9)	1.3941	(3.366 8)
Perceived difficulty * Reward ads	3.7028***	(1.002 9)	2.8767**	(1.142 6)	6.2510***	(1.017 6)
Perceived achievement^{-1} * Reward ads	−3.6199***	(1.000 1)	−5.1636***	(1.012 0)	−3.7358***	(1.002 7)
Mismatch between challenge and skill Δa_{it}	−3.2059*	(1.851 9)	0.4074	(1.157 1)	0.3997	(1.157 0)
Thresholds						
State 1			2.1253	(1.396 1)	2.1326***	(1.528 1)
State 2	−11.8505***	(1.393 2)			15.5840***	(1.000 2)
State 3	−10.8095***	(1.612 0)	−9.7361	(0.882 1)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20216.88			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.3 Alternative User Sample

In the main section, I select players who join the platform after Jan 1 in 2018 (the beginning date of the data collection period) to track their complete activities after downloading the app. I also try another sampling strategy, that is randomly select 1715 players without requiring that they start to play the game after Jan 1. The results are presented in Table A.3. As shown in the table, the main findings generally remain consistent.

A.4 Alternative Model to Model Transition Probabilities

In the conventional hidden Markov model, a player probabilistically belongs to an engagement state where 1 is the lowest state and n is the highest state. Thus, the states (1, 2, . . . , n) are ordinal dependent variables and I use the ordinal regression model to describe the transition process where ordered logit model is a widely adopted one in the literature (Netzer et al. 2008, Huang et al. 2019, Zhang et al. 2019). The ordered logit model assumes the error term to follow a type I extreme value distribution and therefore the likelihood can have a closed form solution. An alternative option is to use the ordered probit model to model transition probabilities where the error term follows a normal distribution. Specifically, the state transition matrix from period t to $t+1$ is defined as follows.

$$
\begin{aligned}
q_{it}(j,n) &= \Phi(\mu_n - \alpha_j * X_{it} - \eta_i) - \Phi(\mu_{n-1} - \alpha_j * X_{it} - \eta_i) \\
q_{it}(j,j+1) &= \Phi(\mu_{j+1} - \alpha_j * X_{it} - \eta_i) - \Phi(\mu_j - \alpha_j * X_{it} - \eta_i) \\
q_{it}(j,1) &= \Phi(\mu_1 - \alpha_j * X_{it} - \eta_i) - \Phi(\mu_0 - \alpha_j * X_{it} - \eta_i)
\end{aligned} \tag{A.1}
$$

where μ specifies the engagement transition cut points, and $\mu_n \geqslant \mu_{n-1} \geqslant ... \geqslant \mu_j \geqslant ... \geqslant \mu_0$. In addition, $\mu_0 = -\infty$ and $\mu_n = \infty$. X_{it} is a vector of covariates that affect state transition, α_j is a vector of corresponding coefficients, and η_i is the player specific random effect to capture the individual unobserved heterogeneity that may affect state transition. The results are reported in Table A.4 and the main conclusions are in general not sensitive to the model specification of the transition process.

Table A.3: HMM with Alternative User Sample

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−0.0794	(1.363 2)	−0.6158	(1.126 1)	0.2235	(0.927 0)
Soft currency purchase	4.3228***	(1.026 9)	4.2153*	(2.525 6)	3.7691***	(1.303 2)
In-app purchases	0.2974	(1.310 7)	−0.6710	(4.955 6)	−0.0888	(1.185 9)
Puzzle difficulty	4.4391***	(1.527 5)	2.8109	(2.233 2)	4.9949	(5.468 5)
Frequency	0.0633	(1.489 6)	0.1689	(1.140 1)	−0.3483	(0.987 5)
Latest intensity	−0.5249	(2.609 5)	0.9090	(3.173 2)	0.6557	(1.070 7)
Historical intensity	−0.8863	(1.608 8)	1.7104	(1.172 6)	2.0753	(3.393 3)
Irregularity	−1.8117	(1.307 3)	−0.1723	(1.006 1)	−5.6636***	(1.842 4)
Weekend	0.0598	(1.062 4)	0.0739	(1.323 5)	−0.0071	(0.943 2)
Work hour	−0.0947	(1.369 1)	−0.0339	(1.388 1)	0.0133	(0.979 8)
Constant	−3.3386*	(1.808 5)	−5.0030***	(1.729 7)	−1.1803	(2.942 7)
θ Dispersion rate	0.4640**	(0.214 2)	−0.5891	(0.965 4)	−1.1782	(1.002 4)
Intercept in the logit probability of playing	−1.6945*	(0.880 0)	0.8065**	(0.382 6)	0.8069***	(2.135 2)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	6.1406***	(2.326 2)	9.6066***	(2.918 0)	4.4817***	(1.583 8)
Fluctuation of perceived difficulty^2	−5.3207***	(1.001 6)	−2.3029**	(1.000 3)	−3.4562***	(1.025 9)
Reward ads	7.2943***	(1.060 8)	2.1562*	(1.267 0)	13.8117	(8.738 2)
Perceived difficulty	−0.5281	(5.858 6)	−0.5443	(1.096 8)	12.3180***	(4.095 0)
Perceived achievement^{-1}	−4.9477	(4.172 1)	−11.3826***	(1.182 4)	−3.2192**	(1.475 9)
Cumulative playing time	3.1999**	(1.383 4)	4.7039***	(1.243 4)	1.7203	(2.431 1)
Perceived difficulty * Reward ads	2.2640**	(1.000 9)	1.8052*	(1.001 4)	4.2417***	(1.258 4)
Perceived achievement^{-1} * Reward ads	−1.8474*	(1.028 5)	−2.5086**	(1.000 3)	−1.7119*	(1.004 1)
Thresholds						
State 1			−3.8823	(6.240 1)	−1.4109	(1.348 8)
State 2	−6.8720	(4.375 8)			2.2638**	(1.060 5)
State 3	−5.1687	(3.164 8)	−5.0132	(2.628 8)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-47562.62			
Individuals			1715			
Observations			42265			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.4: HMM with Ordered Probit Model State Transitions

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−1.5917	(1.222 9)	−14.9580***	(2.188 8)	−1.3526	(1.046 0)
Soft currency purchase	3.4215***	(1.117 7)	7.4292***	(1.266 4)	2.4282**	(1.008 7)
In-app purchases	−1.0062	(1.335 4)	−9.6371***	(1.003 0)	−0.2769	(1.068 1)
Puzzle difficulty	5.7948***	(1.973 7)	1.4213	(1.168 1)	10.6869***	(1.016 0)
Frequency	−0.0661	(1.038 7)	5.4238***	(1.025 4)	−2.2280**	(1.049 4)
Latest intensity	−1.3135	(1.615 0)	−0.0799	(1.077 7)	0.8073	(1.180 5)
Historical intensity	0.3750	(1.010 6)	0.2432	(1.083 9)	2.1023**	(1.039 3)
Irregularity	−5.4051***	(1.280 1)	−0.3965	(1.006 2)	−8.2325***	(1.000 7)
Weekend	−0.0375	(1.006 9)	−0.0643	(1.152 4)	−0.0048	(0.985 6)
Work hour	−0.0568	(0.991 2)	−0.5756	(1.074 2)	0.0797	(1.000 8)
Constant	−2.1149**	(0.951 2)	−7.1617***	(1.005 7)	0.8048	(0.984 7)
θ Dispersion rate	−0.5549	(0.447 4)	−0.8647	(1.001 0)	−1.3038	(1.081 5)
Intercept in the logit probability of playing	−1.5099**	(0.589 0)	1.5975***	(0.421 7)	1.6150***	(1.228 7)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	8.7417***	(1.067 0)	6.0623***	(2.120 6)	5.4048***	(1.688 4)
Fluctuation of perceived difficulty^2	−9.8490***	(1.000 1)	−4.9258***	(1.279 4)	−7.9705***	(1.234 7)
Reward ads	9.8927***	(3.557 2)	−0.4612	(6.411 1)	5.8269***	(1.051 3)
Perceived difficulty	0.4450	(3.498 9)	−2.0486	(1.533 8)	3.5682	(3.056 2)
Perceived achievement^{-1}	−2.5534*	(1.399 8)	−15.8258***	(4.670 9)	−5.6765***	(1.237 6)
Cumulative playing time	12.5487***	(2.809 0)	2.9524**	(1.256 1)	0.0390	(1.805 4)
Perceived difficulty * Reward ads	2.7022***	(1.027 6)	3.3711**	(1.340 2)	2.8344**	(1.402 4)
Perceived achievement^{-1} * Reward ads	−2.9410***	(1.027 1)	−3.6872***	(1.052 7)	−6.7646***	(1.069 7)
Thresholds						
State 1			1.7890	(1.274 6)	2.9139	(0.977 8)
State 2	−10.5696***	(1.257 9)			1.6911***	(0.431 3)
State 3	−10.1259***	(1.114 0)	−9.7689	(0.967 5)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20198.12			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.5 Alternative Outcomes

In the main model, I use the number of reward ads as the observed outcome. In this section, I also report the results when setting the in-period playing time (Z_{it}) as the observed outcome and assume it follows a log normal distribution. The results are reported in Table A.5, and my findings on variables affecting state transitions generally remain consistent.

A.6 Using Cumulative Average for Variables in State Transition

First order HMM is widely applied in literature when modeling users' latent engagement (Netzer et al. 2008, Huang et al. 2019, Zhang et al. 2019). In this study, I use the first order HMM mainly due to the tractability of computation. However, I conduct a robustness check, in which I use the cumulative average values until period t when constructing variables affecting state transition from t to $t+1$ (i.e., X_{it}). In this way, X_{it} contains information from the first period till t, which captures the effects from previous states to some extent. The results are reported in Table A.6, and my findings on variables affecting state transitions still hold.

A.7 Estimation with Different Number of States

In the main model, I find 3 hidden states outperforms all other model specifications because it has a smaller BIC and AIC. In this section, I also report the results using 2 hidden states and 4 hidden states.

Table A.5: HMM with Alternative Outcomes

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−0.0759	(1.282 1)	0.0095	(1.012 6)	−0.0473	(1.034 3)
Soft currency purchase	1.8347	(1.137 3)	1.7472*	(1.050 0)	2.6979**	(1.054 4)
In-app purchases	−0.6807	(1.326 5)	−0.4727	(1.082 4)	−0.1903	(1.385 8)
Puzzle difficulty	4.6628***	(1.728 4)	4.4645	(4.004 2)	5.2533***	(1.331 4)
Frequency	−0.2673	(1.127 9)	0.8801	(1.157 0)	−1.3917	(1.126 4)
Latest intensity	0.4084	(1.096 9)	0.2616	(1.019 3)	0.0500	(2.847 3)
Historical intensity	0.3915	(1.258 4)	0.2175	(1.531 1)	0.4204	(3.145 0)
Irregularity	−4.5794	(6.190 7)	0.1539	(1.977 8)	−6.9029**	(3.278 0)
Weekend	0.0054	(0.868 2)	−0.0329	(1.038 9)	0.1009	(1.029 5)
Work hour	−0.0071	(0.907 0)	−0.1087	(1.098 0)	−0.1112	(1.258 0)
Constant	3.9239***	(1.300 1)	3.9244***	(1.000 7)	5.0969	(1.571 2)
σ standard deviation	0.6995	(0.499 2)	0.6493	(0.948 5)	0.6881	(0.936 2)
Intercept in the logit probability of playing	−1.4094*	(0.802 7)	1.8701	(1.016 5)	2.7859	(9.173 0)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	9.3363***	(1.273 2)	10.5954***	(1.145 5)	11.3160***	(1.018 6)
Fluctuation of perceived difficulty^2	−10.9006***	(1.000 5)	−3.3774***	(1.003 0)	−6.2825***	(1.000 2)
Reward ads	7.6586***	(1.340 2)	6.9481***	(1.066 8)	4.7716***	(1.012 4)
Perceived difficulty	1.2136	(1.498 5)	−3.5520	(2.818 4)	6.9939***	(2.340 1)
Perceived achievement^{-1}	−5.7180**	(2.239 0)	−10.9931***	(1.023 8)	−3.3249***	(1.119 5)
Cumulative playing time	11.9178***	(1.032 2)	5.1015***	(1.294 2)	1.4093	(4.043 2)
Perceived difficulty * Reward ads	3.0288***	(1.003 1)	3.5855***	(1.000 1)	5.4352***	(1.002 9)
Perceived achievement^{-1} * Reward ads	−2.8885***	(1.000 1)	−4.4062***	(1.000 1)	−4.3448***	(1.000 5)
Thresholds						
State 1			3.2418***	(1.253 2)	3.2427***	(1.006 7)
State 2	−11.8593***	(1.879 9)			37.5530***	(1.000 1)
State 3	−6.9619***	(2.096 6)	−6.9593***	(1.016 9)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20042.33			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.6: Using Cumulative Average for Variables in State Transition

Parameter	State 1 (low)		State 2 (medium)		State 3 (high)	
Variables Affecting Observed Outcome						
Coin balance	−1.5623	(1.383 9)	−15.3755***	(1.759 9)	−1.4302	(1.666 0)
Soft currency purchase	3.2744***	(1.083 4)	7.8982***	(1.019 0)	2.4254**	(1.144 3)
In-app purchases	−0.5229	(2.275 2)	−10.1675***	(1.146 0)	−0.2180	(1.054 9)
Puzzle difficulty	4.1097***	(1.311 8)	4.2680	(2.904 6)	10.0182**	(3.959 0)
Frequency	0.0414	(1.723 4)	2.9548	(3.485 8)	−2.5508**	(1.026 0)
Latest intensity	−0.9472	(2.410 1)	−0.4213	(1.250 7)	0.6755	(1.188 9)
Historical intensity	0.8881	(1.584 9)	−0.0875	(1.740 3)	2.0794	(1.386 7)
Irregularity	−6.8033***	(2.157 6)	−0.8239	(1.029 1)	−11.4887***	(2.039 6)
Weekend	0.0086	(1.022 9)	−0.0387	(1.089 2)	−0.0161	(1.095 0)
Work hour	−0.0793	(1.126 3)	−0.5932	(1.210 9)	0.0871	(1.057 3)
Constant	−2.0571	(1.260 8)	−4.9525	(4.438 2)	1.2524	(1.383 1)
θ Dispersion rate	0.5498	(0.378 2)	−0.8100	(0.969 1)	−1.2848	(1.022 5)
Intercept in the logit probability of playing	−1.5241***	(0.254 2)	1.5747***	(0.310 7)	1.5794***	(1.459 5)
Variables Affecting State Transition						
Fluctuation of perceived difficulty	11.3327***	(1.760 9)	13.4045***	(3.984 4)	12.6825***	(2.527 6)
Fluctuation of perceived difficulty^2	−10.4996***	(1.000 7)	−5.5796***	(1.160 4)	−6.6478***	(1.097 1)
Reward ads	12.5240***	(2.619 1)	−6.0338	(7.398 3)	10.2157**	(4.516 7)
Perceived difficulty	5.6956**	(2.798 0)	−6.7228***	(1.504 0)	5.7715*	(3.061 5)
Perceived achievement^{-1}	−4.2664***	(1.017 3)	−30.9795***	(6.230 3)	−10.6300***	(1.029 8)
Cumulative playing time	17.5639***	(5.795 3)	4.4104	(2.983 3)	0.4373	(1.755 4)
Perceived difficulty * Reward ads	4.4240***	(1.032 9)	2.2250*	(1.348 0)	5.2762***	(1.011 7)
Perceived achievement^{-1} * Reward ads	−3.6365***	(1.001 8)	−5.3449***	(1.031 8)	−4.6816***	(1.000 2)
Thresholds						
State 1			2.8536**	(1.362 1)	4.5919	(1.038 8)
State 2	−13.2191***	(1.386 6)			13.5086***	(1.000 1)
State 3	−12.5088***	(1.551 4)	−11.7810	(1.065 9)		
Individual random effect	Yes		Yes		Yes	
Log Likelihood			-20210.97			
Individuals			1232			
Observations			22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.7: HMM with 2 States

Parameter	State 1		State 2	
Variables Affecting Observed Outcome				
Coin balance	−0.1921	(1.160 8)	−7.7980***	(1.025 3)
Soft currency purchase	3.3169**	(1.412 2)	5.0074***	(1.106 7)
In-app purchases	−1.8417	(1.158 0)	−1.6142	(1.748 9)
Puzzle difficulty	2.2510**	(1.140 4)	6.0210	(4.974 9)
Frequency	1.8033	(1.400 1)	−1.1817	(1.085 9)
Latest intensity	0.0188	(1.034 1)	−0.1509	(1.014 6)
Historical intensity	0.4825	(1.100 7)	−0.0769	(1.049 5)
Irregularity	−1.9046*	(1.008 4)	−8.4054***	(1.005 7)
Weekend	−0.1437	(1.001 0)	−0.0610	(1.142 1)
Work hour	−0.1494	(1.074 8)	−0.1379	(1.090 3)
Constant	−3.8106**	(1.668 4)	−0.8623	(0.997 2)
θ Dispersion rate	0.5092	(0.321 9)	−1.1509	(0.987 1)
Intercept in the logit probability of playing	−1.3528***	(0.390 5)	2.2431***	(0.709 6)
Variables Affecting State Transition				
Fluctuation of perceived difficulty	8.8592***	(1.042 3)	10.3291***	(1.837 2)
Fluctuation of perceived difficulty^2	−10.9200***	(1.010 7)	−7.8154***	(1.375 8)
Reward ads	7.5047***	(1.776 5)	2.6534	(5.428 9)
Perceived difficulty	3.6075*	(1.865 9)	−0.8796	(1.125 0)
Perceived achievement^{-1}	−3.7858***	(1.203 7)	−17.5834***	(6.731 7)
Cumulative playing time	17.2073*	(8.846 0)	1.7915	(1.378 4)
Perceived difficulty * Reward ads	3.6983***	(1.026 8)	4.0536***	(1.225 0)
Perceived achievement^{-1} * Reward ads	−3.3624***	(1.030 9)	−4.4753***	(1.015 4)
Thresholds				
State 1			3.0437**	(1.546 5)
State 2	−13.5289***	(2.396 0)		
Log Likelihood	-22520.22			
Individuals	1232			
Observations	22325			

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.
The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.8: HMM with 4 States

Parameter	State 1		State 2		State 3		State 4	
Variables Affecting Observed Outcome								
Coin balance	−0.7512	(1.046 0)	−15.0642***	(2.356 2)	−7.7061**	(3.736 7)	−0.7731	(1.038 7)
Soft currency purchase	3.3734***	(1.089 3)	4.0763***	(1.450 6)	5.2046***	(1.267 6)	2.8976**	(1.223 7)
In-app purchases	−1.2715	(1.098 5)	−9.5532***	(1.001 1)	−1.9760	(2.053 8)	−0.7516	(1.008 0)
Puzzle difficulty	6.2935***	(2.053 5)	1.9261	(1.302 5)	2.7629*	(1.605 1)	5.0454***	(1.312 9)
Frequency	−0.6687	(1.281 1)	3.5541	(5.517 8)	1.4379	(1.030 5)	−1.3305	(0.926 8)
Latest intensity	−0.2524	(1.468 1)	−1.7698	(3.063 3)	0.0293	(1.506 3)	0.4116	(1.075 3)
Historical intensity	0.4059	(1.168 2)	−1.4537	(3.033 4)	0.3775	(2.403 9)	1.0338	(1.011 0)
Irregularity	−5.8156***	(1.774 5)	−0.4888	(1.126 3)	−4.6663***	(1.076 8)	−7.5103***	(1.932 8)
Weekend	−0.1351	(1.416 7)	0.1157	(1.118 0)	0.1020	(1.034 0)	−0.1848	(0.963 4)
Work hour	−0.0575	(1.135 7)	0.2600	(1.066 8)	−0.2540	(1.067 4)	−0.0595	(0.984 0)
Constant	−1.2187	(1.081 2)	−1.5332	(5.939 1)	−2.8668	(1.923 8)	1.8280**	(0.857 1)
θ Dispersion rate	0.4136***	(0.090 0)	−1.5020	(1.290 0)	−0.7630	(0.992 2)	1.1042	(0.899 8)
Intercept in the logit probability of playing	−1.5381*	(0.913 5)	1.4439***	(0.316 7)	1.4451***	(2.523 7)	1.4460***	(2.364 7)
Variables Affecting State Transition								
Fluctuation of perceived difficulty	8.3550***	(1.361 0)	9.9753***	(1.090 9)	10.2950***	(1.008 3)	9.1979***	(1.821 0)
Fluctuation of perceived difficulty^2	−12.0051***	(1.000 1)	−10.7568***	(1.010 6)	−4.9887***	(1.144 3)	−6.8029***	(1.056 1)
Reward ads	12.6848**	(5.357 2)	2.0465**	(1.006 9)	1.7111	(2.651 2)	12.1058**	(5.115 0)
Perceived difficulty	−1.1747	(3.380 9)	0.7819	(3.733 2)	−1.4395	(5.217 0)	4.2342***	(1.303 8)
Perceived achievement^{-1}	−5.1737***	(1.794 6)	−8.3179***	(1.003 5)	−22.6904***	(1.034 0)	−9.4016***	(2.394 1)
Cumulative playing time	5.2490***	(1.405 1)	3.8748	(3.108 1)	4.6714**	(1.994 2)	2.1019	(2.369 7)
Perceived difficulty * Reward ads	3.9894***	(1.187 1)	2.9123***	(1.004 7)	3.3563***	(1.044 5)	6.1814***	(1.043 4)
Perceived achievement^{-1} * Reward ads	−3.5241***	(1.019 3)	−4.0928***	(1.008 2)	−5.0130***	(1.003 0)	−3.8104***	(1.001 4)
Thresholds								
State 1			2.8936**	(1.138 0)	3.1275	(3.391 1)	4.6494	(1.850 6)
State 2	−2.6589**	(1.213 1)			−2.4257	(1.070 5)	1.9262	(0.963 3)
State 3	−0.5492	(1.033 9)	−0.3320	(1.594 3)			10.8310**	(1.012 8)
State 4	−0.1619	(1.038 5)	0.4715	(1.194 0)	0.4818	(3.280 4)		
Log Likelihood				-20319.84				
Individuals				1232				
Observations				22325				

Note: The following rescaling is performed: Fluctuation of perceived difficulty is mean-centered and divided by 10. Perceived difficulty and cumulative playing time (in minutes) are scaled down by a factor of 1000. The perceived achievement, reward ads, latest intensity, historical intensity and irregularity are scaled down by a factor of 100. Puzzle difficulty, coin balance, soft currency purchase are scaled down by a factor of 1000.

The standard error is reported in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.8 Statistics for Testing Coefficients Difference

I check whether the estimated coefficients vary across states and whether the differences in coefficients are statistically significant. The asterisk symbols in Table A.9 denote that the coefficients between two engagement states are statistically different. For example, I test whether the effect of reward ads in low engagement state is significantly different from that in medium engagement state, and the t-test statistics is 3.66, which is significant at the 0.01 level.

Table A.9: Statistical Tests to Compare Differences in Coefficients

T statistics	L to M	L to H	M to H
Fluctuation of perceived difficulty	−0.53	0.00	0.53
Fluctuation of perceived difficulty^2	−2.25**	−0.38	1.89*
Reward ads	3.66***	1.57	−2.68***
Perceived difficulty	5.39***	0.46	−4.54***
Perceived achievement^{-1}	10.80***	3.64***	−8.11***
Cumulative playing time	5.75***	7.51***	2.46**
Perceived difficulty * Reward ads	2.58***	0.65	−1.95*
Perceived achievement^{-1} * Reward ads	1.29	−1.04	−2.33**
Intercept to play	−8.81***	2.44**	4.99***
Coin balance	6.50***	−0.19	−6.72***
Soft currency purchase	−2.46**	0.65	2.63***
In app purchases	5.36***	−0.54	−5.78***
Puzzle difficulty	2.63***	−2.49**	−4.31***
Frequency	−4.18***	1.62	5.59***
Historical intensity	0.21	−1.11	−1.39
Irregularity	−1.67*	3.08***	4.75***
Constant	3.97***	−2.25**	−6.56***

Note: The table reports the T-test statistics between two coefficients of two different states. Variables that are not statistically significant in all states are excluded.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix B

B.1 Network Evolution

To model network evolution over time, I track link changes by defining two networks: the formation network (y^+) and the dissolution network (y^-). $y^+_{t-1\to t}$ consists of network y_{t-1} plus the links formed from time $t-1$ to t, and $y^-_{t-1\to t}$ consists of network y_{t-1} minus the links removed from time $t-1$ to t. In the data, I observe y_{t-1} and y_t. Given the observed data, I are able to recover formation network $y^+_{t-1\to t}$ and dissolution network $y^-_{t-1\to t}$.

Table B.1 shows 4 possible transitions between nodes i and j. If there is no link between nodes i and j at both times $t-1$ and t (the first row in Table B.1), then the value of $y^+_{t-1\to t}$ is set to 0 in the formation network, indicating that there is no link formation from $t-1$ to t. However, I are not able to come to any conclusion regarding link dissolution because there is no link that can be dissolved between nodes i and j from time $t-1$ to t (denoted by - in Table B.1). If the link between nodes i and j exists at time $t-1$ but no longer exists at time t (the third row in Table B.1), then the value of $y^-_{t-1\to t}$ is set to 0 in the dissolution network, indicating that the link is dissolved from $t-1$ to t. If the link exists both at time $t-1$ and at time t (the last row in Table B.1), then the value of $y^-_{t-1\to t}$ is set to 1 in the dissolution network, meaning that the link is sustained from time $t-1$ to t. However, I cannot infer anything about formation (denoted by - in Table B.1) when $y_{t-1} = 1$, as the link between nodes i and j already exists at time $t-1$. Note that a value of 1 in $y^-_{t-1\to t}$ means that the link is sustained and that a value of 0 means that the link is dissolved. In contrast, a value of 1 in $y^+_{t-1\to t}$ means that the link is formed.

Table B.1: Network evolution

y_{t-1}	y_t	$y^-_{t-1\to t}$	$y^+_{t-1\to t}$
0	0	-	0
0	1	-	1
1	0	0	-
1	1	1	-

B.2 Estimation with Time Fixed Effects

Link formation and dissolution may be affected by some events that are time dependent, for example, some policy change on the platform, the availability of crypto currencies on the platform, or a regulatory change in the social trading industry. During the sample period (2016 and 2017), no changes were made to social trading regulations in general, and all existing regulation remained intact.

The relevant regulators classify social trading as portfolio management per the Markets in Financial Instruments Directive (MiFID). In particular, the European Securities and Markets Authority (ESMA), the EU's securities market regulator, announced in 2008 that financial operators active on a social trading network could exercise "investment discretion by automatically executing the trade signals of third parties", which implied that brokers and market-makers active in that field were assimilated into the group of other financial intermediaries that need ad hoc authorization for portfolio management per the Markets in Financial Instruments Directive (MiFID). Subsequently, the same authorization requirements were confirmed by ESMA in 2012. Whenever a service provider makes an investment through an automated algorithm in view of trade signals coming from third parties—in relation to MiFID financial instruments—this implies that the provider has to perform some consequent duties related to a suitability assessment, the completion of business obligations, and information standards for both clients and authorities.

The corresponding directive, Directive 2004/39/EC, was first introduced in 2004. The main objective of MiFID is to create a European financial market that encourages honest competition between the participating companies and, at the same time, increasing consumer protection. MiFID has been in force since January 31, 2007 and was superseded by MiFID II on January 3, 2018. MiFID II did not bring meaningful changes to the regulation of social trading.

Similarly, the Financial Conduct Authority (FCA) specified that the service of social trading falls within Article 4(1)(9) of MiFID. This article defines "portfolio management" as "managing portfolios in accordance with mandates given by clients on a discretionary client-by-client basis where such portfolios include one or more financial instruments." In copy trading and mirror trading, investment decisions are implemented with no intervention by the client other than an agreement ("mandate") between the service provider and the client on the discretionary service provided.[1] This interpretation has not changed over time.

Despite the lack of regulatory changes during the sample period, it is possible that link formation and dissolution are affected by other events that are time dependent. To mitigate this concern, I include time fixed effects, which allows us to control for time-specific peculiarities. I estimate the extended model and present the results in Table B.2. The estimation results are generally consistent with the results from the main model.

B.3 Estimation with Leader Fixed Effects

As mentioned in Section 4.2.2, in my context, the information provided on the platform is rich and highly transparent. I observe the information that is observed by followers on the platform, which may affect followers' link formation and dissolution decisions. I have access to the complete transaction and social communication histories of each trader, and the platform does not allow for a private chat channel. In addition, given the size and international reach of the platform, I believe that personal relationships between investors are very unlikely and affect at most very few investors. The data include 106,007 unique

[1] `https://www.fca.org.uk/firms/copy-trading`, last accessed Jul. 8, 2022

traders of more than 140 nationalities. In addition, only approximately 14.6 percent of the total links are between members of the same nationality. Considering this large size and international reach, it is unlikely that (many) investors know each other outside of the platform.

However, some investors may decide to advertise their trading via other social media channels such as YouTube and include links to their eToro profile page on YouTube. Despite the various control variables included in my network analysis, some determinants of link formation and dissolution may be unobserved, at least to researchers. For example, it is possible that leaders' advertising activities could be found on other social media sites, which may affect followers' following decision. However, due to the anonymity of the data, I are unable to match data from other social media sites to the trading data. To mitigate this concern, I include leader fixed effects to capture a leader's general propensity to advertise their trading on social media. I believe that such general propensities are rather stable over time. Some investors are, in general, willing to advertise their investment strategies on alternative social media, while others are not.

I model only the link formation process because leader advertisements on other channels are more likely to affect the link formation process. I estimate the model using the same Chamberlain approach described in Section 4.2.2. To guarantee computational tractability with two random effects (follower-specific and leader-specific), I keep all leaders and randomly sample 250 followers from the sample used for the main model. The results are reported in Table B.3, and my main findings hold qualitatively.

Table B.2: Estimation Results with Time Fixed Effects

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.2653***	(0.027 1)	0.1090***	(0.029 0)
Leader's post quality	0.5212***	(0.025 0)	0.0510**	(0.021 3)
Leader's number of replies	0.0971***	(0.028 2)	0.0140	(0.031 7)
Leader's comment received positive score	1.2130***	(0.142 0)	0.4852***	(0.148 9)
Leader's comment received negative score	−5.2520***	(0.778 1)	−1.9831***	(0.479 2)
Follower's post quantity	0.0511	(0.049 3)	−0.5559***	(0.055 6)
Follower's post quality	−0.3151***	(0.081 3)	0.0066	(0.077 4)
Financial performance				
Leader's average profit	0.1069***	(0.011 2)	0.0505***	(0.018 7)
Leader's std. dev. profit	−0.3588	(0.451 6)	−1.5600**	(0.780 2)
Follower's average profit	−0.0028	(0.010 6)	0.0551***	(0.012 6)
Follower's std. dev. profit	−0.2900	(0.503 2)	−8.7381***	(0.596 1)
Leader's average holding time	0.2492***	(0.045 4)	0.0587	(0.068 2)
Leader's lottery preference	0.6185*	(0.355 5)	1.0449***	(0.401 9)
Leader's HHI	−1.0280***	(0.092 9)	−0.3822***	(0.096 7)
Demographics				
Nationality	0.7866***	(0.070 6)	0.3577***	(0.081 0)
Age	0.1132**	(0.050 9)	0.0582	(0.055 2)
Homophily (male)	0.9514***	(0.136 2)	−0.1272	(0.118 9)
Homophily (female)	−0.8487**	(0.417 1)	0.2573	(0.320 9)
Image	2.3251***	(0.425 8)	0.9411*	(0.483 2)
Bio	2.8937***	(0.193 1)	−0.1689	(0.193 7)
Experience	−0.0724*	(0.038 0)	0.1277***	(0.044 3)
Wealth	0.0120	(0.033 2)	0.0666*	(0.038 1)
Income	0.0038	(0.040 1)	0.0276	(0.044 9)
Risk	−0.0809	(0.049 8)	−0.0252	(0.057 9)
Network structure				
Leader's popularity	0.0110***	(0.000 4)	0.0017***	(0.000 4)
Leader's activity	−0.0401***	(0.012 5)	0.0312***	(0.011 9)
Follower's popularity	−0.0570***	(0.013 0)	−0.0084	(0.007 2)
Follower's activity	−0.0047	(0.007 8)	−0.1032***	(0.009 7)
Transitivity	0.0856***	(0.025 5)	−0.0256	(0.066 5)
Constant	−1.5666	(6.033 5)	2.9661***	(0.696 8)
Time fixed effects	Yes		Yes	
Log Likelihood	-13,039.84		-9,000.62	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table B.3: Estimation Results with Leader Fixed Effect

Variable	Formation	
Social communication		
Leader's post quantity	0.0936	(0.081 9)
Leader's post quality	0.1600**	(0.069 2)
Leader's number of replies	0.1521*	(0.090 7)
Leader's comment received positive score	1.1432***	(0.336 5)
Leader's comment received negative score	−7.1147***	(1.947 7)
Follower's post quantity	0.1544***	(0.053 6)
Follower's post quality	−0.6185***	(0.106 8)
Financial performance		
Leader's average profit	0.1196***	(0.039 3)
Leader's std. dev. profit	−4.2705**	(1.827 8)
Follower's average profit	−0.0615***	(0.019 5)
Follower's std. dev. profit	−0.9975	(1.168 5)
Leader's average holding time	0.0713	(0.168 0)
Leader's lottery preference	−0.7884	(0.984 3)
Leader's HHI	−0.1140	(0.274 4)
Demographics		
Nationality	1.5814***	(0.161 9)
Age	0.1193	(0.123 4)
Homophily (male)	0.3063	(0.362 0)
Homophily (female)	−0.3198	(0.820 4)
Image	1.6256***	(0.625 1)
Bio	0.9437***	(0.347 3)
Experience	−0.0871	(0.063 4)
Wealth	0.1410**	(0.056 7)
Income	0.0263	(0.065 4)
Risk	−0.1986**	(0.085 1)
Network structure		
Leader's popularity	0.0001	(0.001 5)
Leader's activity	−0.0371	(0.032 0)
Follower's popularity	−0.0722***	(0.014 8)
Follower's activity	0.1165***	(0.012 3)
Transitivity	0.1580***	(0.031 2)
Constant	27.8232**	(10.900 7)
Log Likelihood	-3,121.92	
Observations	1,592,411	

Notes: The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

B.4 Alternative Data Sample

I estimate the main model using a random subsample of all users on the platform. In this appendix, I take another random sample and estimate the main model to address a potential concern about the external validity of the findings. The results are reported in Table B.4, and the findings are generally consistent with those from the main model.

B.5 Examples of Social Communication Texts

I provide some examples of posts, comments, and replies in Table B.5. As shown in the table, the example posts were written by leaders. Leaders may want to advertise their performance, share their trading strategies or simply welcome new followers. Followers can ask for clarification in the comments (see comment 1), provide positive feedback to the leader (see comment 2), or complain about the leader's performance by leaving a negative comment (see comment 3). Finally, a leader can reply to a follower's question (see reply 1) or share his/her insights about the market by replying to the comment (see reply 2).

B.6 Estimation with Post Quality-Quantity Interaction Term

In this appendix, I add an interaction term between the post quantity and post quality of a leader, keeping everything else the same as in the main model, to investigate the moderating effect of post quality. I estimate the extended model and show the results in Table B.6. I find that post quantity has a positive moderating effect on post quality in the link formation process, whereas the interaction term is statistically not significant in the link dissolution process.

Table B.4: Estimation Results with Alternative Sample

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3015***	(0.026 7)	0.1025***	(0.027 6)
Leader's post quality	0.5340***	(0.024 6)	0.0637***	(0.020 4)
Leader's number of replies	−0.0024	(0.027 1)	−0.0090	(0.029 7)
Leader's comment received positive score	1.0002***	(0.140 4)	0.4649***	(0.145 6)
Leader's comment received negative score	−3.8560***	(0.686 2)	−2.8281***	(0.455 4)
Follower's post quantity	−0.0794*	(0.044 7)	−0.3592***	(0.048 9)
Follower's post quality	−0.2615***	(0.081 8)	−0.1304*	(0.073 5)
Financial performance				
Leader's average profit	0.0898***	(0.011 8)	0.0429**	(0.016 7)
Leader's std. dev. profit	−2.5028***	(0.488 9)	−1.2995*	(0.694 0)
Follower's average profit	−0.0252	(0.015 5)	0.0938***	(0.014 9)
Follower's std. dev. profit	−1.3589**	(0.632 7)	−8.5795***	(0.587 5)
Leader's average holding time	0.3524***	(0.042 3)	−0.0783	(0.061 9)
Leader's lottery preference	0.1861	(0.345 3)	0.7816**	(0.387 6)
Leader's HHI	−0.5040***	(0.086 3)	−0.3184***	(0.087 1)
Demographics				
Nationality	0.6705***	(0.066 9)	0.1022	(0.078 3)
Age	0.0134	(0.050 2)	−0.0408	(0.053 9)
Homophily (male)	1.2767***	(0.148 8)	−0.0717	(0.115 2)
Homophily (female)	−1.5970**	(0.713 6)	0.4126	(0.416 7)
Image	2.6399***	(0.423 9)	1.0131**	(0.500 9)
Bio	2.6410***	(0.157 3)	0.0152	(0.179 7)
Experience	0.0277	(0.038 5)	0.0780*	(0.039 9)
wealth	−0.0128	(0.033 3)	0.1109***	(0.034 4)
Income	0.0056	(0.039 0)	−0.0138	(0.040 5)
Risk	−0.0513	(0.048 7)	−0.1273**	(0.051 3)
Network structure				
Leader's popularity	0.0089***	(0.000 4)	0.0016***	(0.000 4)
Leader's activity	−0.0479***	(0.012 5)	0.0283**	(0.011 7)
Follower's popularity	−0.0489***	(0.011 9)	−0.0033	(0.005 7)
Follower's activity	0.0036	(0.005 9)	−0.0147**	(0.006 2)
Transitivity	0.0825***	(0.023 1)	0.0439	(0.060 9)
Constant	−18.7072***	(4.972 9)	1.3530**	(0.674 5)
Log Likelihood	-14,202.79		-9,541.34	
Observations	11,000,237		20,796	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table B.5: Examples of Social Communication

Post	**Text**
1	After being very long out I made some big mistakes. Now I trusted my instinct and made every trade without the knowledge from someone else. With this I turned back in the green and closed 2016 in green! aIsome
2	The $EURUSD is a buy for me for the next few months... been at its lowest in years lately & I will definitely be looking for buy positions
3	@adelaya Hi, thank you for following. I wish you happy successful trading :)
Comment	**Text**
1	Why does it seem to fall in the after hours charts, though? Shouldn't it surge? I'm a newbie.
2	Your current investments are looking aIsome!
3	I has lost so much money since i copied u and i never taste the earning sIet
Reply	**Text**
1	@Seregaomsk Additional funds will be used only when I open new deals, to distribute them to open positions while the only option is to stop copying and then copy with a new amount.
2	The beginning of the fall was undoubtedly connected with the general correction in the market, after which they are still only recovering. The current fall, in my opinion, is largely speculative, since there have been no negative indicators, news or decisions regarding YNDX lately. In this regard, I plan to keep them for now, as I look forward to recovery in the coming weeks.

Notes: Some social communications were not originally in English, and thus, I present the translated versions.

Table B.6: Estimation Results with Post Quality-Quantity Interaction Term

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.2942***	(0.027 9)	0.1226***	(0.028 8)
Leader's post quality	0.4140***	(0.035 6)	0.0678**	(0.028 3)
Leader's number of replies	−0.0263	(0.031 0)	0.0037	(0.033 1)
Leader's comment received positive score	1.2167***	(0.141 1)	0.5555***	(0.146 6)
Leader's comment received negative score	−4.6982***	(0.761 1)	−2.1387***	(0.471 9)
Follower's post quantity	0.0657	(0.047 4)	−0.5301***	(0.053 6)
Follower's post quality	−0.3339***	(0.079 7)	−0.0049	(0.074 9)
Leader's post quality * leader's post quantity	0.0636***	(0.016 6)	−0.0184	(0.014 9)
Financial performance				
Leader's average profit	0.0891***	(0.013 5)	0.0729***	(0.018 4)
Leader's std. dev. profit	−1.8736***	(0.555 1)	−2.7737***	(0.768 7)
Follower's average profit	0.0013	(0.011 4)	0.0441***	(0.012 0)
Follower's std. dev. profit	−0.6637	(0.509 0)	−8.1489***	(0.558 0)
Leader's average holding time	0.3451***	(0.044 1)	0.0615	(0.066 1)
Leader's lottery preference	0.1675	(0.363 0)	1.0977***	(0.395 7)
Leader's HHI	−0.6133***	(0.090 6)	−0.1403	(0.092 9)
Demographics				
Nationality	0.7466***	(0.070 4)	0.3440***	(0.079 8)
Age	0.1235**	(0.050 5)	0.0514	(0.054 3)
Homophily (male)	0.9849***	(0.136 0)	−0.1562	(0.116 5)
Homophily (female)	−0.8621**	(0.416 7)	0.2938	(0.317 8)
Image	2.4808***	(0.423 6)	0.7285	(0.465 2)
Bio	2.9513***	(0.187 3)	−0.1537	(0.188 6)
Experience	−0.0445	(0.037 1)	0.1459***	(0.040 4)
Wealth	0.0244	(0.032 4)	0.0617*	(0.034 7)
Income	−0.0091	(0.039 3)	0.0240	(0.040 9)
Risk	−0.0838*	(0.048 8)	−0.0397	(0.052 7)
Network structure				
Leader's popularity	0.0078***	(0.000 4)	0.0015***	(0.000 4)
Leader's activity	−0.0354***	(0.012 4)	0.0379***	(0.011 9)
Follower's popularity	−0.0544***	(0.012 7)	−0.0091	(0.006 5)
Follower's activity	0.0026	(0.007 1)	−0.0771***	(0.008 8)
Transitivity	0.0810***	(0.025 7)	−0.0206	(0.065 6)
Constant	−16.0395***	(5.662 2)	2.0885***	(0.644 4)
Log Likelihood	-13,613.02		-9,151.50	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

B.7 Alternative Estimation Using Conditional Logit Estimator

The Chamberlain correlated random effects applied in the main model require the assumption that η_i follows a conditional normal distribution depending on X_i with a constant variance, which is equivalent to a conventional random effects model that controls for the correlation function. Thus, the coefficients on the time-invariant observables (C_i) can be estimated. To correct for the incidental problem, another way to estimate the follower-specific unobservables η_i is based on the conditional logit estimator (Wooldridge 2010), which allows η_i to be arbitrarily correlated with X_i. However, under the conditional logit estimator, η_i and time-invariant covariates C_i cannot be identified simultaneously. Thus, C_i should be excluded.

I use the link formation process to illustrate the model implementation. The utility of follower i from forming a link with leader j between period $t-1$ and period t is defined as follows:

$$y^*_{ijt} = \boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij} + \eta_i + \epsilon_{ijt}, \tag{B.7.1}$$

where the definitions of notations are the same as those in the main model, Equation (4.6).

For the link formation process, I define

$$y_{ijt} = \begin{cases} 1 & y^*_{ijt} > 0 \\ 0 & \text{otherwise}, \end{cases} \tag{B.7.2}$$

where y_{ijt} is a binary variable that is equal to 1 if follower i forms a link with leader j between period $t-1$ and period t. The link dissolution process is defined in the same fashion. In the dissolution process, y_{ijt} is equal to 1 if follower i dissolves the link with leader j in period t.

I denote as n_i the sum of all binary outcomes for follower i's following status over all the periods. That is, $n_i = \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} y_{ijt}$, where T_i is the number of periods during which follower i exists on the platform and J_{it} is the number of leaders that follower i can potentially follow in period t. Each follower i has a corresponding vector with length $T_i \times J_{it}$. B_i is the set of all possible vectors in which n_i elements are equal to 1 and $(T_i \times J_{it} - n_i)$ elements are equal to 0. In other words, B_i represents all the possible scenarios in which follower i forms

n_i links with the potential J_{it} leaders over T_i periods. Mathematically,

$$B_i = \{b \in \{0,1\}^{\{T_i \times J_{it}\}} | \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} b_{jt} = n_i\}, \tag{B.7.3}$$

where b is one realization or scenario among all the possible scenarios and b_{jt} denotes an element in the vector b.

The conditional probability of y_i given n_i is defined as follows:

$$Pr(y_i | X_{it-1}, W_{jt-1}, n_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{e^{(y_i \times (\boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij}))}}{\sum_{b \in B_i} e^{(b \times (\boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij}))}}. \tag{B.7.4}$$

From Equation (B.7.4), I observe that the conditional probability does not depend on η_i. Thus, the conditional log likelihood is also independent of η_i and can be written as follows:

$$CLL(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{i=1}^{I_t} \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} \ln[Pr(y_i | X_{it-1}, W_{jt-1}, n_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})], \tag{B.7.5}$$

where I_t is the total number of followers in period t, T_i is the number of periods that follower i exists on the platform, and J_{it} is the number of leaders that follower i can potentially follow in period t. I estimate the model by maximizing its overall log-likelihood value. The estimation results are reported in Table B.7. I find that the results are consistent with the findings in the main model.

Table B.7: Estimation Results Using Conditional Logit Estimator

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3184***	(0.026 8)	0.1167***	(0.028 4)
Leader's post quality	0.5167***	(0.025 0)	0.0438**	(0.020 8)
Leader's number of replies	0.0211	(0.027 7)	−0.0114	(0.031 2)
Leader's comment received positive score	1.1088***	(0.139 5)	0.6159***	(0.148 5)
Leader's comment received negative score	−4.8155***	(0.754 4)	−2.0508***	(0.472 0)
Follower's post quantity	0.0405	(0.062 4)	−0.4604***	(0.067 1)
Follower's post quality	−0.2816***	(0.087 4)	0.0585	(0.080 0)
Financial performance				
Leader's average profit	0.0906***	(0.013 6)	0.0734***	(0.018 5)
Leader's std. dev. profit	−2.1217***	(0.554 3)	−2.7319***	(0.765 1)
Follower's average profit	0.0298	(0.019 7)	0.0085	(0.019 6)
Follower's std. dev. profit	−3.9422***	(1.273 8)	−8.2172***	(1.279 6)
Leader's average holding time	0.3466***	(0.044 6)	0.0774	(0.067 3)
Leader's lottery preference	0.2098	(0.361 3)	1.1073***	(0.398 9)
Leader's HHI	−0.5972***	(0.090 4)	−0.1268	(0.093 5)
Demographics				
Nationality	0.7418***	(0.070 7)	0.3390***	(0.079 4)
Age	0.1198**	(0.050 7)	0.0548	(0.054 4)
Homophily (male)	0.9848***	(0.135 9)	−0.1385	(0.117 0)
Homophily (female)	−0.8769**	(0.416 6)	0.2625	(0.312 1)
Image	2.4737***	(0.423 9)	0.4781	(0.542 4)
Bio	2.9224***	(0.186 9)	−0.2559	(0.200 8)
Network structure				
Leader's popularity	0.0077***	(0.000 4)	0.0014***	(0.000 4)
Leader's activity	−0.0359***	(0.012 4)	0.0392***	(0.012 1)
Follower's popularity	−0.0364**	(0.014 7)	−0.0060	(0.008 7)
Follower's activity	−0.0282***	(0.007 4)	−0.0849***	(0.010 4)
Transitivity	0.0927***	(0.029 0)	−0.0218	(0.063 5)
Log Likelihood	-11,339.00		-5,522.04	
Observations	4,074,702		15,083	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.

Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.