What is the Value of Free Players and Free Goodies in Online Video Games? An Empirical Study of Microtransactions

Hai Che Cheng Chou Peng Liu*

Preliminary draft. Do not circulate or cite.

Abstract

Massively multiplayer online (MMO) games usually adopt free-to-play business model, in which companies earn profits from gamers' in-game purchases (commonly known as microtransactions). We develop a structural model of a player's simultaneous choice of the hours spent on playing the game and the consumption of in-game items with unobserved player heterogeneity in the opportunity cost of time, price sensitivity, and taste of the game, and with social interactions within the gaming community. We prove the completeness and coherency of the model. Using player-level panel data, we show the identification of all structural parameters specifying a paid player's preference and propose a novel estimation method. Our estimates show that without the participation of free players, the average in-game expenditure among paid players will drop by 58% in our specific MMO game. We propose and evaluate a counterfactual reward program, in which players will be rewarded virtual goods for free after playing the game for certain hours.

Keywords: free-to-play, microtransaction, massively multiplayer online games, social interaction

1 Introduction

Many online video games adopt a microtransaction business model (also called "free-to-play" or F2P model), in which the game is free and the players can purchase optional virtual items to enhance their gaming experience. Game companies earn profits from players' in-game purchases in

*Hai Che is an Associate Professor of Marketing at School of Business, University of California Riverside, USA. Email: chehai@ucr.edu. Cheng Chou is an Assistant Professor of Economics at School of Business, University of Leicester, UK. Email: cchou@le.ac.uk. Peng Liu is an Assistant Professor of Marketing at Leavey School of Business, Santa Clara University, USA. Email: pliu2@scu.edu. Authors are listed in alphabetical order.

F2P model. F2P model is prevalent in massively multiplayer online (MMO) games using free entry to accumulate a large players base, which is essential for the success of an MMO game. In 2019, free-to-play MMO games generated $87.1 billion revenue worldwide, accounting for 80 percent of digital game revenues.[1] The COVID-19 pandemic has contributed to substantial increases both in gamer population as well as in time spent on playing MMO games.[2]

According to a survey of Fortnite (the highest gross revenue F2P video game in 2019) users, 71% of Fortnite players spent at least 6 hours each week on playing the game, 69% of the players spent money on in-game purchases (with an average total spending $84.67), while the rest 31% never spent even if they are active players.[3] For the specific game we are studying, only 22% of the active players paid for in-game items, while the rest 78% of players did not make any in-game purchases for the three months data duration. Why does anyone spend real dollars on something that is totally virtual when playing is free? What is the economic value of these free players who never spent a dime or are not spending in the game to a game platform? Can companies increase sales if players are rewarded with free virtual goodies for their participation and long hours played?

To answer these questions, we develop a structural model of a player's simultaneous choice of the hours spent on playing the game and the consumption of in-game items. Our model accommodates the unobserved player heterogeneity in the opportunity cost of time, price sensitivity, and taste of the game. We also model the social interaction effect—individual player's utility of playing the game depends directly on the participation choices of other players (free or paid players) in the gaming community, which is a crucial feature of MMO games. We prove the completeness and coherency of the model and show the identification of all structural parameters specifying a paid player's preference. Particularly, for each paid player, we can identify and estimate her opportunity cost of time, price sensitivity, and taste of the game. These parameters are valuable to managers in pricing and micro-management of each player and her gaming experience. For free players, who never spent on microtransactions, we cannot identify their price sensitivity (unless very strict assumptions are imposed), and we cannot separately identify their opportunity cost of

[1] "2019 Year In Review: Digital Games and Interactive Media", Superdata, Nielsen, Accessed December 30, 2020, https://www.superdataresearch.com/reports/p/2019-year-in-review.

[2] For instance, Verizon reported an 82% increase in video game traffic over pre-COVID levels.A recent global survey conducted by Statista shows 60% of respondents stated that they were playing more MMO during the pandemic.

[3] Mansoor Iqbal, "Fortnite Usage and Revenue Statistics", Business of Apps, Updated December 5, 2020, https://www.businessofapps.com/data/fortnite-statistics/, and Mike Brown, "The Finances of Fortnite: How Much Are People Spending on This Game?", lendedu, June 26, 2018, https://lendedu.com/blog/finances-of-fortnite/.

time and taste of the game (instead we only know their difference). To estimate the model, we propose a novel sequential estimation method that is easier and numerically more stable than the conventional maximum likelihood estimation method, which is difficult to implement due to the presence of corners solutions of hours played and in-game consumption and the large number of player activity observations.

We estimate the proposed model using player-level panel data obtained from a large game company in China. The data contain daily player-level gaming activities between January and March 2011 (81 days in total). After exploring players' daily activities data and conducting reduced-form analyses, we find that empirically hours played and in-game consumption are complementary to each other, hence ignoring one of them will cause bias in estimating consumers preference. Thus, it is essential to jointly model the decisions of hours played and in-game consumption. However, modeling the two decisions simultaneously is difficult. Under very weak conditions, we prove that there are multiple equilibria when a player maximizes her utility with respect to the hours played and in-game consumption simultaneously. So certain equilibrium refinement is necessary, and we propose an intuitive selection rule to address this issue.[4] Both unobserved heterogeneity and social interaction effect are particularly important for the study of MMO games adopting a F2P business model. Since the game is free, it is expected that the population of players are very heterogeneous as illustrated by the distribution of the hours played and in-game consumption in our data (see Figure 1). Our long player-level panel data provide the opportunity of identifying the unobserved player heterogeneity treated as various fixed effects in our model. Using players' decisions about hours played and in-game consumption, we can separately identify these unobserved heterogeneity. For example, one's price sensitivity should not affect her hours played given her in-game purchases. Social interaction is another important issue given that many activities in MMO games are simply unplayable without an active gaming community. For example, certain complex tasks require players with similar skill level to cooperate or compete with each other. To model social interaction effect, we let individual player's subjective quality of playing the game depend on

[4] We assume that players follow the following sequential rule of decisions. First, she decides whether or not it is optimal to play the game at all (positive hours played as opposed to zero hours played). If she found zero hours played is optimal, she does not play the game and does not spend on microtransactions. Otherwise, she moves on to decide whether or not it is optimal to make any in-game purchase (positive amount spent as opposed to no purchase at all). If she found that no spending is optimal, she does not spend and play the game until her marginal utility of playing the game equals the marginal cost, which is the unit cost of leisure. If she found that positive spending is optimal, she purchase in-game items and play the game to maximize her utility.

the participation of other players, among whom the majority are free players, in addition to other player's characteristics. The dependence of one player's quality on the participation of other players reflects the positive network externalities (Katz and Shapiro, 1985; Brock and Durlauf, 2001).

Results from the model estimation reveal how free players' presence and their participation could be beneficial to the paid players and improve the game company's revenue and profits through the channel of social interaction effect. In our data, 78% of the active players did not make any purchases even if they were active players during the three-month study period. Understanding the value of free players in online video games, especially those with both free and paid players, is valuable for game companies and marketing researchers. Despite the potential value, little is known from academic research on how to leverage the presence of free players in MMO games. Our estimation results show that if the current free players did not play the game at all, the average spending on microtransactions among paid players would have dropped by 58.1% and their average hours played would have dropped by 61.4%. On the other hand, if the game participation rate among the free players (the average daily participation rate is 35.8%) was the same as the rate among the paid players (50.0%), the amount spent by paid players would have increased by 6.8% percent, the average hours played among all players would have increased by 31.4%.

Given the great value of the free players shown in our results, it is useful to consider some counterfactual programs that could improve the engagement of gamers, particularly free players. We propose and evaluate a counterfactual reward program, in which any players can obtain virtual items for sale by just playing the games for certain hours. The amount of the reward can be managed by the game company.[5] Our counterfactual analysis shows that game companies must be strategic in the generosity of such reward program in order to maximize its impact on the sales of in-game items. When the reward program is too generous, players will just obtain in-game items by playing the game rather than purchasing, and the game can be a hit without being profitable. Our counterfactual analysis could help guide managers to design their reward program to find the optimal generosity.

In section 2, we discuss the literature. Section 3 presents the institutional background information regarding MMO games, focusing on salient data patterns to guide our model specification. We then develop a structural model of players' choices about playing video games with interactions

[5] In the gaming community, players usually spend hours on "farming" certain weapons or resource to upgrade their characters. The efficiency of farming is determined by the game company, and such farming scenario is similar to our counterfactual reward program.

between free and paid players in section 4, and present our estimates about the value of free players. In section 5, we prove the identification of our structural model, and supplement the estimation details. In section 6, we study a counterfactual reward program, in which players can farm for in-game items. We then conclude the paper in section 7. The additional technical details including the proofs of theorems are in the appendix.

2 Literature Review

Our study is related to three streams of literature. First, the paper is related to an emerging stream of research on the freemium business model of digital products (Cheng and Liu, 2012; Lee and Tan, 2013; Foubert and Gijsbrechts, 2016; Hoang and Kauffman, 2018).In particular, Gu, Kannan and Ma (2018) conducted a randomized field experiment with an online content provider that offers book titles in a PDF version for free and sells the paperback version for a premium. They find that book titles with a free PDF version lead to more sales than those that do not and identify the attraction effect and the compromise effect. Lee, Kumar and Gupta (2019) study a freemium cloud storage product and focus on understanding the effect of a referral program on consumer's usage and upgrade decisions. They find that while a referral program reduces the number of consumers who would upgrade, it makes up the loss by bringing in many new consumers who will potentially upgrade. Shi, Zhang and Srinivasan (2019) conduct a theoretical analysis of the freemium model. Adopting the screening model framework, they found a freemium business model is an equilibrium strategy when network externality exists and when the premium product provides its consumers with larger utility gains from the firm's user base. Our paper adds to this literature by providing an econometric modeling framework to study the effects of a large free-player user base on the paid user's participation and consumption decisions. Following Shi, Zhang, and Kannan's theory model, we allow two types of players who consumer the free and premium product. We also allow the players to have different quality valuation by modeling both observed and unobserved heterogeneity. After obtaining the structural model estimates, we demonstrate through counterfactual experiments that while a high percentage of video game players who did not ever purchase during the sample period, their participation in the game encourages paid consumers to participate and spend more in the game.

Second, our paper relates to the economics literature on network externality models. In the context of online video games, the size of the player base may exert strong direct positive network effects on individual players. Katz and Shapiro (1985) define network effects as "the value of (a) membership to one user (which) is positively affected when another user joins and enlarges the network". They also describe the source of positive consumption externalities as the utility that a user derives from consumption of goods, which increases with the number of other agents consuming the (same) good. Our empirical model framework follows these definitions. In the empirical literature, many econometric models of social interactions based on the network externality theory have been developed (see Durlauf and Ioannides (2010) as well as Kline and Tamer (2020) for reviews of this literature). Among them, Brock and Durlauf (2001) developed a binary choice model of social interaction by incorporating terms reflecting individuals' desire to conform to others' behavior. They show that the social utility component represents a version of the binary choice models with externalities. Lee, Li and Lin (2014) extend Brock and Durlauf (2001) binary choice complete network (or social interaction) model with homogeneous rational expectations to a general network model with heterogeneous rational expectations.

Finally, our paper is related to the marketing literature on the online video game market. Many of the existing studies on the video game market Nair (2007); Dubé, Hitsch and Chintagunta (2010); Ishihara and Ching (2019) focus on consumer's decision of purchasing and upgrading the video games. Nevskaya and Albuquerque (2019) propose a continuous-time demand model to study players' excessive usage (playing) behavior. Through counterfactual experiments, they found improving reward schedules and imposing time limits lead to shorter usage sessions and longer subscription periods. Jo et al. (2020) adopt a causal inference model to study the effects of a usage regulation policy on gamer behavior and found usage restriction laws may deter light gamers from potentially excessive gaming but do not work well to dissuade heavy gamers. Our study differs from the above studies in that we zoom in on the role of free players in video games by studying the effects of their presence on the paid players' participation (play or not), playing time, and consumption behaviors. Game companies could potentially use the behavioral primitive estimates obtained from our proposed structural model to improve their game feature designs, thus encouraging higher levels of participation and consumption.

3 Online Video Games and Data

We first introduce some institutional details about the online video game studied here. We then discuss the data and point out important observations from reduced form estimation that will guide our model specification.

3.1 Massively Multiplayer Online Game in This Paper

The game we are studying is a massively multiplayer online role-playing game launched by a leading game studio in China in July 2010. It is a free-to-play game, as any player can download and play the game for free. The only revenue for the game company comes from players' purchase of in-game items, such as weapons and costumes. These in-game consumption directly improves players' gaming experience by allowing players to perform better, completing their tasks more easily etc. However, in-game purchases are not necessary to level up or to participate in various activities in this game. Depending on the expiration dates of in-game items, game company charges different prices. There are three different expiration dates: seven days, thirty days, and ever-lasting. Most items without expiration dates costs more than ¥30 RMB (about $4.5 USD). In Table 1, we show the distribution of sales of different in-game items on the first day in sample categorized by their prices. This table shows that 66.5% of the in-game items sold are non-durable goods.

Social interaction is an essential feature of our game. There are two playing modes in the game: player-versus-environment (PVE) and player-versus-player (PVP).[6] In both modes, players complete adventure missions while interacting with a large number of players in a virtual world. Certain complex tasks require players to cooperate with each other in a team. Players usually form teams with random players by a matchmaking algorithm, guild members, and online friends.[7] Such a massively multiplayer setting created a vibrant gaming community. Individual player's gaming experience, or what we will call "subjective quality" of gaming in our model, depends on whom she teams with, which ultimately depends on the population of the active players in her gaming session. For example, a player who logs in during conventional working/school hours may have limited availability of players to match.[8] Unbalanced games due to bad matchmaking is detrimental for gaming experience. Although free players do not contribute directly to the revenue of the game company via in-game purchases, active free players generate positive indirect network

[6] In the PVE mode, players complete missions, fight monsters and villains in instanced dungeons, and accumulate experience points. In the PVP mode, players practice and improve game skills in group or one-on-one combats.

[7] In our game, players can make friends and join player guilds. A guild is a player-initiated organization that helps facilitate players to collaborate and compete with one another.

[8] In the game, there are four different types of classes: warrior, archer, sorceress and cleric. Players can choose their character avatars and interact with each other. The matchmaking algorithm matches players based on characters levels, skills and internet connections.

Table 1: In-Game Items Sales Percentage on the First Sampled Day

Price (¥RMB)	Unit of Sales	Percentage (%)
≤ 10	10807	31.53
10–20	3839	11.20
20–30	8142	23.76
30–40	1007	2.94
40–50	4761	13.89
50–60	413	1.20
60–70	107	0.31
70–80	82	0.24
80–90	2235	6.52
90–100	1594	4.65
> 100	1287	3.76

Note: ¥1 RMB is about $0.15 USD.

externality creating an active gaming community, which has value for the platform. A quote from a game "whale" (big spenders in online and social games) who spent $20,000 in the past five years in response to an interview question "Is that important to you when you are picking games to play, having a very active community?", highlights how avid gamers view active gaming community in their decisions about in-game consumption.

> Finding a community is really important because it tells me the game has a healthy heartbeat and there is going to be a reason to invest my dollars into something that is totally intangible and can totally poof at any moment.[9]

Below we will first provide some "reduced-form" evidence supporting this observation, then we will develop and estimate a structural model of players choice of hours playing the game and in-game consumption in the presence of social interaction effect.

3.2 Data

The data set used in this study is a panel data of 35,810 players during the period January 10, 2011–March 31, 2011 (81 days). The data consist of players' daily time spent on gaming (measured in

[9] "Interview with the Video Game Whale: Splashing the Cash", Eurogamer, accessed December 25, 2020, `https://www.eurogamer.net/articles/2017-10-20-interview-with-the-whale`.

Table 2: Data Summary

	Free Players[1]	Paid Players
Average Daily Gaming Hours	0.51	0.87
	(0.53)	(0.69)
Average Daily In-Game Spending (¥RMB)		2.07
		(5.60)
Average Spending Among Friends (¥RMB)	0.38	0.95
	(2.98)	(4.44)
Average Spending Within Guild (¥RMB)	0.07	0.21
	(0.64)	(1.06)
Initial Account Tenure (Days)	62.24	70.66
	(45.56)	(52.13)
Initial Level (Max = 40)	31.69	33.57
	(4.60)	(5.04)
Percentage of Female Characters	0.26	0.39
Percentage of Players With a Guild	0.43	0.85
Number of Players	27,775	8,035
Total Number of Days	81	81

[1] Free players are those who never spent on in-game items during the entire sampling period. One player will be called a paid player if she is not a free player. Standard deviations are parentheses. ¥1 RMB is about $0.15 USD.

seconds), purchase histories at daily frequency, and various players characteristics, including login history, current level, the number of friends appearing online, gender of their characters, account tenure (days), spending by players' friends, whether or not a player belongs to a guild, and average spending within the guild. The data were provided by the anonymous game company who owns this game title.

The players in our sample are "stable" players, whose average account tenure at the beginning of our sample (January 10, 2011) is 64.17 days. We chose stable players because their preference for the game is presumably "fixed" during our sampling period. This helps control unobserved consumer heterogeneity. We excluded those players, whose hours played is above top 1 percentile, to ensure that our results reflect the behavior of a typical gamer not the professional gamers or game content streamers. The sampling period was selected because it covers Chinese New Year

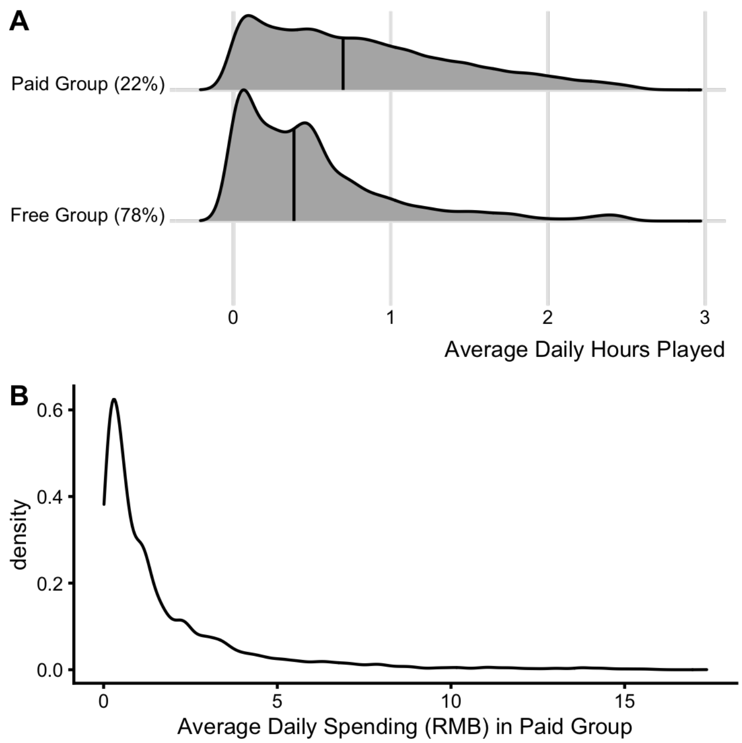

Figure 1: Distribution of Hours Played and In-Game Purchases

holiday, February 2, 2011–February 8, 2011, and there were two major sales promotion (January 27, 2011 and March 23, 2011) during this period. Figure A.1 shows the variation of aggregate price index of in-game goods during the sampling period.[10] The exogenous price variation and holidays (that affect the opportunity cost of time spent on gaming) provide us the sources of identification.

In Table 2 we present summary statistics regarding the data. There are two columns "Free Players" and "Paid Players"—"free players" (78% of the population) are those who never spent on in-game items, and "paid players" (22%) are the rest. Comparing these two groups, "paid players" show more engagement in the game—they have longer hours played, more friends in the game etc. Figure 1 displays the distribution of the hours played and average daily spending in the game (for the group of paid players only). Most notable is the heavy tails of these distributions. In addition, the multiple modes in the distribution of hours played (particularly, the free group) also suggest the importance of accounting observed and unobserved player heterogeneity. Because we observe the login activities of all players in the game, we can calculate the percentage of players who log in and play the game for each day in the sampling period. Figure 2 shows the time

[10]Our price P_t is essentially Laspeyres price index. See Appendix A for the details about how we constructed the price index from in-game expenditure data.

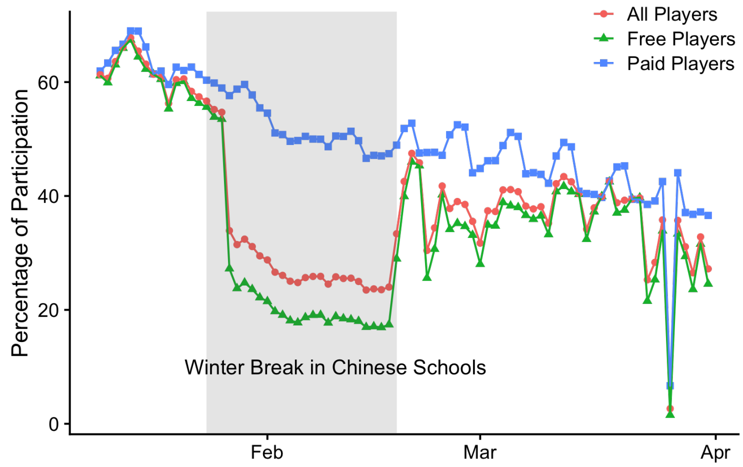

Figure 2: Participation Rate in the Gaming Community

Note: Suppose there are 100 registered players, and 50 of them play the game on day t, the participation rate will be 50% on day t.

series of the participation rate in the gaming community, which shows noticeable variation in the sampling period. In particular, we observe substantial drop in the participation of free players during the Winter Break of Chinese schools. Since the gaming population does not change much, the participation rate is proportional to the number of gamers playing the game on each day, which affects the social interaction of players in the game. For this reason, we will also call the participation rate "social interaction (effect)" hereafter.

Before discussing our structural model, it is useful to consider two questions from simple reduced form regressions. In our structural model, players will make simultaneous choices about the daily hours played (denoted by H_{it}) and the consumption (quantity) of in-game goods (denoted by X_{it}). Here player is indexed by i, and day is indexed by t. The first question is that what is the relationship between hours played and in-game consumption? Are they complementary or substitutes? The second question is that we have claimed that social interaction effect affects individual gamer's choice, do we observe such effect empirically? We consider the following linear simultaneous equation of $(H_{it}, X_{it})'$ for the group of paid players (for free players, their X_{it} are always zero):

$$H_{it} = \delta_{H1} X_{it} + \delta_{H2}' Z_{it} + \nu_{H,it}, \qquad \text{and} \qquad X_{it} = \delta_{X1} H_{it} + \delta_{X2}' Z_{it} + \nu_{X,it},$$

where Z_{it} is a vector of observed player characteristics (see Table 3 and its footnote). Particularly, we impose the following exclusion condition to identify the above model: for the hours equation,

Table 3: Hours Played and In-Game Consumption—Reduced Form Estimates

	Hours Played	In-Game Consumption
In-Game Consumption	0.019	
	(0.001)	
Hours Played		1.429
		(0.192)
Social Interaction	1.523	0.194
	(0.018)	(0.329)
Weekend	0.176	
	(0.004)	
Holiday	0.160	
	(0.007)	
Log of Price Index		−14.574
		(0.420)
Average Spending by Friends		0.343
		(0.004)
Average Spending in Guild		0.177
		(0.017)
Num. Of Friends Who Logged In	0.013	0.000
	(0.000)	(0.003)
R^2	0.076	0.010
Num. obs.	622131	622131

Note: The other common control variables include account tenure, account tenure-squared, and current level in game. The F test statistics for the first step of 2SLS of the hours and in-game consumption equations are 3231.5 and 1169.79, respectively.

we assume that the composite price of in-game goods, average spending among friends, average spending in the player's guild have no direct effect on current hours played; for the consumption equation, we assume that the dummy variables of weekends and holidays do not affect directly the current consumption of in-game goods. Table 3 reports the estimates using 2 stage least squares (2SLS). We have three observations. (a) In-game consumption is complementary not substitute of hours played. This can be seen from the estimates $\hat{\delta}_{H1} = 0.019$ and $\hat{\delta}_{X1} = 1.429$—they are both significant. (b) Social interaction effect has bigger direct effect on the hours played (the coefficient of "Social Interaction" in hours equation is 1.523 and significant) than its effect on the in-game consumption (0.194 and insignificant). This does not mean that social interaction does not affect consumption—it does by affecting hours played which further affects in-game consumption. (c) Gamers are quite price sensitive based on the big negative coefficient of log price index. (d) Conventional "social network" variables, like the number of friends who logged in, have relatively small effect compared with the social interaction (or sometimes called neighborhood effect in the literature) effect. Our structural model estimates will corroberate these findings.

4 Model

We develop a model of players' choices about their hours playing the game and the consumption of in-game goods (in terms of quantity). The goal of this section is to show how we can calculate the value of free players provided that we had estimated this model. We postpone the discussion about the identification, estimation, and some extensions to the next section.

4.1 Basic Setup

We observe n randomly sampled players from the gaming community of our MMO game over T days. The objective here is to characterize the individual choices about the hours played, and the quantity of in-game consumption and the vector of the entire gaming community choices. For each day $t = 1, \ldots, T$, player $i = 1, \ldots, n$ maximizes the following utility function

$$U_{it} = X_{it0} + \psi_{it}(X_{it}) \ln(1 + H_{it}),$$

subject to the budget constraint below. Here $X_{it0} \geq 0$ is a numeraire good, $H_{it} \geq 0$ is the number of hours played on day t, and the multiplier ψ_{it} is the subjective "quality" of playing game as an entertainment. The subjective quality ψ_{it} will be an increasing strictly concave function of the consumption of in-game goods $X_{it} \geq 0$. The quality will also depend on other characteristics about this player, and the expected choices made by the gaming community. Particularly, the corner solution $H_{it} = 0$ means not to play the game at all, and $X_{it} = 0$ means no in-game purchase. Because $\psi_{it}(X_{it})$ is strictly concave in X_{it}, gaming utility $\psi_{it}(X_{it}) \ln(1+H_{it})$ is strictly quasiconcave in $(X_{it}, H_{it})'$.

Note that when $H_{it} = 0$, the utility from playing the game, $\psi_{it}(X_{it}) \ln(1 + H_{it})$, is always zero regardless of the value of quality hence regardless of the value of in-game consumption X_{it}. This implies that a rational player will not purchase in-game goods if she finds that it is optimal that she does not play the game at all (for example, due to high opportunity cost of time). In other words, when $H_{it} = 0$ is optimal, the optimal choice of X_{it} must also be zero provided that the prices of in-game items are positive.[11] This simple observation will help solve and estimate the model. On the other hand, we will let $\psi_{it}(X_{it})$ be greater than zero even if a player did not purchase any in-game goods, i.e. $X_{it} = 0$. Consequently, players may find that it is optimal that they just play the game for free (without purchasing any in-game items). Players face the following budget constraint,

$$X_{it0} + P_t X_{it} + \omega_{it} H_{it} = Outlay,$$

where P_t is a composite price of in-game goods, ω_{it} is the "price" or opportunity cost of playing the game for one hour as leisure, which is usually intercepted as wage rate, and *Outlay* is the total budget or outlay, which is irrelevant to the choice $(X_{it}, H_{it})'$ due to the quasilinear form of utility function. The quasilinear form utility implies that there is no income effect on the consumption of gaming as entertainment, which is reasonable for casual players. For professional gamers, such as Twitch streamers whose actual job is gaming, this assumption is inappropriate. For this reason, our sample excluded those players whose hours played are within the top 1 percentile.

We now consider the specification of the opportunity cost of time spent on playing the game, ω_{it}. Such daily opportunity cost depends on the unobserved player heterogeneity in the opportunity cost of leisure, whether or not day t is a weekend ($Weekend_t$) or holiday ($Holiday_t$),[12] and the unobserved random shock that shifts the time cost (ε_{it}^{time}). Consider

$$\ln \omega_{it} = \omega_{1i} + \omega_2' Z_t^{time} + \varepsilon_{it}^{time}. \tag{1}$$

[11] In data, we only observe 0.27% players (less than 3 out of 1,000 players) who purchased while not playing the game at all. The reason why we still observe in-game purchases, though tiny, while players do not play could be that these players are irrational, or they are testing accounts created by the game developers.

[12] Our sampling period contains Chinese New Year in 2011.

Here $Z_t^{time} \equiv (Weekend_t, Holiday_t)'$, and $Weekend_t$ and $Holiday_t$ are both dummy variables. Apparently, if we observe other covariates that affect the opportunity cost of leisure, like weather, we can redefine Z_t^{time} to include them. Later, we will assume that ε_{it}^{time} follows a normal distribution, hence the time cost ω_{it} follows a log-normal distribution. The log-normal specification enforces the opportunity cost of time to have heavy tail, which is desirable given the heavy tails of the hours played in our data (see panel A of Figure 1). In the actual estimation (section 5.2), we will extend this specification of time cost, so that there could be a lump-sum "login" cost associated with playing the game[13]; after the "login", players pay the per unit cost of time.

The subjective quality ψ_{it} depends on the unobserved player heterogeneity in price sensitivity and preference for the game, the quantity of in-game purchases X_{it}, social interaction or neighborhood effect S_{it}^e (defined in the next paragraph), other observed players characteristics Z_{it}^{qual} (such as the hours played in the previous session), and quality shock ε_{it}^{qual}. Consider

$$\ln \psi_{it} = \alpha_{1i} \ln(1 + X_{it}) + \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_{it}^e + \varepsilon_{it}^{qual}, \tag{2}$$

where $\alpha_{1i} \in (0, 1)$, so that ψ_{it} is monotone increasing and strictly concave in X_{it} for each player. It will also be clear later that α_{1i} determines players price elasticity at the intensive margin. We used $\ln(1+X_{it})$ rather than $\ln X_{it}$ so that the corner solution $X_{it} = 0$ is possible. It is for the same reason why we used $\ln(1 + H_{it})$ earlier. Moreover, the intercept term γ_{1i} determines the intrinsic preference about playing the game. When we start discussing the gamers purchase choice (eq. (7) below), it will be clear that when α_{1i} is small enough and/or γ_{1i} is small enough, players will not purchase any in-game items. Intuitively, players may not purchase for two reasons. First, they may find the game is not that interesting, hence spend less time on playing the game, which makes the return of in-game goods consumption smaller. This is reflected by small intercept γ_{1i} of the log-quality equation. Second, players may find the in-game goods too expensive. This is reflected by small coefficient α_{1i} associated with log-purchase, which determines a player's price elasticity. The random term ε_{it}^{qual} will also have a normal distribution, so the conditional distribution of quality will be log normal.

Lastly, we specify the social interaction effect S_{it}^e. We let S_{it}^e be the player's expectation of the share of players who play the game on day t. Players' beliefs are rational satisfying the self-

[13] Even if a player has an account, it typically takes time to set up the computer or game console, to load the game content etc. Indeed, one big selling point of the new game console generation is the faster load times, see "PlayStation 5 vs. Xbox Series X: The Next-Gen Game Console Brawl", PCMag UK, November 12, 2020, `https://uk.pcmag.com/comparison/129931/playstation-5-vs-xbox-series-x-the-next-gen-game-console-brawl`.

consistency condition, whose precise definition will be given in a moment after we clarify the conditional choice probability (CCP) of playing the game. Simply put, by rational expectation, we will let the expected S_{it}^e be the actual observed percentage of participating in the game, denoted by S_t.

4.2 Players' Choices

Our players make the following sequence of decisions in order to maximize their utility. A player first decides whether or not to play the game on day t at all—we call it a *binary "participation"* decision, which will be denoted by D_{it}^H. Section 4.2.1 below shows that the participation decision can be described as a binary choice model with social interactions (Brock and Durlauf, 2001). If a player participates, the player then decides whether or not to purchase any in-game goods on day t—we call it a *binary "purchase"* decision, and denote it by D_{it}^X. Lastly, the player decides the hours played and the consumption of in-game goods (if any) at the intensive margin. Theorem 1 at the end of this section shows that there is a unique solution of our model.

For the purpose of exposition, define

$$Y_{it}^H \equiv \ln(1 + H_{it}) \qquad \text{and} \qquad Y_{it}^X \equiv \ln(1 + X_{it}).$$

Apparently, $Y_{it}^H \geq 0$ and $Y_{it}^X \geq 0$, and $Y_{it}^H = 0$ and $Y_{it}^X = 0$ correspond to the corner solutions $H_{it} = 0$ and $X_{it} = 0$, respectively.

4.2.1 Hours Played

Define a dummy variable $D_{it}^H = 1$ if $H_{it} > 0$ and $D_{it}^H = 0$, otherwise. We say player i participates in the game on day t if $D_{it}^H = 1$. Participation means playing the game—we call it "participation" to emphasize its binary nature and differentiate it from the continuous choice of the hours played. The participation choice is based on the utility maximization; $D_{it}^H = 0$ occurs when the corner solution $H_{it} = 0$ is optimal. By the Kuhn-Tucker conditions, the corner solution of hours is optimal if and only if the marginal utility of hours played H_{it} divided by the time cost ω_{it}, evaluated at $H_{it}^* = 0$ and the associated optimal X_{it}^*, is not greater than the marginal utility of the numeraire, which is the unit. As for the optimal X_{it}^* associated with $H_{it}^* = 0$, note that when $H_{it}^* = 0$, the utility from the game $\psi_{it}(X_{it}) \ln(1 + H_{it}^*)$ is always zero, hence the optimal in-game purchase X_{it}^* (corresponding to $H_{it}^* = 0$) must also be 0, otherwise one can improve the utility by trading in-game goods with the numeraire. Following this argument and the model setup, it is easy to see that the binary participation decision is simply,[14]

$$D_{it}^H = \boldsymbol{I}\left[(\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_{it}^e + (\varepsilon_{it}^{qual} - \varepsilon_{it}^{time}) > 0\right], \tag{3}$$

where $\boldsymbol{I}$ is an indicator function. It has the familiar form of binary discrete choice model. Let $p(Z_{it}, S_{it}^e, \gamma_{1i}, \omega_{1i}) \equiv \mathrm{E}(D_{it}^H \mid Z_{it}, S_{it}^e, \gamma_{1i}, \omega_{1i})$ be the conditional choice probability function, where $Z_{it} \equiv (Z_t^{time}, Z_{it}^{qual}, P_t)$. Though price P_t does not appear here, it will be convenient to let Z_{it} include P_t as well.

Self-consistency of beliefs or rational expectations require that the subjective belief S_{it}^e equals the objective expected value of the percentage of players in the gaming community who participate on day t. More formally, we require that

$$S_{it}^e = S_t = \int p_i(Z_{it}, S_{it}^e = S_t, \gamma_{1i}, \omega_{1i})\, \mathrm{d}\hat{F}_Z\, \mathrm{d}F(\gamma_{1i}, \omega_{1i}), \tag{4}$$

where $\hat{F}_Z$ is the empirical probability distribution of Z_{it}, and $F(\gamma_{1i}, \omega_{1i})$ is the distribution function of $(\gamma_{1i}, \omega_{1i})'$. This equation is identical to the self-consistency condition in Brock and Durlauf (2001), where they studied the binary choice models with social interaction. It then follows from their result (Proposition 1), which mainly applies the Brouwer fixed-point theorem, that a self-consistent equilibrium always exist. Due to our extremely large sample size (near 3 million), we will let $S_{it}^e = S_t = n_t/n$ and ignore the sampling error, where n is the total number of players who ever appeared in the sampling periods, and $n_t = \sum_{i=1}^n D_{it}^H$ is the number of players who participated in the game on day t. We have seen the plot of the time series S_t in Figure 2 (the line corresponding to all players).

When $D_{it}^H = 1$, player will choose a positive number of hours played to maximize the utility. The condition is that the player will keep playing until $\partial U_{it}/\partial H_{it} = \omega_{it}$ (marginal utility equals marginal cost) or equivalently $\ln(\partial U_{it}/\partial H_{it}) = \ln \omega_{it}$. It can be shown that this condition implies that $Y_{it}^H = \ln \psi_{it} - \ln \omega_{it}$. Recall that $Y_{it}^H \equiv \ln(1 + H_{it})$. This gives rise to the choice of hours played:

$$Y_{it}^H = D_{it}^H \times \tilde{Y}_{it}^H, \tag{5}$$

[14] The utility function is $U_{it} = X_{it0} + \psi(X_{it}) \ln(1 + H_{it})$. So the marginal utility (MU) of hours played is $\partial U_{it}/\partial H_{it} = \psi(X_{it})/(1 + H_{it})$, and the MU of numeraire is $\partial U_{it}/\partial X_{it0} = 1$. When $H_{it}^* = 0$, the optimal X_{it}^* is 0, and the resulted MU of hours played becomes $\psi(X_{it}^* = 0) = \exp(\gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_{it}^e + \varepsilon_{it}^{qual})$ by the specification of $\psi(X_{it}) = (1 + X_{it})^{\alpha_{1i}} \exp(\gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_{it}^e + \varepsilon_{it}^{qual})$. Player will participate, i.e. $D_{it}^H = 1$, if the MU of hours divided by the unit cost of playing time ω_{it} is greater than the MU of numeraire (whose price is normalsized to 1), which equals 1. That is $D_{it}^H = \boldsymbol{I}(\psi(X_{it}^* = 0)/\omega_{it} > 1) = \boldsymbol{I}(\ln \psi(X_{it}^* = 0) > \ln \omega_{it})$. We have $\ln \psi(X_{it}^* = 0) = \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_{it}^e + \varepsilon_{it}^{qual}$, and $\ln \omega_{it} = \omega_{1i} + \omega_2' Z_t^{time} + \varepsilon_{it}^{time}$.

where the latent $\tilde{Y}_{it}^{H} = \ln \psi_{it} - \ln \omega_{it}$, i.e.,

$$\tilde{Y}_{it}^{H} = (\gamma_{1i} - \omega_{1i}) + \alpha_{1i} Y_{it}^{X} + \gamma_2' Z_{it}^{qual} - \omega_2' Z_{t}^{time} + \beta S_t + (\varepsilon_{it}^{qual} - \varepsilon_{it}^{time}). \tag{6}$$

The multiplication by D_{it}^{H} is to enforce that the hours played is zero if a player had decided not to play the game at all on day t. Note that because $\alpha_{1i} > 0$ (by assumption), $\tilde{Y}_{it}^{H} \geq 0$ if $D_{it}^{H} = 1$.

4.2.2 Expenditure Choice

Similar to the definition of D_{it}^{H}, define a dummy variable $D_{it}^{X} = 1$ if $X_{it} > 0$ (equivalently, $Y_{it}^{X} > 0$) and $D_{it}^{X} = 0$, otherwise. Following her participation decision D_{it}^{H}, players make a binary purchase decision D_{it}^{X}. If $D_{it}^{H} = 0$, assume $D_{it}^{X} = 0$. That is if the consumer had decided not to participate and play the game, she will not purchase any in-game goods either. Provided that the player had decided to participate, $D_{it}^{H} = 1$, her purchase decision is also based on the comparison between the marginal utility of in-game purchase divided by price P_t and the marginal utility of the numeraire, which is one.[15] $D_{it}^{X} = 0$ occurs when the corner solution $X_{it} = 0$, or equivalently $Y_{it}^{X} = 0$, is optimal, which would be the case if the marginal utility $\partial U_{it}/\partial X_{it}$, evaluated at $X_{it}^{*} = 0$ and optimal H_{it}^{*} (hence optimal Y_{it}^{H*}) associated with $X_{it}^{*} = 0$, is greater than price P_t. According to eq. (5) and (6), the optimal hours when $X_{it}^{*} = 0$, i.e. $Y_{it}^{X*} = 0$, is

$$Y_{it}^{H*} = (\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_{t}^{time} + \beta S_t + (\varepsilon_{it}^{qual} - \varepsilon_{it}^{time}).$$

We thus have

$$D_{it}^{X} = \boldsymbol{I}\left[\ln(\alpha_{1i}) + \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t + \ln Y_{it}^{H*} - \ln P_t + \varepsilon_{it}^{qual} > 0\right], \tag{7}$$

from $D_{it}^{X} = \boldsymbol{I}\big(\ln \partial U_{it}/\partial X_{it} > \ln P_t\big)$. When $D_{it}^{X} = 1$, player will keep buying until $\partial U_{it}/\partial X_{it} = P_t$ or equivalently $\ln(\partial U_{it}/\partial X_{it}) = \ln P_t$. It is easy to verify that such decision rule leads to the following model about in-game purchase.

$$Y_{it}^{X} = D_{it}^{H} \times D_{it}^{X} \times \tilde{Y}_{it}^{X}, \tag{8}$$

where the latent variable $\tilde{Y}_{it}^{X}$ is as follows,

$$\tilde{Y}_{it}^{X} = \frac{1}{1-\alpha_{1i}} \times \left[\ln \alpha_{1i} + (\gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t) + \big(\ln Y_{it}^{H} - \ln P_t\big) + \varepsilon_{it}^{qual}\right]. \tag{9}$$

[15]The marginal utility of in-game goods is $\partial U_{it}/\partial X_{it} = \alpha_{1i}(1+X_{it})^{\alpha_{1i}-1} \exp(\gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t + \varepsilon_{it}^{qual}) Y_{it}^{H}$. Here we have imposed the self-consistency of belief condition, so that $S_{it}^{e} = S_t$.

The multiplication by D_{it}^{H} and D_{it}^{X} is to enforce that in-game purchase quantity will zero, if the player had decided not to play the game at all on day t ($D_{it}^{H} = 0$) or they had found that it is optimal to not purchase any in-game goods ($D_{it}^{X} = 0$). The equations about the binary purchase choice (eq. (7)) and the continuous consumption (eq. (9)) show the dual roles of α_{1i}: (a) it affects the purchase at the extensive margin—a higher α_{1i} increases the likelihood of the event $D_{it}^{X} = 1$; (b) it affects the purchase at the intensive margin—α_{1i} determines the elasticity provided that consumers decide to purchase (note that in eq. (9), the coefficient associated with log price $\ln P_t$ is $(1 - \alpha_{1i})^{-1}$), a higher α_{1i} leads to a higher elasticity.

One important issue is that whether or not the binary participation model eq. (3), binary purchasing model eq. (7), and the nonlinear simultaneous model eq. (5) and (8) together have a solution (the existence of an equilibrium), and whether or not the solution is unique (uniqueness of equilibrium). The existence is called coherency condition and the uniqueness is called completeness following the terms of Tamer (2003). Amemiya (1974) was the first to study the coherency and completeness of *linear* simultaneous Tobit model. His results only apply only to linear simultaneous Tobit model. The more general results in Lewbel (2007) provide conditions for simultaneous qualitative dependent variable models, e.g. simultaneous probit models, but his results require at least one dependent variable is binary. We thus cannot apply Lewbel's results either. In the following theorem, we prove the coherency and completeness.

Theorem 1 (Coherency and Completeness)**.** *If $\alpha_{1i} \in (0, 1)$, the binary participation model eq.* (3)*, binary purchasing model eq.* (7)*, and the nonlinear simultaneous model eq.* (5) *and* (8) *together are coherent and complete.*

Theorem 1 provides the foundation for our identification and estimation results. If the model is not coherent, it must be misspecified. If the model in incomplete, the point identification of the structural parameters is likely to fail, and we may have to consider set identification. Even worse, the counterfactual analysis, in which the value of free players is calculated, will be difficult.

Remark 1 (Multiple equilibria)**.** *Careful readers may have noticed that our quasilinear utility function $U_{it} = X_{it0} + \psi_{it}(X_{it}) \ln(1 + H_{it})$ is not guaranteed to be strictly quasiconcave, even though the part of utility from playing games, $\psi_{it}(X_{it}) \ln(1 + H_{it})$ is strictly log concave, hence strictly quasiconcave.*[16] *Indeed, it turns out in general the utility function $U_{it}(X_{0it}, X_{it}, H_{it})$* cannot *be*

[16]It can be shown that one sufficient condition for $U_{it}(X_{0it}, X_{it}, H_{it})$ being strictly concave is that $\ln(1 + H_{it}) > \alpha_{1i}/(1 - \alpha_{1i})$, which may fail when H_{it} is small enough.

concave provided that it satisfies the two intuitive conditions (see Proposition B.1 in Appendix B): (a) the in-game purchases does not affect the utility of gaming provided that the player did not spend time on playing, and (b) the in-game purchases increases the marginal utility of hours playing games.

If the utility function is not strictly quasiconcave, how can we conclude that the optimal decision is unique as required by the completeness condition? The uniqueness relies on the sequential nature of player's decision making, where a player first decides whether or not to participate, then whether or not to purchase any in-game goods, and finally she decides the optimal hours and in-game consumption. Intuitively, it is possible that $(H_{it} = 0, X_{it} = 0)$ *(no play, no purchase) and* $(\tilde{H}_{it}, \tilde{X}_{it})'$ *(for some positive hours and in-game purchase) both maximize the utility. But our sequential decision making process acts as an equilibrium refinement rule, that rules out the second equilibrium.*

4.3 Value of Free Players

There are two groups of players, free and paid groups, in data. Free group consists of those players who never spent money on in-game goods in our sampling period. Paid group consists of those who purchased in-game items at least once in our sample. We are interested in evaluating the value of free players from the perspective of their social interactions with paid players. We claim that we have estimated the model. The next section will supplement the details about the identification and estimation. Assuming this claim for the moment, we show how to calculate the value of free players.

The value of free players in this model comes from their impact on the social interaction effect. Consider the case when none of the free players participate in the game, and the resulted social interaction effect becomes

$$S_t^{cf} = \frac{n_{paid,t}}{n}.$$

We then can simulate the resulted counterfactual participation, hours played and consumption by the paid players (see section 5.3 for details). It is straightforward to solve these counterfactual decisions numerically from eqs. (3) to (9). Theorem 1 guarantees that there is a unique solution in this counterfactual exercise. Without Theorem 1, one may find multiple solutions of the expenditure by the paid players, leading to ambiguous conclusions about the value of free players.

Let $X_{it}^{cf} \equiv X(S_t^{cf}, Z_{it}, \varepsilon_{it})$ and $H_{it}^{cf} \equiv H(S_t^{cf}, Z_{it}, \varepsilon_{it})$ be the counterfactual choices by a paid player. Here $\varepsilon_{it} \equiv (\varepsilon_{it}^{qual}, \varepsilon_{it}^{time})'$. The notation is to emphasize the dependence on the counterfactual social interaction effect S_t^{cf}, the observed player characteristics Z_{it}, and the unobserved

Table 4: The Value of Free Players

	Free Players In	Free Players Out
$\mathrm{E}(X)$	2.34	0.98
$\mathrm{E}(H)$	0.70	0.27
$\Pr(X > 0)$	0.0328	0.0138
$\Pr(H > 0)$	0.522	0.235
$\mathrm{E}(X \mid X > 0)$	88.00	87.61
$\mathrm{E}(H \mid H > 0)$	1.34	1.17

Note: The results here are based on 50 replications using the entire group of paid players (7,791 players over 81 days).

characteristics ε_{it}. Like the treatment effect literature, we can report the average (treatment) effect

$$V_{ATE}(S_t^{cf}) = \int\int X(S_t^{cf}, Z_{it}, \varepsilon_{it}) - X(S_t, Z_{it}, \varepsilon_{it}) \,\mathrm{d}\,\hat{F}_Z \,\mathrm{d}\,F_{\varepsilon_{it}}$$

The term $V_{ATE}(S_t^{cf})$ reflects the value of the group of free players.

Using the estimated model (again we postpone the estimation details to the next section), we ran the counterfactual experiments, in which free players, though exist in the game, do not play, and report the results in Table 4. We found that the average spending drops by 58.1% and the average hours played drops by 61.4%. The table also shows that the effects on in-game purchase is mostly at the extensive margin—it has little effect at the intensive margin. The effect on the hours played is also mostly on the extensive margin.

5 Identification and Estimation

We will first show which parameters are identified using our long panel data, then describe the model estimation method and results that leads to our calculation of the value of free players.

5.1 Identification

Our identification arguments proceed in two steps. In the first step, we focus on the relatively simple binary participation choice model eq. (3), where we can identify a big chunk of gaming utilities up to scale. Partially knowing the utility of gaming, we then in the second step identify the demand function for in-game goods for the group of paid players.

Assumption 1. *Assume that* $\varepsilon_{it} \equiv \left(\varepsilon_{it}^{qual},\ \varepsilon_{it}^{time}\right)'$ *is independent of the observable characteristics* Z_{it} *and social interaction effect* S_t*, and* ε_{it} *follows a bivariate normal distribution.*

For Assumption 1, it should be remarked that the social interaction effect $S_t = S_{it}^e$ (under the self-consistency of beliefs) is the *expected* percentage of game participation in the gaming community. So it does not involve utility shocks ε_{it}, which can also be seen from its formula in eq. (4). Let $\sigma_{qual}^2 \equiv \mathrm{Var}(\varepsilon_{it}^{qual})$, and let $\tilde{\sigma}^2 \equiv \mathrm{Var}(\varepsilon_{it}^{qual} - \varepsilon_{it}^{time})$. It will be useful to define

$$\varepsilon_{it}^* \equiv \frac{\varepsilon_{it}^{qual} - \varepsilon_{it}^{time}}{\tilde{\sigma}} \sim N(0,1).$$

Let $\rho \equiv \mathrm{corr}(\varepsilon_{it}^{qual}, \varepsilon_{it}^*)$. Apparently, knowing the joint distribution of $\left(\varepsilon_{it}^{qual}, \varepsilon_{it}^*\right)'$ is equivalent to knowing the joint distribution of $\left(\varepsilon_{it}^{qual}, \varepsilon_{it}^{time}\right)'$. Under Assumption 1, the participation choice D_{it}^H model in eq. (3) can be viewed as a probit model with fixed effect $(\gamma_{1i} - \omega_{1i})$, which is rephrased below for the convenience of reading,

$$D_{it}^H = \boldsymbol{I}\left(\tilde{\sigma}^{-1}\left[(\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_t\right] + \varepsilon_{it}^* > 0\right). \tag{10}$$

In addition to the usual identification up to scale issues in discrete choice models, the presence of social interactions S_t raises an additional concern of identification, since it depends on various neighborhood-level variables (the whole gaming community is one neighborhood), including the dummy variables of weekend and holiday, price etc, and these neighborhood-level variables are also regressors in the linear index of this probit model. Unlike the linear-in-means model studied in Manski (1993) (see Graham and Hahn (2005) for simpler exposition), Brock and Durlauf (2001) proved that because S_t, which is integrated CCP function, is nonlinear in neighborhood-level variables, the social interaction effect is identified (up to scale $\tilde{\sigma}$).

As a summary, we clarify what we can and cannot identify using only participation choices. With our long panel data (we observe 81 days for each player),[17] we can identify $\gamma_{1i} - \omega_{1i}$ (for all $i = 1, \ldots, n$), γ_2, ω_2, and β up to scale $\tilde{\sigma}^{-1}$, that is $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$, $\tilde{\sigma}^{-1}\gamma_2$ etc. Note that we only identify $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$, and cannot separate γ_{1i} and ω_{1i}. Intuitively, we cannot separately

[17] We have two pieces of evidence supporting the claim that our panel data are "long" enough along with the time dimension. First, non-linear fixed effects models, like our probit model with fixed effect, suffer the incidental parameter bias problem when the number of periods T in panel data is small. Analytical bias correction has been proposed in the literature (Hahn and Newey, 2004; Fernández-Val, 2009). We found that our estimates with and without the bias correction are very close. This suggests that our panel data are long enough so that the incidental parameters bias is not concern. The second evidence is that our fixed effects can be estimated very precisely (the average t-value is 26.1).

identify the unobserved time cost ω_{1i} and unobserved preference of playing games γ_{1i} using only participation choices—one player may choose not to play the game because she has low interests in playing video games (low γ_{1i}) or her opportunity cost of playing the game is too high (high ω_{1i}). We will rely on purchase decisions to identify γ_{1i} (gaming preference), because unobserved heterogeneity in time cost ω_{1i} does not directly affect the in-game purchase after controlling the hours played. Also, the marginal utility of in-game goods, that is α_{1i} in the quality eq. (2), is not identified from participation choice alone. Intuitively, if a player did not participate in the game, it is impossible to learn her preference regarding in-game goods.

We next discuss the identification of the simultaneous nonlinear models eq. (5) and (8). The identification is a concern due to the simultaneity. After the estimation of participation stage, we have recovered many parameters. It helps rephrase the original simultaneous nonlinear models and obtain a simpler form. Collapsing the recovered terms from the participation stage, we define

$$W_{it}^X \equiv \tilde{\sigma}^{-1}\left(\gamma_2' Z_{it}^{qual} + \beta S_t\right), \qquad W_{it}^H \equiv \tilde{\sigma}^{-1}\left[(\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_t\right].$$

Using W_{it}^X and W_{it}^H, we can rewrite eq. (5) and (8) as follows,

$$Y_{it}^X = D_{it}^H \times D_{it}^X \times \tilde{Y}_{it}^X, \qquad\qquad Y_{it}^H = D_{it}^H \times \tilde{Y}_{it}^H,$$

with

$$\tilde{Y}_{it}^X = \frac{1}{1-\alpha_{1i}} \times \left[\ln\alpha_{1i} + \gamma_{1i} + \tilde{\sigma}W_{it}^X + \ln Y_{it}^H - \ln P_t + \varepsilon_{it}^{qual}\right]$$

and

$$\tilde{Y}_{it}^H = \alpha_{1i} Y_{it}^X + \tilde{\sigma} W_{it}^H + \tilde{\sigma}\varepsilon_{it}^*.$$

If we could identify the coefficients associated with the above regressors, we will be able to separately identify γ_{1i} (preference of the game) and ω_{1i} (opportunity cost of leisure time), and α_{1i} (marginal utility of in-game purchase) for each player. Then $\tilde{\sigma}$, σ_{qual}, and $\rho \equiv \text{corr}(\varepsilon_{it}^{qual}, \varepsilon_{it}^*)$ are identified from the residuals.

Depending on whether or not a player i purchased in-game items at all in our sample, we have positive and negative identification results for paid players and free players, respectively. We start with negative results for free players. Suppose player i is a free player, who never purchased in our data, that is $D_{it}^X = 0$ and $Y_{it}^X = 0$ for all $t = 1, \ldots, T$. We cannot identify α_{1i} and γ_{1i} for any free player i, since there is no variation of dependent variables D_{it}^X or Y_{it}^X. But this does not affect our identification of the value of free players, because in our model, their value hinges on their participation of the game, which requires only knowing the already identified $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$.

For paid players, we have positive news—we can identify all γ_{1i}, ω_{1i} and α_{1i}. Theorem 2 below shows that the above nonlinear simultaneous model is identified for any paid players. Like the conventional linear simultaneous model, the usual exclusion conditions turn out to be sufficient to identify the model. The exclusion conditions are satisfied here. For the consumption equation, equation about $\tilde{Y}_{it}^{X}$, log-price $\ln P_t$ is the excluded variable that affects the expenditure but not the hours played. For the hours equation, equation about $\tilde{Y}_{it}^{H}$, weekend and holiday dummy variables in Z_t^{time} are the excluded variable that affects the hours played but not the consumption. Amemiya (1974) showed that such exclusion conditions can identify linear simultaneous Tobit model. Our identification of nonlinear simultaneous equation will rely on the more recent results by Matzkin (2015). Our sequential estimation method below will make the identification even more transparent.

Theorem 2 (Identification of Paid Players' Preference)**.** *Suppose i is a paid player, and $\alpha_{1i} \in (0, 1)$ so that Theorem 1 hold, and Assumption 1 hold. We can identify all parameters specifying the preference of this paid player i, including $\alpha_{1i}, \gamma_{1i}, \omega_{1i}, \gamma_2, \omega_2$, β, σ_{qual}, $\tilde{\sigma}$ and ρ.*

5.2 Estimation Steps

Though it seems straightforward to estimate the model by the maximum likelihood estimation method (MLE), a joint MLE turns out to be computationally complicated if feasible at all because of various truncation implied by the sequential decision process, and it does not transparently deliver identification of the deep structural parameters. Here we propose a sequential estimation method that is fast, robust and easy to implement, though it is less efficient than the full MLE.[18] Due to our extremely large sample size, efficiency is not a concern to us or companies who own much larger data.

We first give an outline of our estimation method. Our estimation proceeds in three steps. In the first step, we estimate the participation choice model using probit models. From the first step, we can estimate $\theta_1 \equiv \tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i} - \omega_{login}, \gamma_2, \omega_2, \beta)'$. Here ω_{login} is a lump sum login cost affecting the cost of participation. For simplicity, we only write $\gamma_{1i} - \omega_{1i}$ to denote $\gamma_{1i} - \omega_{1i}$ for all players $i = 1, \ldots, n$. In the second step, we use 2SLS to estimate the equation of hours played. From this step, we can estimate $\theta_2 \equiv (\omega_{login}, \tilde{\sigma}, \alpha_{1i})'$. Note that α_{1i} is only identified for paid players. In the

[18] We did simulation studies to evaluate the finite sample performance of the proposed estimation method. The results are available upon request.

third step, we estimate the binary purchase choice model among the paid players. From this last step, we can estimate $\theta_3 \equiv (\gamma_{1i}, \sigma_{qual}, \rho)'$. Knowing $\theta \equiv (\theta_1', \theta_2', \theta_3')'$, we have completely identified the preference of paid players, and can run counterfactual analysis when they face different social interaction effect.

Step 1: Game participation choices

The participation choice D_{it}^H is based on the decision rule eq. (10), which is extended below. We consider

$$D_{it}^H = \boldsymbol{I}\left(\tilde{\sigma}^{-1}\left[(\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_t\right] - \tilde{\sigma}^{-1}\omega_{login} + \varepsilon_{it}^* > 0\right). \qquad (11)$$

Compared with the original participation model eq. (3), we have a lump sum login cost $\omega_{login} \geq 0$, which only appears in the participation decision. After login, ω_{login} does not affect the opportunity cost of leisure anymore. It is straightforward to estimate the probit model eq. (11) using fixed effect estimator. From this stage, we have estimates $\hat{\theta}_1$ of $\theta_1 = \tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i} - \omega_{login}, \gamma_2, \omega_2, \beta)'$. It helps define

$$\tilde{W}_{it}^H \equiv \tilde{\sigma}^{-1}\left(\gamma_{1i} - \omega_{1i} + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_t\right) - \tilde{\sigma}^{-1}\omega_{login}.$$

The only difference between $\tilde{W}_{it}^H$ and W_{it}^H is the presence of login cost ω_{login}, and we have $W_{it}^H = \tilde{W}_{it}^H + \tilde{\sigma}^{-1}\omega_{login}$. Using $\hat{\theta}_1$, we have generated regressors $\hat{W}_{it}^X$ and $\hat{\tilde{W}}_{it}^H$ by substituting the unknown parameters in their definition with the estimates $\hat{\theta}_1$.

Step 2: Hours played

Note that provided that $D_{it}^H = 1$ (player i participates in the game), we have that

$$Y_{it}^H = \alpha_{1i} Y_{it}^X + \tilde{\sigma} W_{it}^H + \tilde{\sigma}\varepsilon_{it}^* = \omega_{login} + \alpha_{1i} Y_{it}^X + \tilde{\sigma}\tilde{W}_{it}^H + \tilde{\sigma}\varepsilon_{it}^*.$$

The second identity follows from $W_{it}^H = \tilde{W}_{it}^H + \tilde{\sigma}^{-1}\omega_{login}$. We have estimated $\tilde{W}_{it}^H$ in the first step, so it is known as $\hat{\tilde{W}}_{it}^H$ in this step. To derive an estimating equation, we need to correct for the sample selection due to the choice $D_{it}^H = 1$ and also address the endogeneity of Y_{it}^X (the endogeneity comes from the simultaneity). Let Z_{it}^{IV} be a vector of instrumental variables (IV) that affect consumption but are mean independent of the participation shocks ε_{it}^*. We have

$$\mathrm{E}(\varepsilon_{it}^* \mid Z_{it}^{IV}, \tilde{W}_{it}^H, D_{it}^H = 1) = \lambda(\tilde{W}_{it}^H),$$

where $\lambda(\cdot) \equiv \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio. We then have the conditional moment equation for any player i,

$$\mathrm{E}\Big(Y_{it}^H - \omega_{login} - \alpha_{1i} Y_{it}^X - \tilde{\sigma}\big[\hat{\tilde{W}}_{it}^H + \lambda(\hat{\tilde{W}}_{it}^H)\big] \;\Big|\; Z_{it}^{IV}, \tilde{W}_{it}^H, D_{it}^H = 1\Big) = 0,$$

from which we can estimate α_{1i} (for each paid player i), ω_{login}, and $\tilde{\sigma}$. For any free player j, her Y_{jt}^X always equals zero, hence α_j for a free player is not estimated. In practice, one just need to create a dummy variable for each player to take account of player specific intercept, and run 2SLS of the linear regression of Y_{it}^H on Y_{it}^X, $\hat{\tilde{W}}_{it}^H + \lambda(\hat{\tilde{W}}_{it}^H)$, and player dummies using the interaction terms between player dummies and Z_{it}^{IV} (and the exogenous $\hat{\tilde{W}}_{it}^H$) as IV and using the data of all players who had participated in the game, i.e. $D_{it}^H = 1$. Let $\hat{\omega}_{login}$, $\hat{\tilde{\sigma}}$, and $\hat{\alpha}_{1i}$ denote the estimates. Also let $\hat{W}_{it}^H \equiv \hat{\tilde{W}}_{it}^H + \hat{\tilde{\sigma}}\hat{\omega}_{login}$ be the estimate of W_{it}^H.

The primary instrumental variables we used are the log-price $\ln P_t$.[19] In our model, the log of price affects the consumption of in-game goods, but is independent of ε_{it}^*. Price variation in our data comes from the promotion activities. There were two major promotions in our sampling period. One was to celebrate the Chinese New Year, and the other one was to accompany a major release of new contents.

Step 3: Binary purchase choices in paid group

In this step, we will estimate $\theta_3 \equiv (\gamma_{1i}, \sigma_{qual}, \rho)'$ based on consumers binary purchase decisions eq. (7), which can be written as follows using the notation of W_{it}^X and W_{it}^H,

$$D_{it}^X = \boldsymbol{I}\left[\gamma_{1i} + \zeta_{it} + \ln(W_{it}^H + \varepsilon_{it}^*) + \varepsilon_{it}^{qual} > 0\right],$$

where

$$\zeta_{it} \equiv \ln \alpha_{1i} + \ln \tilde{\sigma} + \tilde{\sigma} W_{it}^X - \ln(P_t),$$

Note that ζ_{it} is observable given our estimates of $\hat{\theta}_1$ and $\hat{\theta}_2$. Let $\hat{\zeta}_{it}$ be the estimate.

The objective is to estimate $\theta_3 \equiv (\gamma_{1i}, \sigma_{qual}, \rho)'$ by MLE using the data $(D_{it}^X, \hat{\zeta}_{it}, \hat{W}_{it}^H)'$ for those players, who are from the paid group and participate in the game. Again, the participation creates selection issue. We have that

$$\Pr(D_{it}^X = 1 \mid Z_{it}, D_{it}^H = 1) = \Pr(D_{it}^X = 1 \mid Z_{it}, \varepsilon_{it}^* > -\hat{W}_{it}^H).$$

[19] We also used log of price index yesterday, average spending among friends, and average spending in the player's guild (0, if the player did not join a guild) as IV. The results are similar with and without these extra instrumental variables.

Standard calculation shows that the likelihood function is

$$\begin{aligned}\ell(\gamma_{1i}, \sigma_{qual}, \rho) &= \Pr(D_{it}^X = 1 \mid Z_{it}, D_{it}^H = 1) \\ &= \int_0^\infty \left[1 - \Phi\left(\frac{-(\gamma_{1i} + \zeta_{it} + \ln(v)) - \sigma_{qual}\rho(v - \hat{W}_{it}^H)}{\sigma_{qual}\sqrt{1-\rho^2}}\right)\right] \frac{\phi(v - \hat{W}_{it}^H)}{1 - \Phi(-\hat{W}_{it}^H)}\, \mathrm{d}\, v.\end{aligned}$$

We have applied the change of variables, so that this integral can be easily computed using Gauss-Laguere quadrature (otherwise, the limits of integral vary with $\hat{W}_{it}^H$). Having the likelihood, we can estimate σ_{qual}, ρ, γ_{1i} (for paid players) by MLE.

5.3 Counterfactual Simulation to Evaluate the Value of Free Players

After the model estimation, we can simulate $(D_{it}^H, D_{it}^X, Y_{it}^H, Y_{it}^X)$ from the model. This is useful to evaluate the goodness of fit and more importantly to do counterfactual analysis. The simulation follows the same decision rule that we have assumed players will follow. For concreteness, we focus on the counterfactual experiment that the social interaction effect S_t equals

$$S_t^{cf} = \frac{n_{paid,t}}{n}.$$

If we let $S_t^{cf} = S_t$, we can simulate $(D_{it}^H, D_{it}^X, Y_{it}^H, Y_{it}^X)$ from this model, and examine the goodness of fit our model specification (reported in Table 7).

Below we describe the steps that we used to simulate $(D_{it}^{H,cf}, D_{it}^{X,cf}, Y_{it}^{H,cf}, Y_{it}^{X,cf})$. The simulation follows the following steps.

Step 1. Compute $W_{it}^{X,cf}$, $W_{it}^{H,cf}$ and ζ_{it}^{cf},

$$\begin{aligned}W_{it}^{X,cf} &\equiv \hat{\tilde{\sigma}}^{-1}\left(\hat{\gamma}_2' Z_{it}^{qual} + \hat{\beta} S_t^{cf}\right), \\ W_{it}^{H,cf} &\equiv \hat{\tilde{\sigma}}^{-1}\left[(\hat{\gamma}_{1i} - \hat{\omega}_{1i}) + \hat{\gamma}_2' Z_{it}^{qual} - \hat{\omega}_2' Z_t^{time} + \hat{\beta} S_t^{cf}\right], \\ \zeta_{it}^{cf} &\equiv \hat{\gamma}_{1i} + \ln \hat{\alpha}_{1i} + \ln \hat{\tilde{\sigma}} + \hat{\tilde{\sigma}} W_{it}^{X,cf} - \ln(P_t).\end{aligned}$$

For each player i and each day t, draw $(\varepsilon_{it}^{qual,cf}, \varepsilon_{it}^{*,cf})$ from the bivariate normal distribution

$$\begin{pmatrix}\varepsilon_{it}^{qual,cf} \\ \varepsilon_{it}^{*,cf}\end{pmatrix} \sim N\left(0, \begin{pmatrix}\sigma_{qual}^2 & \rho\sigma_{qual} \\ \rho\sigma_{qual} & 1\end{pmatrix}\right)$$

Step 2. Simulate participation decisions $D_{it}^{H,cf} = \boldsymbol{I}\left(W_{it}^{H,cf} + \varepsilon_{it}^{*,cf} > \hat{\omega}_{login}\right)$. If $D_{it}^{H,cf} = 0$ (not participation), let $(Y_{it}^{H,cf}, Y_{it}^{X,cf})' = (0,0)'$.

Step 3. Simulate purchase decisions $D_{it}^{X,cf} = \boldsymbol{I}\left(\zeta_{it}^{cf} + \ln(W_{it}^{H,cf} + \varepsilon_{it}^{*,cf}) + \varepsilon_{it}^{qual,cf} > 0\right)$ for those whose $D_{it}^{H,cf} = 1$. If $D_{it}^{X,cf} = 0$ (not purchase), let $Y_{it}^{X,cf} = 0$, and $Y_{it}^{H,cf} = \ln(\hat{\tilde{\sigma}}(W_{it}^{H,cf} + \varepsilon_{it}^{*,cf}))$.

Step 4. Simulate consumption decision. For those whose $D_{it}^{H,cf} = D_{it}^{X,cf} = 1$, find $Y_{it}^{X,cf}$ that solves

$$(1 - \hat{\alpha}_{1i})Y_{it}^{X,cf} = \zeta_{it}^{cf} + \ln\left(\hat{\alpha}_{1i}Y_{it}^{X,cf}/\hat{\tilde{\sigma}} + W_{it}^{H,cf} + \varepsilon_{it}^{*,cf}\right) + \varepsilon_{it}^{qual,cf}.$$

Theorem 1 guarantees that there exists a unique solution of $Y_{it}^{X,cf}$. And let $Y_{it}^{H,cf} = \hat{\alpha}_{1i}Y_{it}^{X,cf} + \hat{\tilde{\sigma}}(W_{it}^{H,cf} + \varepsilon_{it}^{qual,cf})$.

5.4 Estimation Results

We report the estimation results in Table 5. The social interaction effect is significant and has big effect. Note that we have controlled the variables about a player's in-game social network (including the number of in-game friends who logged in, whether or not the player joined a clan/guild). To facilitate the interpretation of the estimate of social interaction effect, we create Table 6. We discover that when the participation rate among gamers increases 1 percent (e.g. suppose the current rate is 50%, and it increases to 51%), gamers will play 2.21% longer than before, and the consumption of current paid players will increase by 0.48%. These effects are big given that the average day-to-day variation of the social interaction effect in our sample is 2.83%. Partically, if the game participation rate among the free players (the average daily participation rate is 35.8%) could be as high as the rate among the paid players (50.0%), the consumption by paid players will increase by 6.8% percent, the average hours played among all players will increase by 31.4%.

The estimate of the average coefficient α_{1i} among paid players that affects the price elasticity is 0.114, which by eq. (9) implies that the price elasticity at the intensive margin is approximately -1.13.[20] The high price elasticity is consistent with the low frequency of purchasing in data (on a random day, the percentage of purchasing among the group of paid players is 4.2% on average).

To demonstrate the goodness of fit of our model, in Figure 3, we plot the distribution of average daily hours played (panel A) together with the distribution of estimated fixed effect $\gamma_{1i} - \omega_{1i}$ (panel B and C). It is interesting to note that the distribution of fixed effect shares similar pattern with the distribution of actual hours played in data—they both have long tail and multimodal. In addition, in Table 7 we compare some key statistics calculated from using the actual data and from simulated data using our estimated model. Overall, the fitness is satisfying. The fitness of

[20] When $D_{it}^X = 1$, eq. (9) implies that the partial derivative $\partial Y_{it}^X/\partial \ln P_t = -1/(1 - \alpha_1) = -1.13$. Note that $Y_{it}^X = \ln(1 + X_{it})$, hence $\partial Y_{it}^X \approx \Delta X_{it}/X_{it}$ when X_{it} is large. In our sample, $\mathrm{E}(X_{it} \mid D_{it}^X = 1)$ is 43.6.

Table 5: Estimation Results of Hours Played and In-Game Purchase

	Parameters	Estimates
Step 1: Participation Choices[1]	Weekend	-0.152
		(0.005)
	Holiday	0.439
		(0.007)
	Social Interaction Effect	3.706
		(0.011)
	Weekend $\times$ Free Grp	0.076
		(0.005)
	Holiday $\times$ Free Grp	-0.622
		(0.008)
	Hrs Played Before	0.444
		(0.004)
Step 2: Hours Played[2]	Login Cost (ω_{login})	0.524
		(0.002)
	Mean of α_{1i} among paid players	0.108
		(0.009)
	$\tilde{\sigma} = \text{Std}(\varepsilon_{it}^{qual} - \varepsilon_{it}^{time})$	0.218
		(0.001)
Step 3: Purchase Choice	Mean of γ_1 among paid players	-5.955
		(0.639)
	$\text{Std}(\varepsilon_{it}^{qual})$	7.663
		(0.520)
	Num. obs.	2784467

[1] All reported parameters in step 1 are scaled by $\tilde{\sigma}^{-1}$. The other control variables include hours played yesterday, in-game expenditure yesterday, account tenure, account tenure-squared, current level in game, number of in-game friends who logged in, whether or not the player joined a clan/guild. Analytical bias correction was applied to fixed effect probit model in the first step (Fernández-Val, 2009). Figure 3 shows the distribution of the estimated fixed effect $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$.

[2] The IVs for in-game consumption include log of price index, log of price index yesterday, average spending among friends, average spending in the player's guild (0, if the player did not join a guild). The F test statistic for the first step of 2SLS is 873.68.

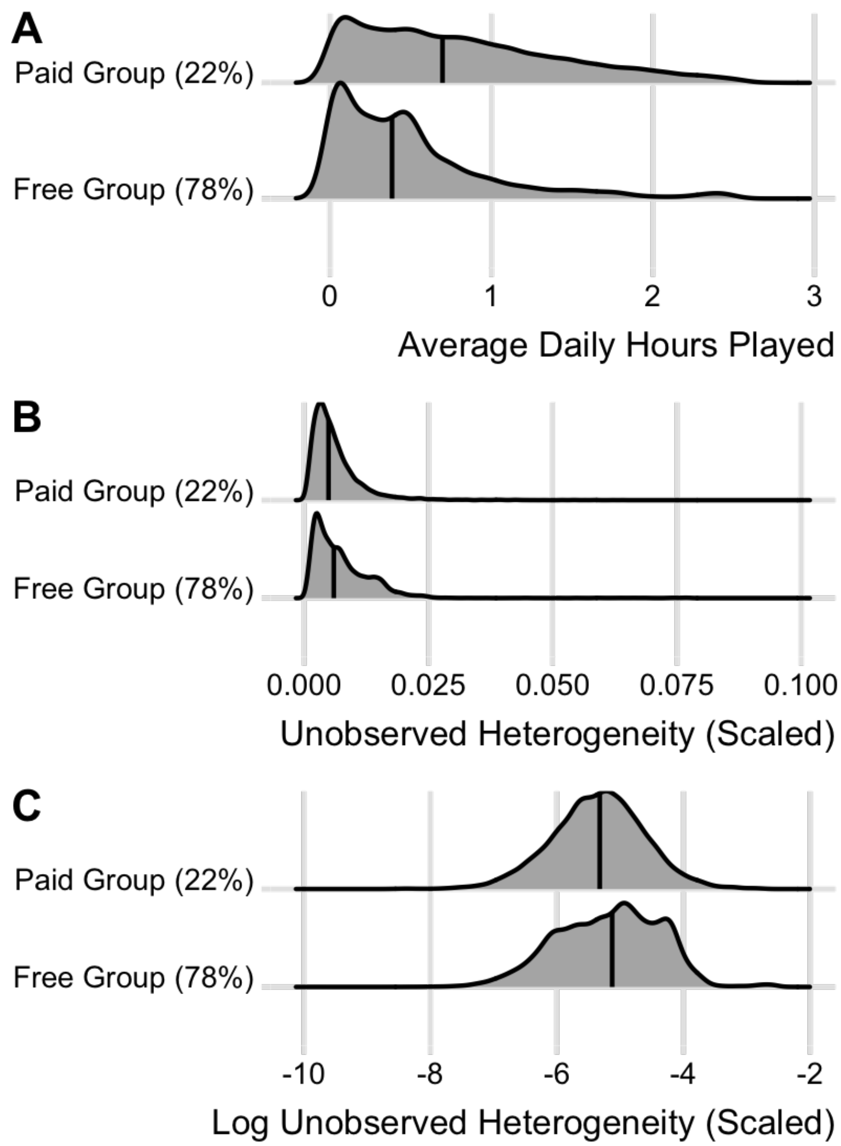

Figure 3: Kernel Density of Daily Hours and Unobserved Heterogeneity

Note: The log of unobserved heterogeneity refers to the fixed effect $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$ in panel probit model of the game participation model eq. (11). "Scaled" means that we estimated $\tilde{\sigma}^{-1}(\gamma_{1i} - \omega_{1i})$. The average estimated fixed effect is -5.22, and average the standard error of estimated fixed effect is 0.20.

Table 6: Effect When the Social Interaction Effect Increases 1%

	Elasticities (%)
Hours played	2.212
Likelihood of game participation ($\Pr(D_{it}^{H}=1)$)	1.927
Hours played (intensive margin)[2]	1.334
Consumption	0.479
Likelihood of purchase ($\Pr(D_{it}^{X}=1)$)[3]	2.219
Consumption (intensive margin)	1.963

[1] The elasticity of hours played at the intensive margin refer to the percentage of change of hours played as the social interaction effect increase by 0.01 for those players who had decided to play, i.e. $D_{it}^{H}=1$. The elasticity of consumption at the intensive margin is defined similarly for those who had decided to purchase.

[2] The elasticities of purchase and consumption are for the group of paid players only. Free players never purchase, hence their elasticities are not identified.

expected consumption (rows of $\Pr(X>0)$ and $\mathrm{E}(X)$) is worse than the fitness of hours played. This is likely to due to low frequency of purchasing in data. Using the estimated model, we can do counterfactual analysis with different levels of social interaction effect, and the results are reported in Table 4 and have been discussed earlier.

6 The Value of Free Goodies

We have showed that the participation of free players creates value for a game company by increasing the consumption of paid players. One natural question is how can we incentivize the free players to participate in the game? One approach is to reward players with in-game items for sale if they play the game longer. So that players are motivated to "farm" in-game rewards by playing the game longer. The rationale why such a reward program could be profitable for a game company is that it incentivizes both free and paid players to play longer. Longer hours played not only increase the social interaction effect but also makes the in-game purchases more rewarding since players use the purchased goods more often or longer. On the other hand, "farming" is a substitute of "paying" in such a reward program. Depending on the generosity of the reward program, players may just want to farm a weapon rather than buying it. The good thing is that the generosity is under the

Table 7: Goodness of Fit: Paid Group of Players

	Data[1]	Model Simulation[2]
$\Pr(X > 0)$	0.042	0.033
$\Pr(H > 0)$	0.506	0.522
$\mathrm{E}(Y_{it}^X \equiv \ln(1 + X))$	0.14	0.14
$\mathrm{E}(X)$	1.82	2.34
$\mathrm{E}(Y_{it}^H \equiv \ln(1 + H))$	0.41	0.43
$\mathrm{E}(H)$	0.81	0.70

[1] These values are computed using all data about the group of paid players.

[2] The values reported here are the average based on 20 sets of simulation from our model. For each simulated (Y_{it}^X, Y_{it}^H), we excluded the simulated samples that were above 99% percentile of (Y_{it}^X, Y_{it}^H).

control of a game company, who can change of the value of the reward. Below, we will modify the basic model to incorporate the above reward program, and use our estimates to evaluate the effect of the program parameterized at different level of generosity.

In order to model the reward program, we modify the subjective quality specification to become the following,

$$\ln \psi_{it} = \alpha_{1i} \ln(1 + X_{it} + \kappa H_{it}) + \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t + \varepsilon_{it}^{qual}, \tag{12}$$

where κ is a new parameter reflecting the generosity of reward, and κ is under the control of the game studio. This specification says that κH_{it} is a perfect substitute of X_{it}. By construction, the price index P_t on the first day is 1. Taking $\kappa = 0.5$ for example, the above specification is saying that the reward a player will get after playing 2 hours of the game, $0.5 \times 2 = 1$, is equivalent to an in-game item for sale that costs ¥1 RMB (on the first day when $P_t = 1$) in terms of utility. A greater κ means more generous reward for the same amount of time playing the game.

Given our estimates of players' preference, we can solve consumers optimal choices of hours played and in-game consumption corresponding to any particular value of κ. The new quality specification eq. (12) changes the marginal utility of hours H_{it} and consumption X_{it}. Using similar technique, we can solve the player's utility maximization problem. First, the binary participation choice is unchanged, and we still have eq. (3) as our participation rule. Second, provided that the

Table 8: Effects of Different Generosity on Hours Played and Consumption

Generosity	Mean Consumption (All Players)	Mean Consumption (Paid Players)	Mean Hours	$\Pr(X > 0)$	$\Pr(H > 0)$
$\kappa = 0$	3.65	115	0.857	0.0382	0.603
$\kappa = 0.25$	3.71	125	0.978	0.0360	0.603
$\kappa = 0.50$	3.72	133	1.07	0.0344	0.604
$\kappa = 0.75$	3.68	138	1.14	0.0330	0.604
$\kappa = 1.00$	3.64	145	1.20	0.0315	0.603
$\kappa = 3.00$	3.58	184	1.48	0.0256	0.603

Note: For each value of the generosity parameter κ, we simulate each player's (free and paid) choices. The values reported are the average based on 50 replications.

player chose to participate in the game, her hours played H_{it} satisfies the following equation,

$$(\gamma_{1i} - \omega_{1i}) + \gamma_2' Z_{it}^{qual} - \omega_2' Z_t^{time} + \beta S_t + \tilde{\sigma}\varepsilon_{it}^* - (1 - \alpha_{1i}) \ln(1 + X_{it} + \kappa H_{it}) \\ + \ln\left[\alpha_{1i}\kappa Y_{it}^H + (1 + X_{it} + \kappa H_{it})(1 + H_{it})^{-1}\right] = 0. \quad (13)$$

Note that when $\kappa = 0$, the above equation becomes our current decision rule of hours played in eq. (6). Third, the binary purchase rule is

$$D_{it}^X = \boldsymbol{I}\left[\ln \alpha_{1i} + \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t + \ln Y_{it}^{H*} - (1 - \alpha_{1i}) \ln(1 + \kappa H_{it}^*) - \ln P_t + \varepsilon_{it}^{qual} > 0\right], \quad (14)$$

where H_{it}^* is the hours played when $X_{it} = 0$, and $Y_{it}^{H*} = \ln(1 + H_{it}^*)$. H_{it}^* can be solved from eq. (13) by letting $X_{it} = 0$. Fourth, the consumption satisfies the following condition,

$$\ln \alpha_{1i} + \gamma_{1i} + \gamma_2' Z_{it}^{qual} + \beta S_t + \ln Y_{it}^H - (1 - \alpha_{1i}) \ln(1 + X_{it} + \kappa H_{it}) - \ln P_t + \varepsilon_{it}^{qual} = 0. \quad (15)$$

We need to solve (X_{it}, H_{it}) from eq. (13) and eq. (15) numerically. One issue is that in order to solve $(D_{it}^H, D_{it}^X, H_{it}, X_{it})$ from the above equations, we need to have α_{1i} and γ_{1i}. This is not a problem for paid players, for whom we have estimated their α_{1i} and γ_{1i}. However, for free players, who never purchased in the game, we cannot identify their α_{1i} and γ_{1i}. We need to assign certain values for these free players, and we used mean of α_{1i} and γ_{1i} among paid players in our analysis below. This could create upper bias in the estimation of in-game purchases.

Assigning different values to the generosity parameter κ, we simulated players choices, and Table 8 reports our findings. We found that when a player will receive an in-game item that would cost ¥1 RMB (about \$0.15 USD) in microtransaction after playing 2 hours, the program creates the largest increase in sales. The average in-game purchases will increase by 1.64% percent, and the hours played will increase by 24.85%. Note that since these rewarded items have been created in the first place, such 1.64% increase in sales are pure profits. The impact on the in-game purchases is two folded. First, the reward program parameterized at the optimal generosity will indeed decrease the likelihood of purchasing (extensive margin) from 3.82% to 3.44%. This is likely because 32% of sales of in-game items in our data cost less than ¥10 RMB (about \$1.5 USD). In the presence of such reward program, players who intended to spend a little money will now just play the game longer in exchange for the free items. Second, the program increases the sales by 15.6% at the intensive margin (i.e. for those who have decided to purchase). There are multiple reasons why we observe such a big impact. (a) Such reward program increases one's own hours played, which is complementary to in-game purchases according to our findings. (b) Such reward program increases the game participation of every player, free or paid. A more active gaming community also incentivizes players to spend more on games. Last but not least, when the reward program is too generous, players will just obtain in-game items by playing the game rather than purchasing. For example, we found that if a player would receive an in-game item that would cost ¥1 RMB by microtransaction after playing one rather the optimal two hours, the decrease in the likelihood of purchasing is so big (from 3.82% to 3.15%) that the average spending is lower than the case without the reward program (from ¥3.65 RMB to ¥3.64 RMB). This will guide managers to design their reward program, otherwise the game can be a hit without being profitable.

7 Conclusion

We develop a structural model to study consumers behavior of playing hours and in-game purchases in the context of an MMO game platform where players can play the game for free but will be charged for optional add-on products such as virtual weapons, pieces of equipment, and costumes. This model helps us understand consumers' demand for virtual goods in a game, which is essential for pricing, customer retention etc. Our model takes account of the social interactions within a gaming community, and our estimates reveal the great value of free players whose game participation creates positive network externality. Using our model, we also study a counterfactual reward program that improves players' engagement. Though we limit our attention to the MMO games, the proposed model can be applied to other video games and freemium markets, where customers will make simultaneous decisions about the usage and consumption of extra paid goods or service.

APPENDIX

A Price Index

Like the real economy, there are many (more than 2,200) virtual goods in our game. We create a single daily aggregate price index to measure the fluctuation of daily prices in the game. Following the standard practice of calculating the Consumer Price Index, we use the Laspeyres price index approach.

To describe our formula, suppose there are K goods in the game. Let $p^t = (p_1^t, \ldots, p_K^t)'$ denote the vector of the prices of these K goods on day $t = 1, \ldots, T$. For day 1, let $q^1 = (q_1^1, \ldots, q_K^1)'$ denote the demand quantities of the K goods on day 1. Given the price p^1 and quantity q^1 on day 1, our price index is

$$P_t \equiv \frac{(p^t)'q^1}{(p^1)'q^1} = \frac{\sum_{k=1}^K p_k^t q_k^1}{\sum_{k=1}^K p_k^1 q_k^1}. \tag{A.1}$$

By construction, $P_1 = 1$ for day 1. Figure A.1 shows the time series of the constructed price index. We observe two big drops in price index (January 27, 2011 and March 23, 2011). The first one (January 27th) was caused by the promotion for the arrival of Chinese New Year, February 3, 2011. The second one (March) was a promotion accompanying the release of a major downloadable content (DLC).

B Proofs

Proof of Theorem 1. For exposition simplicity, we mute the subscript i, t, and (i,t) in the proof. Define

$$A = \frac{\ln \alpha_1}{1-\alpha_1} + \left(\frac{1}{1-\alpha_1}\right)(\gamma_1 + \gamma_2' Z^{qual} + \beta S) - \left(\frac{1}{1-\alpha_1}\right)\ln P + \frac{\varepsilon^{qual}}{1-\alpha_1}$$
$$B = (\gamma_1 - \omega_1) + \gamma_2' Z^{qual} - \omega_2' Z^{time} + \beta S + (\varepsilon^{qual} - \varepsilon^{time}).$$

So that for paid players,

$$D^H = \boldsymbol{I}(B > 0), \qquad \tilde{Y}^H = \alpha_1 Y^X + B, \qquad \tilde{Y}^X = A + \left(\frac{1}{1-\alpha_1}\right)\ln Y^H. \tag{B.1}$$

Case 1: $B \leq 0$. The player will not participate in the game, hence $D^H = 0$. Then it is easy to see that $Y^H = Y^X = 0$ is the only solution.

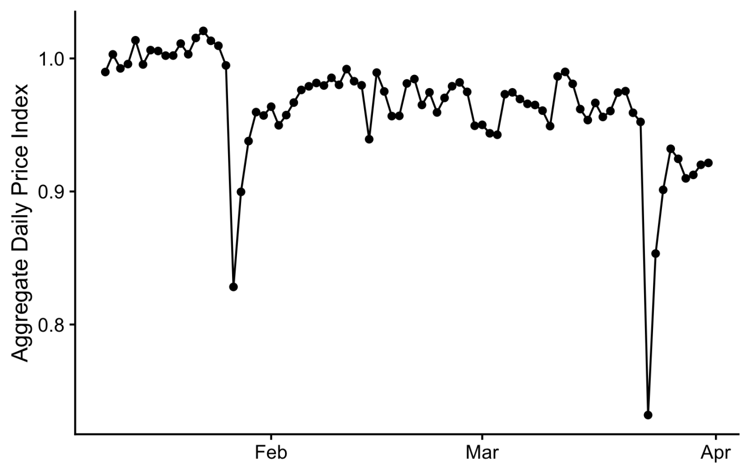

Figure A.1: Aggregate Price Index of In-Game Goods

Note: The price index P_t is calculated using eq. (A.1). There are two big drops in price index (January 27, 2011 and March 23, 2011). The first one (January 27th) was caused by the promotion for the arrival of Chinese New Year, February 3, 2011. The second one (March) was a promotion accompanying the release of a major downloadable content.

Case 2: $B > 0$. When $B > 0$, we must have that $\tilde{Y}^H > 0$ since $\alpha_1 Y^X \geq 0$, hence $D^H = 1$ and $Y^H = \tilde{Y}^H$. Substituting Y^H with $\tilde{Y}^H$ in eq. (B.1), we have

$$\tilde{Y}^X = A + \left(\frac{1}{1-\alpha_1}\right) \ln(\alpha_1 Y^X + B).$$

We then need to verify whether or not

$$Y^X = D^H \times D^X \times \max\left(0, \tilde{Y}^X\right)$$

has a unique solution, when $\tilde{Y}^X$ is defined as above. There can be two scenarios. In first scenario, we have $A + (1-\alpha_1)^{-1}\ln(B) < 0$. In this case, $D^X = 0$ and $Y^X = 0$ is the only solution. The player will decide not to purchase, and the hours played satisfies $Y^H = B$. In the second scenario, we have that $A + (1-\alpha_1)^{-1}\ln(B) \geq 0$, and we know that $D^X = 1$, and $Y^X = 0$ cannot solve the above equation. The solutions, if exist, must be greater than zero, hence $\tilde{Y}^X > 0$. The solutions must satisfy

$$f(Y^X) \equiv Y^X - A - \left(\frac{1}{1-\alpha_1}\right) \ln(\alpha_1 Y^X + B) = 0.$$

Note that

$$\lim_{y \to 0} f(y) = -\left[A + \left(\frac{1}{1-\alpha_1}\right) \ln(B)\right] \leq 0.$$

and $\lim_{y\to\infty} f(y) = \infty$. Apparently, $f(Y^X)$ is a continuous function of Y^X, hence a solution must exist by the Intermediate Value Theorem. Next, it is easy to show that $f(y)$ is strictly convex function when $\alpha_1 \in (0,1)$, so that the root of $f(y) = 0$ for $y > 0$ is unique. The second order derivative of $f(y)$ is

$$\frac{\mathrm{d}^2 f(y)}{\mathrm{d}\, y^2} = \alpha_1 \left(\frac{\alpha_1}{1-\alpha_1}\right) \frac{1}{(\alpha_1 y + b)^2} > 0,$$

provided that $\alpha_1 \in (0,1)$. $\square$

Proposition B.1 below shows that in general, if (a) the in-game purchases do not affect the utility of gaming provided that the player did not actually spend time on playing, and (b) the in-game purchases increases the marginal utility of playing games for one extra unit of time, the utility function $U(X_{0it}, X_{it}, H_{it})$ cannot be concave. Consequently, one must be careful about the existence of multiple equilibria, and some equilibrium refinement is necessary to achieve point identification.

Proposition B.1. *Consider the general quasilinear utility function,*

$$U(X_0, X, H) = X_0 + v(X, H).$$

If $v(X,H)$ is a twice-differentiable function, and

(i) $\lim_{H\to 0} v(X,H) = c$ *for any constant c;*

(ii) $v'_H(X,H)$ *is increasing in* X*, where* $v'_H(X,H) = \partial v(X,H)/\partial H$*,*

$v(X,H)$ *is not globally concave.*

Proof. A twice differentiable function $v(X,H)$ is strictly concave if and only if its Hessian matrix is negative definite. In particular, it is necessary that the determinant of the Hessian is positive. The determinant is

$$\det_v = v''_H(X,H) \times v''_X(X,H) - (v''_{HX}(X,H))^2.$$

By the first condition, $\lim_{H\to 0} v(X,H) = c$, we know that $v''_X(X,H) = 0$ when $H \to 0$. By the second condition, $v''_{HX}(X,H) > 0$ for all (H,X). Together we conclude that $\det_v$ is negative when $H \to 0$, so that v is not concave. $\square$

Proof of Theorem 2. The results of Matzkin (2015) is not directly applicable here, because they require that endogenous variables are strictly monotone in the random errors. Such monotone conditions are not satisfied here due to the presence of corner solutions. Noting that the latent endogenous variables are strictly increasing in the error terms, it is easy to modify her arguments to prove the identification here. Consider a slightly more general model here, which includes our video game as a special case, $Y_1 = \max(\tilde{Y}_1, 0)$ and $Y_2 = \max(\tilde{Y}_2, 0)$, where

$$\tilde{Y}_1 = \beta_1 Y_2 + \alpha_1 X_1 + \varepsilon_1,$$
$$\tilde{Y}_2 = \beta_2 \ln(Y_1) + \alpha_2 X_2 + \varepsilon_2.$$

Following Matzkin's notation, we can write

$$\begin{aligned} \varepsilon_1 &= r^1(\tilde{Y}_1, Y_2, X_1) = \tilde{Y}_1 - \beta_1 Y_2 - \alpha_1 X_1 \\ \varepsilon_2 &= r^2(\tilde{Y}_2, Y_1, X_2) = \tilde{Y}_2 - \beta_2 \ln(Y_1) - \alpha_2 X_2. \end{aligned} \tag{B.2}$$

Define $r(y, x)$ the vector of the functions r^1, r^2. Because the latent variables are only both observed when $Y_1 > 0, Y_2 > 0$, we consider the identification using the information when $Y_1 > 0, Y_2 > 0$. For simplicity, denote the condition $(Y_1 > 0, Y_2 > 0)$ by $Y > 0$. Note that we have

$$f_{Y|X=x,Y>0}(y) = f_\varepsilon(r(y, x)) \left| \frac{\partial r(y, x)}{\partial y} \right|,$$

where $|\partial r(y, x)|\partial y$ is the absolute value of the Jacobian determinant of $r(y, x)$ with respect to y. Note that this equation is the same as eq. (2.3) of Matzkin (2015), except that we consider $f_{Y|X=x,Y>0}$. Using her arguments, we can identify the ratio of derivatives $r^g_{y_j}(y, x_g)/r^g_{x_g}(y, x_g)$ for $g = 1, 2$. Using our parametric specification of eq. (B.2), we can identify $\alpha_1, \alpha_2, \beta_1, \beta_2$. □

References

Amemiya, Takeshi. 1974. "Multivariate regression and simultaneous equation models when the dependent variables are truncated normal." *Econometrica*, 999–1012.

Brock, William A, and Steven N Durlauf. 2001. "Discrete choice with social interactions." *The Review of Economic Studies*, 68(2): 235–260.

Cheng, Hsing Kenneth, and Yipeng Liu. 2012. "Optimal software free trial strategy: The impact of network externalities and consumer uncertainty." *Information Systems Research*, 23(2): 488–504.

Dubé, Jean-Pierre H, Günter J Hitsch, and Pradeep K Chintagunta. 2010. "Tipping and concentration in markets with indirect network effects." *Marketing Science*, 29(2): 216–249.

Durlauf, Steven N, and Yannis M Ioannides. 2010. "Social interactions." *Annual Review Economics*, 2(1): 451–478.

Fernández-Val, Iván. 2009. "Fixed effects estimation of structural parameters and marginal effects in panel probit models." *Journal of Econometrics*, 150(1): 71–85.

Foubert, Bram, and Els Gijsbrechts. 2016. "Try it, you'll like it—or will you? The perils of early free-trial promotions for high-tech service adoption." *Marketing Science*, 35(5): 810–826.

Graham, Bryan S, and Jinyong Hahn. 2005. "Identification and Estimation of the Linear-in-Means Model of Social Interactions." *Economics Letters*, 88(1): 1–6.

Gu, Xian, PK Kannan, and Liye Ma. 2018. "Selling the premium in freemium." *Journal of Marketing*, 82(6): 10–27.

Hahn, Jinyong, and Whitney Newey. 2004. "Jackknife and analytical bias reduction for nonlinear panel models." *Econometrica*, 72(4): 1295–1319.

Hoang, Ai-Phuong, and Robert J Kauffman. 2018. "Content sampling, household informedness, and the consumption of digital information goods." *Journal of Management Information Systems*, 35(2): 575–609.

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

Jo, Wooyong, Sarang Sunder, Jeonghye Choi, and Minakshi Trivedi. 2020. "Protecting consumers from themselves: Assessing consequences of usage restriction laws on online game usage and spending." *Marketing Science*, 39(1): 117–133.

Katz, Michael L, and Carl Shapiro. 1985. "Network externalities, competition, and compatibility." *American Economic Review*, 75(3): 424–440.

Kline, Brendan, and Elie Tamer. 2020. "Econometric analysis of models with social interactions." In *The Econometric Analysis of Network Data*. 149–181. Elsevier.

Lee, Clarence, Vineet Kumar, and Sunil Gupta. 2019. "Designing Freemium: Strategic Balancing of Growth and Monetization." Working paper, Samuel Curtis Johnson Graduate School of Management, Cornell . . . .

Lee, Lung-fei, Ji Li, and Xu Lin. 2014. "Binary choice models with social network under heterogeneous rational expectations." *Review of Economics and Statistics*, 96(3): 402–417.

Lee, Young-Jin, and Yong Tan. 2013. "Effects of different types of free trials and ratings in sampling of consumer software: An empirical study." *Journal of Management Information Systems*, 30(3): 213–246.

Lewbel, Arthur. 2007. "Coherency and completeness of structural models containing a dummy endogenous variable." *International Economic Review*, 48(4): 1379–1392.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

Matzkin, Rosa L. 2015. "Estimation of nonparametric models with simultaneity." *Econometrica*, 83(1): 1–66.

Nair, Harikesh. 2007. "Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games." *Quantitative Marketing and Economics*, 5(3): 239–292.

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

Shi, Zijun, Kaifu Zhang, and Kannan Srinivasan. 2019. "Freemium as an optimal strategy for market dominant firms." *Marketing Science*, 38(1): 150–169.

Tamer, Elie. 2003. "Incomplete simultaneous discrete response model with multiple equilibria." *The Review of Economic Studies*, 70(1): 147–165.