STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2021–2022

Network Effects in the PC Gaming Industry

Haochen Li (41657)

Abstract. Various empirical papers have proved the existence of network effects in many different industries and developed methods to quantify the network effect. The thesis establishes a structural model for the demand system of the PC gaming industry and estimates the reduced-form equations using instrumental variable estimators. The thesis proposes two valid instrumental variables from the labor supply side and the Covid Pandemic. The thesis identifies the existence of indirect network effects in both the platform adoption side and software provision side using the data from the game distributing platform, Steam.

Keywords: Network Effects, Industrial Organization, Digital Market, PC Gaming Industry

JEL: D12, L11, L86

Supervisor:	Andreea Enache
Date submitted:	May 16, 2022
Date examined:	May 25, 2022
Discussant:	Chengkun Li
Examiner:	Karl Wärneryd

Contents

1 Introduction **4**

2 Background: PC Gaming Industry and Platforms **6**

2.1 PC Gaming Industry 6

2.2 PC Gaming Platforms 7

3 Literature Review **9**

3.1 Theoretic Foundation on Network Effects 9

3.2 Network Effects and Video Game Industry 10

3.3 Empirical Researches on Network Effects 11

4 Model of Network Effects **14**

4.1 Hardware Adoption 14

4.2 Software Provision 16

4.3 Estimation Equations 18

4.4 Statistical Tests for IV Estimator 19

5 Data **20**

5.1 Data Sources 20

5.2 Data Description 21

6 Results and Discussions **25**

6.1 Platform Adoption: Basic Model 25

6.2 Software Provision: Basic Model 27

6.3 Platform Adoption: A Valid Instrument from Labor Supply 29

6.4 Software Provision: A Valid Instrument from Covid Pandemic 31

6.5 Robustness Test: Demand System with Weekly New Released Game Titles 33

7 Conclusion **37**

A Appendix **42**

A.1 Time Trend of All Employees in U.S. Computer and Software Industry 42

A.2 Time Trend of Global Weekly Confirmed Cases 42

A.3 An Informal Analysis of Direct Network Effects: CS:GO Players 43

List of Figures

1 Online PC Gaming Market Value Worldwide (2011-2026) 6
2 Digital Video Game Purchases by Store Brand in the U.S. in 2022 8
3 Steam Platform Users by Week (millions) 22
4 Number of Active Titles by Week 23
5 Average Price of Active Titles by Week(deflated by Euro Area CPI (2015 = 100)) . 23
6 Average Age of Active Titles (months) by Week 24
7 All Employees in U.S. Computer and Software Industry by Week 42
8 Global Confirmed Covid Cases by Week 42
9 Log(# Steam Users - black) v.s. Log(# CS:GO Players - red) 43
10 The First Difference of Log(# CS:GO Players) 44
11 Log(# CS:GO Players) and OLS Fitted Line (Time Trend) 45
12 Log(# CS:GO Players) and OLS Fitted Line (Time Trend + lagged(Users)) 46

List of Tables

1 Summary Statistics 21
2 Estimation Results on Platform Adoption with Instruments of Games' and Platform Characteristics 26
3 Estimation Results on Software Provision with Instruments of Games' and Platform Characteristics 28
4 Estimation Results on Platform Adoption with Instruments from Labor Supply Side 30
5 Estimation Results on Software Provision with Instrumetn of Covid Cases 32
6 Estimation Results on Platform Adoption with Weekly New Released Titles 34
7 Estimation Results on Software Provision with Weekly New Released Titles 36
8 Regression Results of ARMA without time trend 47
9 Regression Results of OLS and ARMA with Time Trend 48
10 Regression Results of ARMA with Time Trend and lag(# Users) 49

1 Introduction

The last two decades have witnessed a great proliferation of Internet-based platforms. The successful diffusion of these platforms is often contingent upon the availability of complementary products. For example, the success of online-shopping platforms depends on how many products can be available on the website. Similar statements apply to video game consoles and video games, computer operating systems and software, Music streaming service and music, and so on. The relationship between the adoption of platforms and the availability of complementary products is characterised as network effects.

Farrell and Klemperer (2007) defines the network effects as the case when a user wants compatibility with other users so that they can interact with each other. The interaction creates economies of scope among different participants in the market. The network effects can then be divided into two kinds: direct network effects and indirect network effects. The direct network effects appear among the agents in the same market. For example, a depositor in one bank can benefit from the other depositor in the same bank, since the transfer costs in the same bank is almost zero. There are also negative examples of direct network effects. For example, a car driver may dislike the increase in the number of car drivers in the same commuting routes. The indirect network effects usually appear in a two-sided market: buyers would prefer more sellers on the other side of market, as the variety of the products could increase with the increases of sellers.

When prices cannot fully incorporate the network effects from one person/side's adoption, the network externalities arise. Early literature use the terms "network effects" and "network externalities" in the same way. However, not all network effects are externalities (Liebowitz and Margolis, 1994) . For a single network, the network externalities could cause chicken-and-egg coordination problem and hence cause under-adoption. For competing platforms, the "tipping" and "lock-in" could build entry barrier for new entrants and prevent their entrance.

The video game industry is a classic example of indirect network effects in a two-sided market. Mnay researchers study the hardware/software system between video game console and video games. They find that there are significant indirect network effects in the video game industry. The video game industry includes many sub-sectors, like the console-based video game industry, PC gaming industry, and arcade video games. The mobile games should also be considered as a sub-sector of video game industry. It is estimated that nearly 60 percent of Americans play video games, bringing in annual revenue of over $25 billion for PC gaming alone (in 2016). The current development of the industry could be more prosperous because of its growing trend and the impact of Covid-19. One corresponding example could be the booming of e-sports, which attracts hundreds of millions of audiences all over the world.

In the PC gaming industry, a distributing service named Steam takes up 75 percent of the global market share. As its slogan said, "Steam is the ultimate destination for playing, discussing, and creating games." In 2022, Steam has concurrent online users of 25,449,374 and 7,012,473 are

playing games using the steam services. Steam works as a platform to connect game players and game providers. It is naturally to suspect that the network effects arise within the Steam platform.

This thesis aims to undertake an empirical study on the indirect network effect. The study mainly focuses on the PC gaming industry rather than the whole video game industry. Thanks to the openness of the data associated with the Steam platform, one can easily find data on the users and game titles. Unlike the previous studies using the sales data for the installed base of platform, I can observe the concurrent users in the steam platform directly. That is to say, the main focus of this thesis is the value of usage rather than the value of purchasing.

The main research objective for the thesis is to prove the existence of indirect network effects in the PC gaming industry. If the thesis discovers the existence of indirect network effects, it is of interest to measure the magnitude of the network effects.

To prove the existence of indirect network effects in the PC gaming industry, the thesis establish a model with structural specification, following Berry (1994) and Nair et al. (2004). Then I estimate the reduced-form equations for the demand system of PC gaming industry. The demand system includes platform adoption equation and software provision equation. The estimator used for the reduced-form demand system is Instrumental Variable estimator.

For both equations of platform adoption and software provision, there are two different specifications of instrumental variables, respectively. The first specification mainly follows Clements and Ohashi (2005), using products' and platform's characteristics as instrumental variables. The statistical tests find that the traditional instrumental variables are invalid and correlated with the error terms. Thus, I propose two new instrumental variables from the labor supply side and Covid Pandemic, which are tested to be more valid. The thesis also conduct a robustness check using higher-frequency variables as dependent variable and independent variable.

Finally, the results show that the indirect network effects in the PC gaming industry are significant and substantial. On average, 100 titles increase in a year could result in nearly 0.1 percent increase in the market share of steam platform. Similarly, 1 percent increase in number of users could cause 0.7 percent increase in number of game titles in a year.

The thesis mainly contribute to the literature in two aspects: first, I don't find previous research study the PC gaming industry using the data of Steam platform; second, I propose two new instrumental variables for the study of video game industry.

However, there are also many flaws with this thesis. The method used by this thesis is not as advanced as the research frontier in empirical industrial organizations. The thesis does not structurally estimated the structural model but rather the reduced-form model. The thesis does not consider dynamics and equilibrium conditions for the demand system.

The remainder of the thesis is organized as: Section 2 discusses the related literature; Section 3 establish the structural model and reduced-form for estimating; Section 4 presents the data; Section 5 shows the results; section 6 concludes.

2 Background: PC Gaming Industry and Platforms

2.1 PC Gaming Industry

The PC gaming industry, or computer game industry, is defined as a sub-industry of the whole video game industry, People can play a game via their personal computer rather than a game console like XBOX, or arcade game machine. The main feature for a PC game to be different from a console game is the controller, meaning people use a mouse and keyboard instead of a traditional gamepad. The difference in game controller allows players to operate more precisely and result in the booms of certain game types, FPS (First-Person Shooting) and RPG (Real-Time Strategy).

Along with the growth of hardware technology and operating system software like Microsoft Windows, computer games started to gain popularity among consumers of home video game consoles after the late 1990s. Many people born after 1995 played their first video game "Minesweeper", a pre-installed game with Microsoft Windows after 1992, on computers. Now, instead of using disks, most computer games are distributed digitally, which increases flexibility and efficiency for consumers.

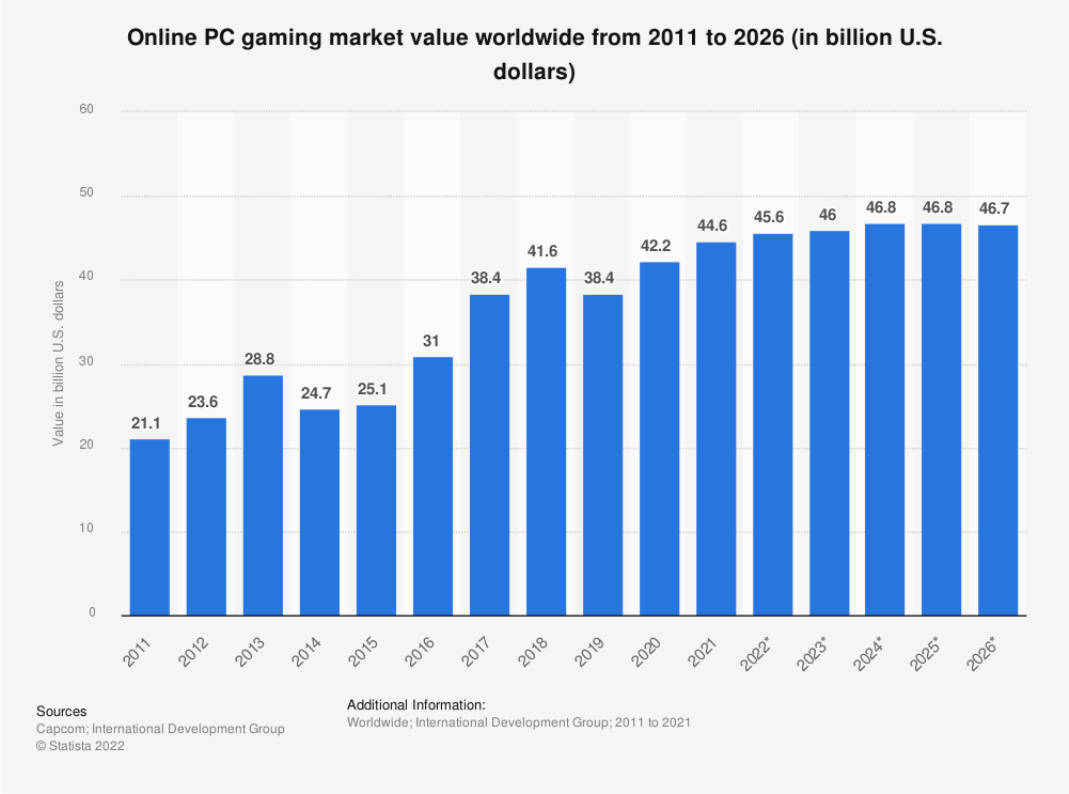


Figure 1: Online PC Gaming Market Value Worldwide (2011-2026)

Personal computers are recognized to have high compatibility. Personal computers are compatible with different game controllers and the operating systems are compatible with older versions.

Compared to other platforms, the share of game developers for personal computers is the most [1]. For game titles first released on the other platforms, game developers often publish these games again after the exclusive periods to make more profits.

Figure 1 illustrate the market value of the online PC gaming market. From 2011 to 2021, the market has grown from 21.1 billion U.S. dollars to 44.6 billion U.S. dollars, with an increase of more than 100 percent. Along with the recovery of game developers and publishers from the pandemic, and increased players attracted by eSports and streaming, the market value is expected to continue to grow.

2.2 PC Gaming Platforms

For nowadays game players who want to buy legitimate-version games, it seems the only choice is to buy games from online game stores. Since the early 2010s, most PC games are distributed digitally, and physical copies of video games seem only to exist for video game consoles.

There are several alternatives for game players to search for games: Steam, Epic Game Store, GOG Galaxy, Origin, Battel.net, Uplay, and Windows Store. Some of these online stores include many third-party games, like Steam and Epic Game Store. However, some only provide services of their first-party games, like Battle.net from Blizzard and Uplay from Ubisoft. Figure 2 shows a survey result of people's game purchases by store brand, including both PC games and console games.

Steam, developed by Valve, is the dominant platform in the PC game market. It come along with Counter-Strike, one of the most popular FPS games in history, in 2003. After nearly a decade, Steam has grown into the most popular video game digital distribution service, where people can purchase over 50 thousand games. Instead of solely providing purchase options, Steam also provides automatic updates, social community, modification workshops, game reviews, and guides. For multiplayer online games, Steam also allows players to quickly match their competitors to start the game as soon as possible.

The main source of Steam's revenue is the 30 percent price cut of every game title sold. In a lawsuit in April 2021, Wolfire Games v.s. Valve Corp., Wolfire Games argued that Steam was utilizing its monopoly power and extracting excessive profits from game developers[2]. The lawsuit estimated that Steam took up 75 percent of the 30 billion U.S. dollars market for PC game sales in 2021. The estimated market share is in line with previous reports[3]. Steam is still ambitious to expand its business to more areas. In 2014, Steam launched its virtual reality service SteamVR. In 2022, Steam developed its own handheld console Steam Deck to run games from the Steam store.

Although Steam is in a dominant position, challengers are threatening its dominance. Epic

[1] Game developers focus on selected gaming platform projects 2022 | Statista
[2] Humble Bundle creator brings antitrust lawsuit against Valve over Steam | Ars Technica
[3] Valve Lines Up Console Partners in Challenge to Microsoft, Sony - Bloomberg (archive.org)

Game Store, developed by Epic Games, launched in December 2018. Epic Games built the popular Unreal Engine for game developers and made success on its first-party game, Fortnite. Tim Sweeney, founder, and CEO of Epic Games thought it was too high for Steam to charge 30 percent of the price, and decided to develop Epic Games' storefront[4]. In 2022, Epic Games sued Apple Inc. for the same reason.

As a new entrant to the market, Epic Games Store decides to establish its market foundation by both securing exclusive games and offering free games to players.The Epic Game Store also decided to lower the price cut to 12 percent. In Epic Game Store's first 10 months, it spent 11.6 million U.S. dollars to distribute 104 million free copies of 42 different titles to players. The investment is continuing. At the same time, Epic Game Stores has grown fast, with 13.2 million concurrent users in 2021, half of the size of Steam. Epic Games is looking forward to taking up 35 to 50 percent of the PC gaming market by 2024.

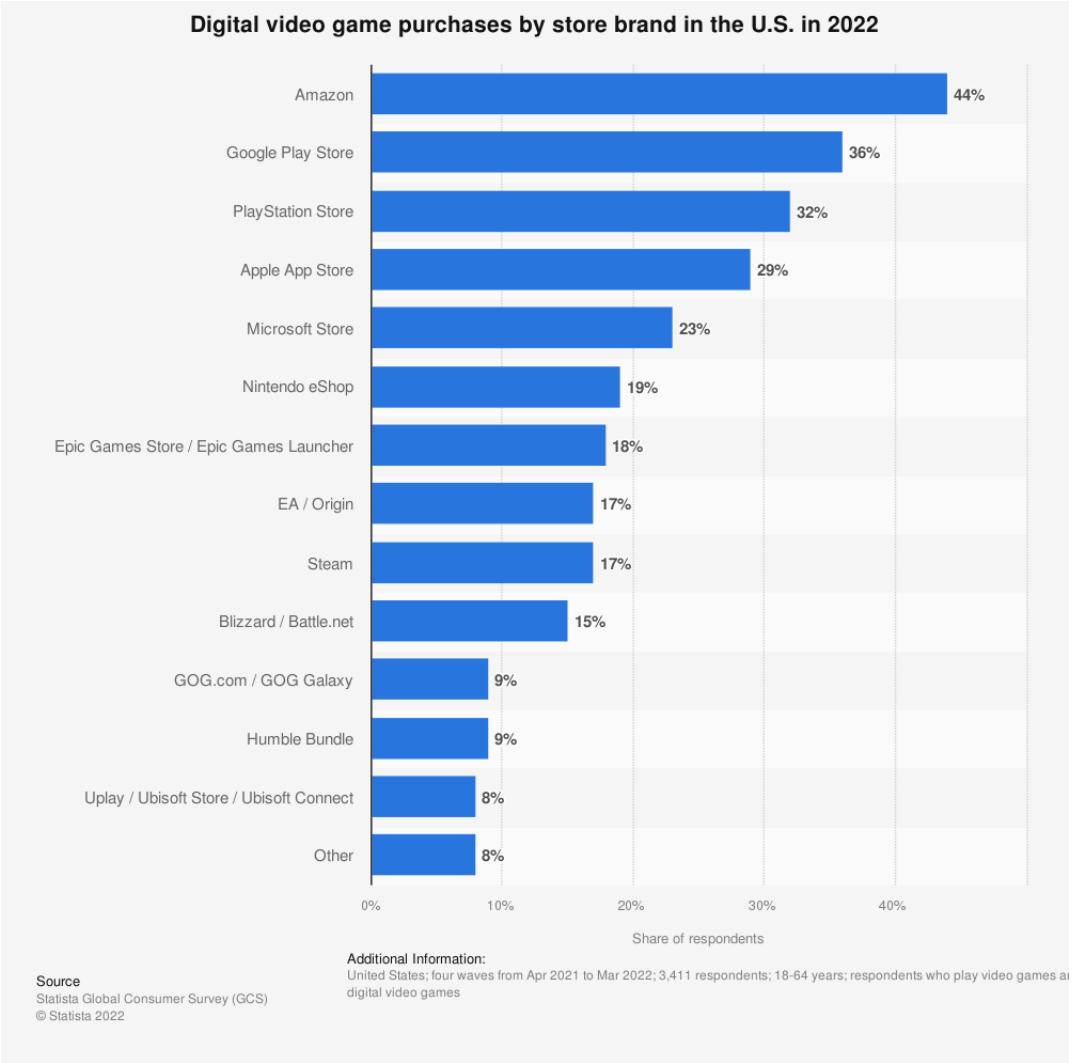


Figure 2: Digital Video Game Purchases by Store Brand in the U.S. in 2022

[4]Epic thinks EGS could make up 35-50% of the PC gaming market by 2024 | Ars Technica

3 Literature Review

3.1 Theoretic Foundation on Network Effects

The phenomenon of network effects is evident, and the discussion threads can be traced back 50 years ago. Economic researchers have discussed the theoretical framework of the network since the early 1980s. Katz and Shapiro (1985) developed their classic model, which captures the network effects associated with consumers' expectations. They set up the conditions where consumers maximize their surplus depending on purchasing brands, and firms maximize their profit by setting the price. Then the model is solved for a fulfilled expectations Cournot equilibrium, and therefore welfare can be calculated. Other than the relationship between consumer purchase decisions and a firm's pricing decision, Farrell and Saloner (1986) also discuss the relationship between installed base and compatible goods. They build a model of users' entrance decisions to explain the possibility of 'excess inertia' or 'excess momentum' when adopting a new technology standard. After the two crucial theoretical frameworks on complementary goods and technology standard,Church and Gandal (1992) propose the first theoretical model of indirect network effects, which examine how the provision of software firms determines the performance of hardware technologies. Their model shows that standardization becomes the equilibrium outcome, and there are multiple equilibria.

Rochet and Tirole mainly focus on the two-sided market and platform competition and include network effects in their studies. Rochet and Tirole (2002) develop a framework for the payment card market, including banks, merchants, and consumers. Their study focuses on the role of the interchange fee or 'access' charge paid by the merchants' banks to the banks who issue cards. They find two scenarios if the payment card association selects the interchange fee: optimal social outcome or overprovision of product. Rochet and Tirole (2003) build a model of platform competition under different business structures, profit-maxing or non-profit-leading. The article compares cases of various industries, including video games, operating systems, live streaming, and so on. Rochet and Tirole (2006) discuss the two-sided market of usage and membership externalities. The model also extended the framework to allow for payments between users. They identify a few factors that determine a two-sided market: transaction costs, platform-imposed constraints, and membership fixed costs.

Rysman (2009) discusses the two-sided market and the platform economics from two aspects: firms' strategies of pricing and openness and public policy of antitrust and regulation. Hagiu and Hałaburda (2014) study the relationship between levels of information on platforms. They develop a model to incorporate market structures of both monopoly and competition. They find that intense information negatively affects the platform through price competition, reflecting that part of the market power of two-sided firms is sourced from asymmetric information. Belleflamme and Peitz (2019) extended the discussion of Hagiu and Hałaburda (2014) into the disclosure of price. They propose a framework where the fee charged to one side is not fully known to another

side of the market. In this case, the users' expectations about participation levels can then determine the market outcomes.

3.2 Network Effects and Video Game Industry

The development of the video game industry is along with the result of the study on network effects. Even though Katz and Shapiro (1985) take the video game industry as an example of network effects, the first empirical paper has not published until 2003. Shankar and Bayus (2003) empirically study network effects in the 16-bit home video game industry between January 1993 and August 1995, where the two main competitors are Nintendo and Sega. They model the network effects as a function of network size (installed base) and network strength (marginal effect of the unit increase on demand). Their findings show that the market has asymmetric network effects between the two competitors. A smaller customer network is associated with higher network strength (Nintendo) and vice versa (Sega).

Many studies take the U.S. video market between 1995 and 2005 as a canonical period, and thus their findings can be compared and organized as guidance for the game industry study. Clements and Ohashi (2005) analyze the importance of indirect network effects in the U.S. 16-bit/32-bit video game market between 1995 and 2002, including eight competing platforms. The paper models the interaction between console adoption decisions by consumers and software supply decisions by firms. By solving the simultaneous equations system, they find that introductory pricing is efficient in the earlier periods of hardware's product cycle and software variety becomes more efficient later. Nair (2007) studies the phenomenon of intertemporal price discrimination by firms. The study proposes a model to include forward-looking consumer behavior and a dynamic pricing algorithm. In the game of forward-looking consumers and forward-looking firms, consumers' behavior significantly affects firms' optimal pricing strategy. The paper finds that firms will bear significant profit loss if they ignore consumers' forward-looking behavior.

Dubé et al. (2010) measure the 'tipping' phenomenon in the U.S. video game console market between 1995 and 2002. 'Tipping' means the market concentration caused by the indirect network externalities. They assume the consumers are single-homing, which is the most case in this market, and follow the framework from Rochet and Tirole (2003) to solve for a Markov perfect Bayesian equilibrium. Their results show that the indirect externality has a significant positive effect on the level of market concentration. Lee (2013) studies software products' vertical integration and exclusivity in the U.S. video game market between 1995 and 2000. As first-party game titles are often hit titles during the product cycle of hardware, the indirect network effects associated with different types of titles could vary. This research builds a dynamic structural model of a two-sided market of consumers' demand for hardware and software and software's demand for hardware platforms. The paper finds that compatibility between software and hardware platforms would increase both software and hardware sales, increasing consumer surplus. Zhou (2017) also studies

the U.S. video game market between 1995 and 2000. The research follows the framework of Lee (2013) and focuses on buyers' purchase decisions and sellers' pricing decisions. The dynamic structure model is solved through a Bayesian Markov chain Monte Carlo estimation approach. The results between models, including dynamic behavior and myopic consumers, differ a lot. Hence, the dynamic behavior has significantly impacted the market for video games.

Most studies on video games have focused on the U.S. video game market between 1995 and 2005, and the data have the same source from the NPD group, which provides market data and statistics. However, there are new samples from other countries and different study periods. Ishihara and Ching (2019) study the used market of Japanese video games. They build a dynamic structural model of consumers' buying and selling decisions. After solving the model, they find that consumer surplus decreases with no used copy market. The elasticities of copies appear to be negatively associated with the retailer inventory.

3.3 Empirical Researches on Network Effects

3.3.1 Newspaper and Magazines

The market of newspapers and magazines are two-sided markets with readers and advertisers. In the market, advertisers' demand for newspapers and magazines can be treated as a function of readership size, which shows indirect network effects. Rysman (2004) studies the Yellow Pages directories. The research treats the Yellow Pages directories as the two-sided intermediary of advertisement and models consumers' demand for advertising and advertisers' demand for readership. The simultaneous equations system is solved through GMM estimation. The results show more directories will increase network effects but decrease market power.

Kaiser and Wright (2006) also study advertisements in German magazines. They present and estimate a two-sided market model, including readership and advertising. Their findings show that advertisers have a higher readership valuation than readers' valuation of the advertisement. Sokullu (2016) gives a nonparametric estimation of the network effects in the German magazine industry. The study finds that the network functions have specifications other than linear or monotonic. Argentesi and Filistrucchi (2007) research Italian newspapers and build a simultaneous equations system of demand and supply where firms jointly maximize their profit. The paper estimates the system through two-stage least squares and measures the market power in the industry. Their empirical results are consistent with the assumption that firms maximize their joint profits.

Chandra and Collard-Wexler (2009) study the effects of mergers in the newspaper industry. They examine the results using the difference-in-difference estimator and find that the market concentration does not necessarily increase the price.Fan (2013) addresses the relationship between mergers and product characteristics by taking the Minneapolis newspaper market as an example. The author develops a structural model of the newspaper industry incorporating adjustments in

price and product characteristics. It finds that mergers lead to changes in price and product characteristics, and ignoring the changes in product characteristics could cause bias in merger effects. Song (2021) build a structural model of the advertising business in the German magazine industry. The findings are along with the previous studies that mergers could lower the price on one side and attract more agents from the other side.

3.3.2 Recordings: CD and DVD

The sound recordings and video recordings industries have a classic hardware/software structure and embody the feature of two-sidedness and network externality. Gandal et al. (2000) conducts a study on the CD players industry between1985 and 1992. The authors identify the elasticity of the availability of software titles by estimating the system of hardware adoption and software supply decisions through instrumental variables. Similar network effects also exist in the DVD players industry. Dranove and Gandal (2003) analyze a particular case where a new technology with compatibility, DIVX, threatens the old DVD format. Although the introduction of DIVX slowed down the sales of DVD players, the effect is short-lived and cannot beat the network effect in the DVD players market.

3.3.3 Banking

Though there are only two participants in the bank industry, depositors and banks themselves, there are network effects thanks to the invention of financial facilities and instruments. Ackerberg and Gowrisankaran (2006) analyze the case of the automated clearing house (ACH) banking industry using a static structural equilibrium model of a two-sided market. The results show that the main driver of the lack of ACH usage is the high fixed costs on the consumer side. Ferrari et al. (2010) study the shared ATM network by estimating a structural model incorporating ATM investment and demand problems. They find that the current investment in ATMs is insufficient. The regulation prohibited cash withdrawal fees and prevented depositors from withdrawing cash from ATMs.

3.3.4 Radio and Theater

Jeziorski (2014) analyses the effect of the merger wave in the two-sided U.S. radio industry between 1996 and 2006. The findings show that increasing market power has a different impact on listeners' and advertisers' welfare, where the former increases slightly but the latter decreases substantially. Caoui (2019) examines the case of the distribution of movies from 35mm film to digital format, where the previously installed bases are not compatible with the new technology. The transition presents the feature of "excess inertia" caused by network externalities.

3.3.5 Internet platforms

Earlier studies mainly focus on hardware/software systems like operating systems. With the development of online platforms and the increased Internet usage, the platform/service systems also attract researchers' attention. The difficulties here are usually not the typical price or quantity in the market but rather just a free-using platform and information-only services. Reshef (2019) examines the effects of entrants on the incumbent firms which incorporate network externalities. The study takes Yelp and Grubhub as an example and shows that the entry decision of firms can expand the market to potential consumers. Through the positive indirect network effects, the high-quality incumbent firms are beneficial from the transition of potential consumers to actual consumers. Farronato et al. (2020) study the impact of online reviews from Yelp on consumer demand and restaurant incentives. The authors extract the signals for the hygiene dimension from the online reviews and estimate the signal effect on consumer choices and restaurant behaviors. They find that the signal effects are above the aggregate level of rating. They also show that restaurants with a higher level of exposure to the Yelp platform are more sensitive to the hygiene regulation which Yelp is sensitive.

3.3.6 Others

Nair and Harris (2004) conducted a study on the personal digital assistants (PDAs) with a classic hardware/software structure. They develop a model to measure the level and importance of indirect network effects with firms competing with incompatible standards. The paper also compares a counterfactual case with no available third-party software with the actual issue. The insufficiency of software provision will slow the growth of the hardware market.

Boik and Corts (2016) studies the U.S. cable television industry, which reveals two-sided features by connecting advertisers and viewers close to newspaper and radio industries. The paper mainly focuses on the role of intermediaries rather than the agents on the two sides. The results show that the stations that interact with cable distributors are more likely to set a higher cable fee for distributors. The stations prefer to compete with other stations through satellite rather than local cables so that these stations can enjoy local advertisers without overlapping.

4 Model of Network Effects

In this section I illustrate the theoretical model I use to analyze the indirect network effects in the PC game market. Following the previous study I discussed in the literature review, I build a canonical hardware/software system solving the hardware adoption problem and software provision problem separately. Although there is no physical hardware product in the steam platform, the platform still possesses similar features with physical hardware products like video game consoles. The indirect network effect can then be identified through the interaction of the hardware demand and software provision.

4.1 Hardware Adoption

I establish the model mainly following Berry (1994), which provides a useful methodology to decompose the unobserved product characteristics into a linear form and then treat these characteristics as the error terms. Like a video game console, the steam platform has no stand-alone utility, I assume that the usage of the platform depends on the number of available game titles. Notice that the data includes the actual number of users. Instead of treating the platform as durable goods, I assume that people enter the market in the beginning of a given time point and leave in the end of the period. This assumption allows me to treat the actual number of users as the installed base, so that I don't have to calculate the installed base by accumulating the number of purchasers. As a free-offering Internet platform, the Steam platform does not have concept of sales, since everyone can download the product for free and uninstall the product anytime.

I use the number of broadband subscribers as the potential users of the Steam platform. This choice may not be consistent with the real case. People can use multiple computers within the Internet provided by the same broadband. Since the numbers of broadband subscribers is on the country level and I aggregate the numbers by year, the household subscribers and institutional subscribers are not identified from the large groups. Therefore, large companies are treated as same as personal users of Internet services. There are also places like Internet Café which provide Internet services for a large number of people with same broadband. However, since the Steam platform users only account for a small group to the broadband subscribers, the bias caused by the misspecification of potential market size can be ignored.

Therefore, I assume a representative user of Internet services choose to maximize the utility function at time t by choosing whether to join the Steam platform:

$$U_{it}^{HW} = x_t\tilde{\beta} - \alpha p_t + \omega N_t + \xi_t + \varepsilon_{it} \tag{1}$$

Where x_t is the matrix of observed characteristics of the platform and $\tilde{\beta}$ is the consumer-specific taste parameters. P_t is the price and then α measure the price sensitiveness of the consumers. Here I assume the α is invariant among the consumers, since the price is zero in my case. N_t is

the number of available number of titles on the platform. ξ_t measures the mean of consumers' valuation of the unobserved characteristics of the platform. And, ε_{it} is the error term.

The consumer i's taste parameter for a specific characteristic k can be decomposed into the following form:

$$\tilde{\beta}_{ik} = \beta_k + \sigma_k \varsigma_{ik} \tag{2}$$

Where β_k is the mean of the taste parameter for characteristic k and ς_{ik} is the consumers i's random preference error term, which is mean-zero and assumed to have a an independently distributed distribution.

Combining (2) and (3), the utility function can then be written as:

$$U_{it}^{HW} = x_t\beta - \alpha p_t + \omega N_t + \xi_t + \varepsilon_{it} + v_{it} \tag{3}$$

Where $v_{it} = \sum_k [x_{kt}\sigma_k\varsigma_{ik}] + \varepsilon_{it}$ is a mean-zero errors standing for the heteroskedasticity of the random taste parameters. Therefore, I can denote the mean utility level by simply fixing the error term to be zero:

$$\delta_t = x_t\beta - \alpha p_t + \omega N_t + \xi_t \tag{4}$$

After specifying the forms of representative consumer's utility function for hardware, the condition for a consumer to join the platform can be written as the comparison between the utility of platform and utility of outside goods:

$$U(x_t, \xi_t, p_t, v_i, N_t, \theta_d) > U(x_{0t}, \xi_{0t}, p_{0t}, v_{0i}, \theta_{0d}) \tag{5}$$

Using the representations of mean utility level, the range of the consumers who choose to join the platform can be denoted as the set of random taste error v_i,

$$A_j(\delta) = \{v_i | \delta_t + v_i > \delta_{0t} + v_{0i}\} \tag{6}$$

Given the distribution of consumers in the market, the market share can be represented with the density function:

$$S(\delta(x, p, \xi, N_t), x, \theta) = \int_{A_j(\delta)} f(v, x, \sigma_v) dv \tag{7}$$

Now I can denote the market share as a function of the observed and unobserved characteristics of the platform.

$$s = S(x, p, N, \xi, \theta) \tag{8}$$

However, the unobserved characteristic, ξ, could be correlated with other observed characteristics and cause endogeneity problem. The endogeneity problem can be solved through transformation of the market share function.

First, assume the unobserved characteristics are known, and the market share can be written as a function with regard to the mean utility level:

$$s = S(\delta) \tag{9}$$

If weak regularity condition holds, I can find the inverse function form for the above equation and there is only one unique $\delta^*(s)$ fulfill the relationship that $s = S(\delta^*(s))$. As the mean utility level has a linear specification form with known variables, the inverse function can then be written as:

$$\delta = S^{-1}(\mathbf{s}) = \mathbf{x}\beta - \alpha\mathbf{p} + \omega\mathbf{N} + \xi \tag{10}$$

Suppose the market share function takes a logit specification. For the logit specification, the error term $\varepsilon_i t$ follows an extreme value distribution with distribution function, $F(\varepsilon) = e^{-e^{-\varepsilon}}$. In addition, the error term is independently and identically distributed across consumers. For simplicity, I assume $\tilde{\beta} = \beta$, meaning that the random coefficients are fixed. Thus, the market share function can be written as:

$$S(\delta) = \frac{e^{\delta}}{e^{\delta_0}} \tag{11}$$

Where δ_0 is the mean utility level of outside goods. By taking logarithm form, the mean utility level can be represented as the difference between market shares of platform and outside goods:

$$ln(s_t) - ln(s_{0t}) = \delta = x_t\beta - \alpha p_t + \omega N_t + \xi_t \tag{12}$$

Thus, I establish the relationship between market share of platform and the observed characteristics of platform, with unobserved characteristics can be estimated as error terms. Then, I can solve the endogeneity problem through instrument variables regression. The mean utility level here works as the bridge to connect the function of logarithm market shares and platform characteristics. Notice that the online platform enjoys special cases with price equal to zero and it is hard to quantify observed characteristics. Thus, the difference of logarithm market shares only depends on the number of game titles and unobserved characteristics.

4.2 Software Provision

I build the model to solve software provision problem through maximizing both consumers' utility and firms' profits. The derivation mainly follows the procedure of Nair et al. (2004) with

simplifying the four-dimension software products to a three-dimension cases.

Suppose the consumer i's utility can be written as a function of all software titles in a given period t. $(q_{i1t}, q_{i2t}, ..., q_{iN_t})$ is the choice set of software products. The consumer i's utility maximization problem can be written as:

$$\begin{aligned} max_{q_{i1t},q_{i2t},...,q_{iN_t}} \quad & U_{it}^{SW}(q_{i1t}, q_{i2t}, ..., q_{iN_t}) \\ s.t. \quad & \sum_{k=1}^{N_t} p_{kt}q_{kt} + z_{it} = y_{it} - p_t \end{aligned} \tag{13}$$

The second line is the budget constraint condition, where p_{kt} is the price for title k, and q_{kt} is the demand for title k. z_{it} stands for the outside goods. y_{it} is the consumer i's income in period t and p_t is the price of complementary platform.

As the consumers prefer more titles to less, it is reasonable to assume that the consumers have a modified CES utility function of software products. Then, the utility function can be written as:

$$U_{it}^{SW} = V(Q_{it}) + z_{it} \tag{14}$$

Where $Q_{it} = (\sum_{k=1}^{N_t}(q_{kt})^{\frac{1}{\beta}})^{\beta}$ is a quantity index, and $V(Q_{it}) = Q_{it}^{\frac{1}{\alpha\beta}}$, $\alpha \geq 1$, $\beta > 1$, is an increasing and concave function of Q_{it}. Given such function specification, the optimal demand can be solved as:

$$q_{ikt}^* = (\alpha\beta p)^{\frac{\alpha\beta}{1-\alpha\beta}}(\frac{P_t}{p_{kt}})^{\frac{\beta}{\beta-1}} \tag{15}$$

Where $P_t = (\sum_{k=1}^{N_t}(p_{kt})^{\frac{1}{1-\beta}})^{1-\beta}$ is a price index for title k. Suppose all the software firms are symmetric and have identical costs, the software market retains a Bertrand equilibrium, where all titles have identical price, p. Under such assumption, the optimal demand can be written as:

$$q_{ikt}^* = (\alpha\beta p)^{\frac{\alpha\beta}{1-\alpha\beta}} N_t^{\frac{\beta(\alpha-1)}{1-\alpha\beta}} \tag{16}$$

In reality, a software firm usually has more than one titles. The development of game titles is time-costing. Thus, I assume the firm can be treated as single-product firm in a given period, which is less than one year. The profit function for this single-product firm with product k is:

$$\pi_{kt} = (p_{kt} - c)\sum_{i=1}^{Y_t} q_{ikt}^* - F_t \tag{17}$$

Where p_{kt} is the price for title k and c is the marginal cost. q_i is the optimal demand for consumer i derived from the consumer utility function above. Y_t is the installed base, which is the users in a given period in this research. F_t is the fixed cost for firms with assumption of identical production technology.

As the number of software companies is large comparing to the platform company, I assume the software companies are competing in a perfect competition market. The first-order condition implies that the firms set the price where marginal revenue equals marginal cost. In a competitive market, the marginal revenue can be represents as a function of price and price elasticity: $MR = p(1 - \frac{1}{\eta})$. The price elasticity, η, can be computed by taking partial derivative of optimal demand equation with regard to price, $\eta_{kt} = -\frac{\partial ln(x^*_{ikt})}{\partial ln(p_{kt})} = \frac{\beta}{\beta - 1}$. The marginal revenue is then $MR = p_{kt}(1 - \frac{1}{\eta_{kt}})) = \frac{p_{kt}}{\beta}$. The condition that marginal revenue equals marginal cost then imply that $p_{kt} = \beta c$. With symmetric demand function and identical marginal cost, the firms reach the Bertrand equilibrium where the price for software products are identical.

Suppose the market has no barrier for entrants, which is consistent with the reality, software firms receive zero profits. By plugging price and demand equation, the profit function of free-entry firms can be written as:

$$\pi^{free}_{kt} = (\beta c - c) Y_t (\alpha \beta p)^{\frac{\alpha\beta}{1-\alpha\beta}} N_t^{\frac{\beta(\alpha-1)}{1-\alpha\beta}} - F_t = 0 \tag{18}$$

By taking logarithm form, the equations can be rearranged as:

$$ln(N_t) = \kappa_t + \varphi ln(Y_t) + u_t \tag{19}$$

Where $\varphi = \frac{\alpha\beta - 1}{\beta(\alpha - 1)}$ and $\kappa_t = \frac{\alpha\beta - 1}{\beta(\alpha - 1)} ln(\frac{c(\beta - 1)(\alpha\beta^2 c)^{\frac{\alpha\beta}{1-\alpha\beta}}}{F_t})$. Now, the software provision decision only depends on the installed base. Notice that N_t is the number of titles available for time period t+1, and Y_t is the installed base of time period t.

4.3 Estimation Equations

I set up the structural model above beginning with the consumers' utility functions for hardware/platform and software products. The ideal and popular estimation strategy to identify the parameters should be calibrating the actual data with the simulation of the structural model. However, with limitations on the data and empirical techniques, I estimate the reduced-form equations instead. The two main equations I analyze are adjusted forms of (12) and (19):

$$ln(s_t) = \omega N_t + ln(s_{0t}) + \xi_t \tag{20}$$

Where s_t and s_{0t} is the market share of platform and outside goods. N_t is the number of game titles. The observed characteristics is omitted because lack of data. That is to say, there are only unobserved characteristics. The price is set to be zero, as the platform product is free-downloaded, which is inconsistent with the feature of digital platform.

$$ln(N_t) = \kappa_t + \varphi ln(Y_t) + u_t \tag{21}$$

Where Y_t is the installed base, which is the average users in a given period.

The market share is calculated by dividing the average users by the potential market size, which is the yearly number of global broadband subscribers. As the dependent variable in one equation is the independent variable in the other equation, there is no doubt that there is endogeneity problem. I solve the endogeneity problem by instrumental variable regressions.

4.4 Statistical Tests for IV Estimator

This section is not relevant to the previous set-up of the structural model and it's reduced-form equations. However, I find that it is necessary to illustrate how the statistical tests for IV estimator work.

Weak Instruments Test: Weak instruments means that the relationship between the instrumental variables and the endogenous variable is not sufficiently strong. In the case of weak instruments, the regular properties of IV estimator does not hold anymore. Hence, the IV estimators are not consistent.

Staiger and Stock (1994) develops asymptotic distribution theory for single instrumental variable estimator and suggests the rule of thumb for first-stage F test is $F > 10$.

Wu-Hausman Test: The null hypothesis for the Wu-Hausman test is that all regressors are exogenous. The method to test the null hypothesis is to compare the coefficients under OLS estimator and under IV estimator using the Hausman test statistics:

$$H = (\hat{\beta_{IV}} - \hat{\beta_{OLS}})' V (\hat{\beta_{IV}} - \hat{\beta_{OLS}}) \tag{22}$$

where V is the inverse of $Var(\hat{\beta_{IV}}) - Var(\hat{\beta_{OLS}})$. H follows to χ^2_k distribution asymptotically, where k is the number of endogenous variables.

Under the null hypothesis, the OLS estimator is more efficient and consistent. If the null hypothesis is rejected, the IV estimator is needed for consistent estimation.

Sargan's Test: When the number of instrumental variables equals the number of endogenous variables, it is called exact-identification. When the number of instrumental variables is larger than the number of endogenous variables, it is called over-identification. Under over-identification case, the exogeneity of the set of instrumental variables can be tested through Sargan's Test.

The null hypothesis is that all the instrumental variables are uncorrelated with the error term. Under the null hypothesis, the Lagrange-Multiplier statistic TR^2 follows the $\chi^2(r)$ distribution, where T is the sample size and r is difference between the number of instruments and the number of endogenous variables.

5 Data

5.1 Data Sources

Valve Corporation, the company that owns and operates Steam, provides an API called the Steam Web API. Many website developers use the API to construct their datasets and provide information on games to the public. The first way to collect data is to directly use the API and request game data, including concurrent players, current price and discount, and many other games-related features. There are currently 171 million apps in the steam database, not all games. So, it is roughly hard to conduct a study on the population. The study did not collect any personally identifiable information and was limited only to Valve's public data regarding games.

As stated above, many external databases use Steam Web API and construct their dataset. The two most wide-used databases are SteamDB and SteamSpy. SteamDB offers time trends of price history and concurrent players history for any specific games, which construct the foundation for the naïve time-series case study in the following sector. SteamSpy estimated informative parameters like the number of owners based on its database and provided a public API for external users. However, since Steam changed its policy regarding user-sharing information, the estimates are not reliable in the sense of academic use.

O'Neill et al. (2016) collected data based on user data and product data via Steam Web API in 2014. They formed a formal and huge dataset and made it available online. The information collected included the associated genre(s), type, developer name, current price, Metacritic rating, release date, and other such information corresponding to every product currently available on Steam.

The data used in this paper is collected through combined methods. The data on the users' side is collected from the SteamDB[5], which has offered a daily number of users since 2004. The data on fixed broadband subscribers are collected from the public database of the International Telecommunication Union (ITU)[6]. The raw data with the yearly number of fixed broadband subscribers is organized by country. I simply aggregate the yearly number of countries to the total amount for potential market size. The dataset of users is rearranged to a weekly frequency. As the daily data have a significant pattern with respect to weekends, the weekly data appear to be smoother with the time trends.

The raw data on the game titles' side is from Kaggle[7], a website that provides data for data analysis excise, where people can upload their own dataset for other users to conduct data analysis. The dataset was collected by Vincente Arce through Steam Web API and Steam Spy API in February 2022. The dataset includes 66414 unique apps, more than half of the total number of titles in Steam stores. Every game title is associated with most features that can be obtained

[5] SteamDB
[6] ITU-D ICT Statistics
[7] Steam and Steam Spy raw datasets | Kaggle

from the official pages of steam stores, like genres, prices, and release dates. After a data cleaning process, the dataset of game titles includes 53725 observations with full information on the titles. The dataset of titles is then rearranged to a weekly level as same as the users' data. Given a time point, I define the active titles as the games released within a year. In the previous study on video games, researchers find that game titles have a short life cycle.

The data on the labor supply in the U.S. computer and software industry is acquired from U.S. Bureau of Labor Statistics[8]. The data on the global weekly confirmed cases is from *Our World in Data* [9]. The consumer price index data is acquired from *Eurostat* [10].

Although the data is available from the beginning of 2004, I selected the study period to start from the first week of August in 2008 and end in the last week of January in 2022. The reason to start from a later time point is that the users' data is not continuous before August 2008. And the reason why the study period ends in January 2022 is simply because of the collection date of game titles' data.

5.2 Data Description

Table 1: Summary Statistics

Descriptions	Mean	Std. Error	Min	Max
Number of Steam users (millions)	10.24	7.19	1.10	27.52
Number of broadband subscribers (millions)	844.23	263.28	414.94	1229.19
Age of Steam platform (months)	139.21	46.84	58.00	220.00
Sales event	0.16	0.30	0.00	1.00
Sales event (Summer and Winter)	0.07	0.23	0.00	1.00
Number of active titles	3493.19	3638.25	97.00	12034.00
Average Price of active titles (deflated by Euro CPI)	8.41	1.77	5.96	12.55
Average Age of active titles (months)	5.16	0.52	3.44	6.40
Number of new released titles	75.74	78.17	0.00	305.00
Average Price of new released titles (deflated by Euro CPI)	8.43	4.30	0.00	58.91
All employees in U.S. computer and software industry (thousands)	228.66	7.78	214.40	253.80
Global Weekly Covid Cases (10 thousands)	53.58	199.21	0	2406.04

Sample size: 705

The dataset contains weekly data from the first week of August 2008 to the last week of January 2022. The total periods are 705 weeks. The data includes variables from the users' side: number of steam users, number of global fixed broadband subscribers, and age of steam platforms. The variables from the software side include the number of active titles released in one year for each week, the average price of active titles, the average age of active titles, the distribution of genres,

[8] U.S. BUREAU OF LABOR STATISTICS
[9] Coronavirus Pandemic (COVID-19) - Our World in Data
[10] Eurostat

and the number of distinct developers and publishers. Table 1 presents the summary statistics for the variables mentioned above. As I extract and compute the variables from the raw data of game titles, I can also create variables with different frequencies and statistics for different sizes of game titles. This could be useful when I try to check the robustness of estimated models.

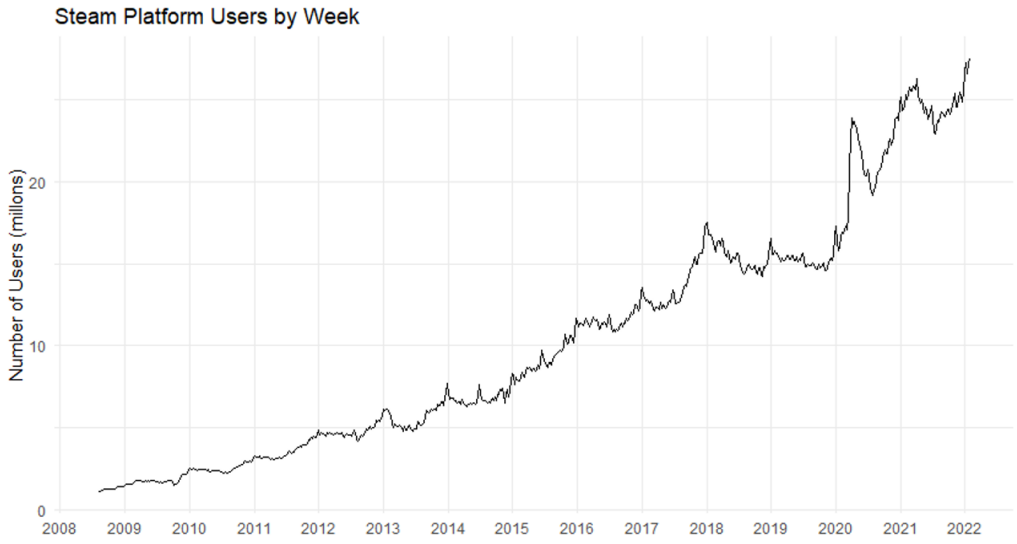


Figure 3: Steam Platform Users by Week (millions)

Figure 3 presents the growth of steam platform users from 2008 to 2022. The number of concurrent users appears to increase with the time trend. Compared to the beginning of the study period, the market size for the Steam platform has increased by 25 times. The growth of steam platform users is not monotonic. It appears to fluctuate around the time trend, implying there could be seasonal effects. It can be noticed that there is an unusual increase at the beginning of 2020 and a quick drop after the increase. The main driver could be the great pandemic of covid-19 when people face lockdowns and have to kill time by playing games.

Figure 4 illustrates the growth in the number of active titles. Remember, the active titles are defined as the game titles released within a year given a time point. The growth pattern in the number of active titles can be decomposed into three periods. From 2008 to 2014, only a few games were released every year. From 2014 to the middle of 2018, the game industry started to show its productivity, and the number of games released increased quickly to around 7500 titles a year. From the middle of 2018 to the beginning of 2022, the productivity of the game industry reached a bottleneck at first but started to continue to increase. The game variety shows such a pattern is not clear from my data. It could be associated with the development of technology and the increase in labor supply in the Internet market. The major of computer science has become one of the most popular majors in universities.

Figure 5 shows the change in the average price of active titles, deflated by the Euro Area's CPI, with 2015 as the basis. Before 2014, the average prices of active titles range from €8 to

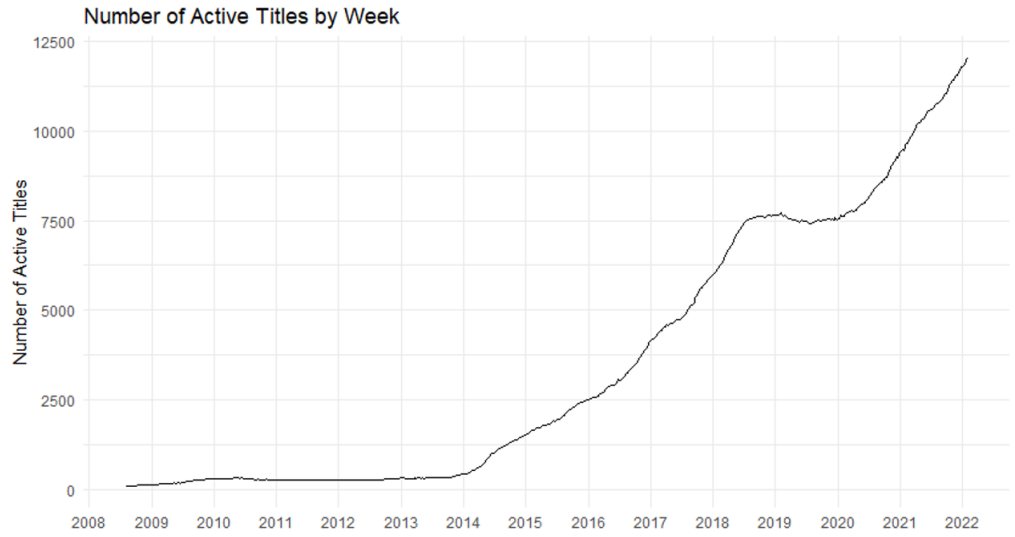


Figure 4: Number of Active Titles by Week



Figure 5: Average Price of Active Titles by Week(deflated by Euro Area CPI (2015 = 100))

€12. After 2014, the average prices of active titles decrease quickly to around €6. This pattern is consistent with the pattern of the number of active titles. With the increase in the number of game titles, the price dropped lower. The relationship suggests that the competition level has become more intense in the software market. Notice that the prices here are only the initial prices of the game titles, which are decided exogenously by the firms and do not change with time. The Steam platform holds the sales for different groups of games during a year, so the actual prices of game titles could vary. But as I only study the active game titles released within a year, a substantial price drop possibilities are slight.

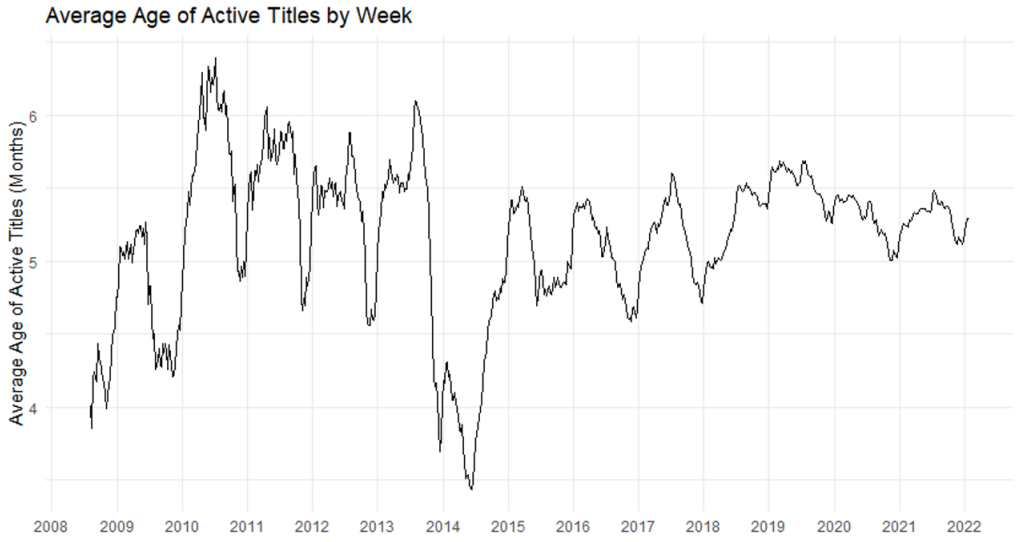


Figure 6: Average Age of Active Titles (months) by Week

Figure 6 presents the change in the average age of active titles by week. The data shows a seasonal effect in which the average age always drops at the end of a year. The pattern illustrates that the game publishers are more likely to publish a game in winter.

6 Results and Discussions

I present the estimation results of the models with platform adoption and software entry equations I proposed in the previous section. I use a two-stage least square (2SLS) estimator to estimate the coefficients for different specifications. For each specification of the models, I report an OLS regression result, the first-stage and the second-stage regressions results and an instrumental variable regressions. For each instrumental variable regression, I conduct three tests (1st stage F statistics test, Wu-Hausman test and Sargan test) and report both statistics and their significance level. The aim for the tests is to check the weak instrument assumption, endogeneity assumption and the over-identification assumption.

6.1 Platform Adoption: Basic Model

Table 2 presents the estimation results from the platform adoption equation. The dependent variable is the difference between the logarithm of platform market share and the logarithm of outside goods market share. The specification is mainly following the Clements and Ohashi (2005). The model includes yearly dummies as control variables to control for the time trends of platform's market share. I don't include other control variables, as the the models have high R^2, which suggests the errors are well explained by the model specification.

The first column reports the OLS regression of the dependent variable on the number of active game titles and control variables. Notice the number of active game titles is divided by 100 to make the coefficients more readable. Without specifications of instrument variables, the result implies that the number of active game titles has statistically significant and positive effect on the difference of market shares.

The traditional instrument variables used for estimating a product includes the product characteristics ((Berry, 1994; Berry et al., 1995; Nair and Harris, 2004)). So I first check if the average age of active games, its squared term and the age of platforms could be a valid combination of instrumental variables. The second column reports the first-stage regression result and suggest that the instrument variables are not weak and have great explanatory power. The third column reports the second-stage regression with predicted value of number of active game titles. The result is consistent with the fourth column, which estimated using instrumental variable regression.

The coefficients on the number of active game titles are relatively low. Because the potential market of global broadband subscribers is large and the changes of platform market share are small.

Table 2: Estimation Results on Platform Adoption with Instruments of Games' and Platform Characteristics

	log(Platform Market Share) - log(Outside Share)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
Number of Active Game Titles	0.006***			0.015***
	(0.001)			(0.001)
Predicted Number of Active Game Title			0.015***	
			(0.001)	
Average Age of Active Game Titles		2.773		
		(3.101)		
Average Age of Active Game Titles^2		−0.115		
		(0.310)		
Age of Platform		0.786***		
		(0.030)		
Constant	−5.786***	−56.220***	−5.797***	−5.797***
	(0.018)	(7.766)	(0.015)	(0.019)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	705	705	705	705
R^2	0.976	0.995	0.980	0.972
1st stage F stats	-	-	-	238.0***
Wu-Hausman	-	-	-	128.8***
Sargan	-	-	-	152.4***

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the platform adoption equation is the logarithm of platform market share minus the logarithm of the outside goods share.
The platform market share is defined as the fraction of the global broadband subscribers who are steam platform users given time period t.
The platform adoption equation includes yearly dummies as control variables, which are not reported in the table. There are no other control variables as the R^2 is high in all models.
The number of active game titles is divided by 100 to better present the coefficients.
The instruments used for the number of active game titles are game titles' characteristics and platform's characteristics: average age of active game titles and its squared term, and platform's age.
Standard errors are reported in the parenthesis.

Compared with the OLS regression results, the results of 2SLS estimation imply that the network effect is underestimated. However, the results may not be valid. The 1st stage F statistics is significant at 99 percent level, which suggests that I can reject the null hypothesis of weak instruments. The Wu-Hausman statistics is also significant at 99 percent level, which implies that the null hypothesis of regressors being exogenous can be rejected. Given the assumption that the instruments are valid, the difference between OLS estimator and IV estimator is significant. The Sargan statistics is also significant at 99 percent level. The null hypothesis that all the instruments are uncorrelated with the error term can be rejected. Now the model is concerned with the exogeneity of the choice of instrument variables. Thus, the results acquired from the IV estimator may not be consistent.

6.2 Software Provision: Basic Model

Table 3 presents the estimation results from the software provision equation. As same as the estimation results of platform adoption equations, the four columns report the results of the OLS regression, first-stage regression, second-stage regression, and instrumental variable regression. The dependent variable is the logarithm of the number of active game titles. As discussed in the previous section, the network effect in the software provision side is measured through the logarithm of installed base. Given a time period t, the installed base for the number of active game titles is defined as the one-year-lagged number of users. That is to say, I assumed the game developers and publishers base their software provision decisions on their previous observations, and they cannot forecast installed bases in the future.

The instrument variables used for estimating the installed bases are the age of steam platforms, its squared term and the age of active game titles. The instrumental variables also need to take lag of one year to be in line with the installed bases.

In both OLS and 2SLS estimators, the coefficients for the logarithm installed base are positive and statistically significant. While the magnitude of the instrument variable regression is larger than the magnitude of the OLS regression. If the instrumental variables are valid, it suggests that the OLS regression underestimate the network effect in the software market, as same as in the platfrom adoption case.

However, the instrumental variables is concerned with their validity. The 1st stage F statistics reported for the IV estimator is significant at the 99 percent confidence level, suggests that the combination of instrumental variables are not weak instruments. The Wu-Hausman test reject its null hypothesis of exogenous regressors, which implies that the OLS estimator is biased, given the assumption that the instrumental variables are valid. Nonetheless, the exogeneity of the instrumental variables is violated. The Sargan statistics is significant at the 99 percent level. Hence, the whole set of instrumental variables is not independent from the error term.

Table 3: Estimation Results on Software Provision with Instruments of Games' and Platform Characteristics

	log(Number of Active Game Titles)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
log(Installed Base)	0.547***			1.347***
	(0.065)			(0.117)
Predicted log(Installed Base)			1.347***	
			(0.098)	
Lagged Age of Platform		0.021***		
		(0.003)		
Lagged Age of Platform^2		−0.00004***		
		(0.00001)		
Age of Active Game Titles		−0.078***		
		(0.008)		
Constant	−2.063**	13.271***	−13.300***	−13.300***
	(0.910)	(0.133)	(1.372)	(1.651)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	651	651	651	651
R^2	0.992	0.994	0.993	0.990
1st stage F stats	-	-	-	127.78***
Wu-Hausman	-	-	-	107.69***
Sargan	-	-	-	84.97***

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the software provision equation is the logarithm of number of active game titles.
The active game titles are defined as the game titles released in a year before the start of the given time period t.
The platform software provision includes yearly dummies as control variables, which are not reported in the table. There are no other control variables as the R^2 is high in all models.
The installed base is the number of users before the start of the given time period t. That is to say, I assume the game developers base decision on their observations rather than forecasts. Similarly, the instruments used for the installed base are lagged values to be consistent with the time of installed base.
The instruments used for the installed base are also game titles' characteristics and platform's characteristics: average age of active game titles, platform's age and its squared term.
Standard errors are reported in the parenthesis.

Now I find that the traditional choice of instrumental variables for both platform adoption problem and software provision problem cannot apply to the PC game industry and steam platform. The invalid instrumental variables cannot generate consistent 2SLS estimator and thus the estimates of network effects in both sides are also biased.

A natural idea is to find valid instrumental variables to replace the invalid instrumental variables in previous estimators. In the following sections I show two different instrumental variables for both sides, one from the labor supply side and the other from the demand side.

6.3 Platform Adoption: A Valid Instrument from Labor Supply

Table 4 presents the estimation results using a combination of instruments variables. The first column is as same as previous section. In the new specification, I exclude the term of platform age. By comparing the Sargan statistics from last IV model and the Sargan statistics from the IV model without term of platform age.

A new choice of instrumental variable is from the labor supply side. From multiple choices of variables of both computersoftware industry and software publisher industry, I select the number of U.S. computer and software employees as the new instrumental variable. The variable from labor supply side has an influence on the cost side of software provision and works as an indicator for cost shifter, another traditional candidates of instrumental variables. Thus, the assumption of exogeneity holds in theory.

Replacing the term of platform age with the new instrument variable, the number of U.S. computer and software employees, the estimated coefficient of the number of active game titles of new IV estimator is much larger than the previous IV specification. The results suggest that the previous models underestimate the network effects. The coefficients of the number of active game titles is positive and significant at the 99 percent confidence level. This finding is consistent with the theoretical assumption for a positive indirect network effect.

To check the validity of the combination of new instrumental variables, I conduct the three tests again. The 1st stage F statistics is just significant at the 90 percent level. If the new combination of instrumental variables is weak instruments, the results of IV estimation could vary from the level of real network effect. The Wu-Hausman statistics is significant at the 99 percent confidence level, suggests that the choice of IV estimator is necessary. The Sargan statistics is not significant, hence I cannot reject the null hypothesis of exogenous instrumental variables. The results from the the three statistical tests prove the validity of the new combination of instrumental variables.

One may argue why not use only instrumental variable of the number of U.S. computer and software employees. On the one hand, the combination of instrumental variable has a marginally significance level in weak instruments test. The instrumental variable could result in inefficient estimation results. Therefore, the IV estimator gain consistency, but lose efficiency. Over-identification can generate more efficient estimates.

Table 4: Estimation Results on Platform Adoption with Instruments from Labor Supply Side

	log(Platform Market Share) - log(Outside Share)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
Number of Active Game Titles	0.006***			0.091***
	(0.001)			(0.026)
Predicted Number of Active Game Titles			0.091***	
			(0.006)	
Average Age of Active Game Titles		8.699**		
		(4.386)		
Average Age of Active Game Titles^2		−0.941**		
		(0.437)		
Number of U.S. Computer and Software Employees		−0.087*		
		(0.046)		
Constant	−5.786***	3.017	−5.886***	−5.886***
	(0.018)	(15.512)	(0.018)	(0.075)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	705	705	705	705
R^2	0.976	0.990	0.981	0.642
1st stage F stats	-	-	-	3.809*
Wu-Hausman	-	-	-	207.835***
Sargan	-	-	-	3.224

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the platform adoption equation is the logarithm of platform market share minus the logarithm of the outside goods share.
The platform market share is defined as the fraction of the global broadband subscribers who are steam platform users given time period t.
The platform adoption equation includes yearly dummies as control variables, which are not reported in the table.
The number of active game titles is divided by 100 to better present the coefficients.
The instruments used for the number of active game titles include: average age of active game titles, measured by month, and its squared term; number of U.S. computer and software employees, meausred by thousand.

On the other hand, the Sargan test can only apply to the case where instrumental variables are more than endogenous variables. By using sole instrumental variable, one can only discuss the exogeneity assumption of isntrumental variable theoretically.

6.4 Software Provision: A Valid Instrument from Covid Pandemic

Table 5 presents the estimation results from the software provision equation with a new instrumental variable of Covid-19 pandemic. The variable used for Covid-19 pandemic is the weekly confirmed cases after January 27, 2020. The instrumental variable from Covid pandemic seems to be an ideal instrumental variable. It is natural to consider Covid as an exogenous impact on the demand side. The Covid pandemic can also influence on the production side. But in the case of game titles supply side, the game developers and publishers are assumed to base their decision on previous observations before the pandemic impact. Thus, the Covid pandemic may not work as a valid instrumental variable when the influence on the production side become substantial. However, whether the Covid pandemic instruments is valid can also depends on the industrial conditions. It is easier for game industry to work from home than most other industries.

The whole set of instrumental variables includes global weekly covid cases (divided by 10,000), sales event and big sales event (summer and winter). All three instrumental variables are lagged by one year to incorporate the time specification for installed bases. The control variables includes deflated average price of active game titles and yearly dummies. The average price of active game titles is deflated by the consumer price index in Euro Area, as the game prices acquired are in Euro.

The coefficients for the instrumental variables in the first-stage regression are significant. And the R^2 of the first-stage regression is high. Both findings suggest that the combination of instrumental variables has enough explanatory power on the endogenous variable, logarithm of installed base. Here I don't use a logarithm of global weekly covid cases. Instead, I choose a non-linear specification of the relationship between installed base and covid pandemic.

I suppose the covid pandemic has a positive impact on the installed base. The hypothesis is inspired by the previous figure of users' evolution. The reason for such hypothesis is obvious: the covid pandemic reduce people's work time through lockdown and working-from-home, thus it would increase people's time spending on leisure. The result from first-stage regression is consistent with the hypothesis that covid pandemaic has positive impact on installed base.

The estimated coefficients for the logarithm of installed base are positive and statistically significant. The finding implies that there is positive network effect on the software provision side, which is consistent with the previous assumption. Although the coefficients from the IV estimator is slightly larger than the one from the OLS estimator, they are not statistically different. This finding suggests that the OLS estimator could also be consistent.

Table 5: Estimation Results on Software Provision with Instrumetn of Covid Cases

	log(Number of Active Game Titles)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
log(Installed Base)	0.534***			0.776***
	(0.050)			(0.188)
Predicted log(Installed Base)			0.776***	
			(0.198)	
Lagged Global Weekly Covid Cases		0.001***		
		(0.0001)		
Lagged Sales Event		0.048***		
		(0.015)		
Lagged Summer and Winter Sales Event		−0.034*		
		(0.019)		
Deflated Average Price	−0.217***	−0.002	−0.216***	−0.216***
of Active Game Titles	(0.011)	(0.008)	(0.011)	(0.011)
Constant	0.353	14.074***	−3.066	−3.066
	(0.717)	(0.085)	(2.795)	(2.655)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	651	651	651	651
R^2	0.995	0.991	0.994	0.995
1st stage F stats	-	-	-	1.55e-10***
Wu-Hausman	-	-	-	0.173
Sargan	-	-	-	0.175

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the software provision equation is the logarithm of number of active game titles.
The active game titles are defined as the game titles released in a year before the start of the given time period t.
The platform software provision includes yearly dummies and deflated average price of active game titles as control variables. The coefficients of yearly dummies are not reported in the table. The average prices for the active game titles are deflated using the consumer price index in the Euro Area with 2015 = 100. The average prices are treated as signals for the costs and qualities of active game titles. The prices are release prices, and they don't change with time.
The installed base is the number of users before the start of the given time period t. That is to say, I assume the game developers base their decisions on their observations rather than forecasts. Similarly, the instruments used for the installed base are lagged values to be consistent with the time of installed base.
The instruments used for the installed base are global weekly covid cases, sales event and two main sales event. The global weekly covid cases are divided by 10000 to better present the coefficients.

To check the validity of the instrumental variables, I report the three statistics used in previous specifications. The first-stage F statistics is significant at the 99 percent confidence level, so the null hypothesis of weak instrumental variables is rejected. The Wu-Hausman test doesn't report significant statistics, which suggest that the regressors are likely to be exogenous, given the instrumental variables are valid. In this case, the IV estimator cannot be varied from the OLS estimator statistically. The Sargan statistics is not significant, which implies that the combination of instrumental variables are exogenous. Thus, the instrumental variables from Covid Pandemic is valid, but the IV estimator could not be better than the OLS estimator.

Now, I have proposed two different instrumental variables, which are both valid and efficient. Under the new specifications of models, the network effect on the platform adoption side is more substantial, but the network effect on the software provision side is less important, compared to the traditional specifications of models.

Next, I will conduct a robustness check using weekly new released game titles instead of weekly active titles. The weekly new released game titles is just the game released in the given time period t. And corresponding installed base is the number of weekly average users. There arises a simultaneous problem. I relax the assumption that game developers and publishers can forecast the trends of users' evolution in the neighbour time periods. Thus, the number of weekly average users can be treated as a proxy for the previous and future installed bases.

6.5 Robustness Test: Demand System with Weekly New Released Game Titles

Table 6 presents the estimation results from the platform adoption equation with higher-frequency data of weekly new released game titles. The OLS estimation has a small and significant coefficient for the weekly number of new released game titles. Suppose the network effect exists, the coefficient could be biased.

The set of instrumental variables used for the platform adoption with higher-frequency endogenous variable includes the platform age and the number of U.S. computer and software employees. The difference between the specifications with weekly new released game titles and previous specification of weekly active game titles is that there are not games' age terms. The reason is because that the average software age in the new specification is identical to one week.

The results for models with weekly new released game titles are similar to the previous specification. Under OLS specification, the coefficient for the number of titles is positive and significant at the 95 percent confidence level. Supposing the instrumental variables are valid, the 2SLS estimator provides a much larger and more significant coefficient for the number of titles. The finding suggests that the OLS estimator underestimate the size of network effect on average.

Table 6: Estimation Results on Platform Adoption with Weekly New Released Titles

	log(Platform Market Share) - log(Outside Share)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
Number of Titles	0.0003**			0.006***
	(0.0002)			(0.001)
Predicted Number of Titles			0.006***	
			(0.0004)	
Number of U.S. Computer and Software Employee		−0.489*		
		(0.251)		
Age of Platform		2.091***		
		(0.219)		
Constant	−5.780***	−0.666	−5.798***	−5.798***
	(0.018)	(65.278)	(0.015)	(0.033)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	705	705	705	705
R^2	0.974	0.937	0.982	0.918
1st stage F stats	-	-	-	48.412***
Wu-Hausman	-	-	-	308.161***
Sargan	-	-	-	0.185

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the platform adoption equation is the logarithm of platform market share minus the logarithm of the outside goods share.
The platform market share is defined as the fraction of the global broadband subscribers who are steam platform users given time period t.
The platform adoption equation includes yearly dummies as control variables, which are not reported in the table.
Compared to the number of active game titles, the number of weekly new released games is much smaller. Thus, I don't divide the number of weekly new released games by 100.
The instruments used for the number of active game titles include: age of steam platforms and the number of U.S. computer and software employees(measured by thousand). The reason why I exclude the average age of new released games is that the games' ages are identical to 1 week.

To check whether the instrumental variables are valid, I report the three statistics for the IV regression. Similar to previous finding with the weekly active game titles, the first-strge F statistics and Wu-Hausman statistics are significant at the 99 percent level. The Sargan statistics is not significant, which suggests that the set of instrumental variables are exogenous.

Table 7 presents the estimation results from the software provision equation with weekly new released game titles. Since the number of new games entry to the market at the same time point as the users enter the platform, I don't have to take lags for the installed base anymore. The simultaneous problem is addressed by relaxing the assumption of game developers and publishers.

The dependent variables is the logarithm of number of weekly released game titles. To avoid the problem with the number equal to zero, the number of weekly released game titles is added by one unit and then take logarithm. This linear transformation should not be concerned with inconsistency in estimation. The control variables are yearly dummies and deflated average price of weekly released game titles.

The set of instrumental variables used for the software provision equation includes only the global weekly confirmed covid cases. The sales event and important sales event (summer and winter) are proved to be correlated with the error term by comparing the Sargan statistics with and without the two instrumental variables. While the dates of sales and big sales can be thought as exogenous, the game titles released in the sales event periods maybe affected by their competing game titles which attend the sales event. When the number of weekly new released games is small, the effects from discounted game titles could be substantial and result in endogeity.

Unfortunately, the coefficients for the interest variable, logarithm number of users, are not significant. This finding suggests a more complicated relationship between the number of users and the number of weekly released game titles. I should consider some more restricted specifications and include more control variables. The difference between the coefficients from OLS estimator and IV estimator is dramatic, with the change of coefficients' directions. Under specification of IV estimator, one percent change in number of users could cause nearly one percent change of number of weekly released game titles.

The first-stage F statistics is significant at the 99 percent confidence level, which suggests that the instrumental variable, global weekly confirmed covid cases, is not weak instrument. The Wu-Hauseman statistics is much larger than I want, implying the inefficiency of IV estimator. The Sargan test doesn't apply to sole instrument. But as I discussed in previous section, the global weekly confirmed cvoid cases should be exogeneous and valid.

The above two estimation results are consistent with the positive network effect assumption, so I suppose the the IV estimators are valid. Thus, I prove the existence of network effects in the demand system of PC gaming industry and magnitude of network effects' size could be measured.

Table 7: Estimation Results on Software Provision with Weekly New Released Titles

	log(Number of Weekly Released Game Titles + 1)			
	OLS	1st Stage	2nd Stage	IV
	(1)	(2)	(3)	(4)
log(Number of Users)	−0.162			0.949
	(0.217)			(1.426)
Predicted log(Number of Users)			0.949	
			(1.399)	
Global Weekly Covid Cases		0.0002***		
		(0.0001)		
Deflated Average Price	0.011**	−0.001	0.012**	0.012**
of Weekly Released Games	(0.005)	(0.001)	(0.005)	(0.005)
Constant	3.380	14.068***	−12.244	−12.244
	(3.058)	(0.019)	(19.687)	(20.056)
Yearly Dummies	Controlled	Controlled	Controlled	Controlled
Observations	705	705	705	705
R^2	0.909	0.990	0.909	0.905
1st stage F stats	-	-	-	4.22e-05***
Wu-Hausman	-	-	-	0.422

*Significance at the 90-percent confidence level.
**Significance at the 95-percent confidence level.
***Significance at the 99-percent confidence level.
The dependent variable for the software provision equation is the logarithm of number of weekly new released game titles.
The weekly new released game titles are defined as the game titles released in the given time period t.
The platform software provision includes yearly dummies and deflated average price of weekly new released game titles as control variables. The coefficients of yearly dummies are not reported in the table.
The average prices for the weekly new released game titles are deflated using the consumer price index in the Euro Area with 2015 = 100. The average prices are treated as signals for the costs and qualities of weekly new released game titles. The prices are release prices, and they don't change with time.
The installed base is the number of users in the given time period t. One may argue that the game developers and publishers should base their decision on the previous observations of installed base. However, I am not sure how many lags should be used for the previous observations. As the changes in neighbour periods are slight, I relax the assumption that game developers and publishers cannot forecast the future installed bases.
The instrument used for the installed base is only global weekly covid cases. The global weekly covid cases are divided by 10000 to better present the coefficients. Thus, the Sargan test statistics doesn't exist.

7 Conclusion

This study mainly aims to find the evidence for existence of indirect network effects in the PC gaming industry. As there is a major platform of game distributing, Steam platform, the PC gaming industry has a classic structure of two-sided market. The indirect network effects should exist in both sides when platform users choose to play games and software providers decide to provide game titles. For platform users, more game titles increase their value for the platform and they intend to stay in one platform because of "excess inertia" and switching costs. For software providers, more platform users stand for larger potential markets. Thus, the providers include the number of platform users into their profit function. Both network effects should be positive and the feedback loop should result in more platform users and more game titles provided. There may not be more game providers since the software market could be high competitive and reach equilibrium.

Following previous theoretical frameworks, I establish a structural model for both sides of the demand system for platform. Next, I estimate the reduced-form equations of platform adoption and software provision separately.

To overcome the endogeity problems in both sides, I choose the Instrumental Variable estimator. For the IV estimator to be consistent and efficient, the set of instrumental variables must be valid. I check the validity of the instrumental variables through three statistical tests: first-stage F test for weak instrument assumption, Wu-Hausman test for endogeity assumption and Sargan test for exogenous instruments assumption.

The results show that the traditional specification of instrumental variables in the video game industry are not valid in the PC gaming industry. Therefore, I find two more instrumental variables from the labor supply side and Covid Pandemic to account for the endogeity in platform adoption side and software provision side, respectively. Both sets of instrumental variables are proved to be valid and pass the three statistical tests. Further, I checked the robustness using higher-frequency variables.

Finally, using the IV estimator, I prove the existence of positive network effects in both platform adoption side and software provision side. On average, 100 titles increase in a year could result in nearly 0.1 percent increase in the market share of steam platform. Similarly, 1 percent increase in number of users could cause 0.7 percent increase in number of game titles in a year. The network effects in both sides are large and significant.

This research mainly contributes to the existed literature in two aspects: first, I don't find any research on the PC gaming industry and digital platform of game distributing; second, I propose two instrumental variables which I don't find in previous studies.

The study discovers the existence of network effects in the PC gaming industry and measures the magnitude of network effects' size under a static, reduced-from framework. However, the mechanism and dynamics of the network effects still remain uncovered. Further studies could

consider the dynamics and research the PC gaming industry through a formal structural parametric estimation. The further researchers could make more parametric specifications for the structural models and calibrate the model to more factors, like weekly real price and game users for each game titles. More empirical industrial models could apply for the PC gaming industry, for example, the BLP model proposed by Berry et al. (1995).

References

Ackerberg, D. A. and Gowrisankaran, G. (2006). Quantifying equilibrium network externalities in the ach banking industry. *The RAND journal of economics*, 37(3):738–761.

Argentesi, E. and Filistrucchi, L. (2007). Estimating market power in a two-sided market: The case of newspapers. *Journal of Applied Econometrics*, 22(7):1247–1266.

Belleflamme, P. and Peitz, M. (2019). Managing competition on a two-sided platform. *Journal of Economics & Management Strategy*, 28(1):5–22.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Boik, A. and Corts, K. S. (2016). The effects of platform most-favored-nation clauses on competition and entry. *The Journal of Law and Economics*, 59(1):105–134.

Caoui, E. H. (2019). Estimating the costs of standardization: Evidence from the movie industry. *Available at SSRN 3523545*.

Chandra, A. and Collard-Wexler, A. (2009). Mergers in two-sided markets: An application to the canadian newspaper industry. *Journal of Economics & Management Strategy*, 18(4):1045–1070.

Church, J. and Gandal, N. (1992). Network effects, software provision, and standardization. *The journal of industrial economics*, pages 85–103.

Clements, M. T. and Ohashi, H. (2005). Indirect network effects and the product cycle: video games in the us, 1994–2002. *The Journal of Industrial Economics*, 53(4):515–542.

Dranove, D. and Gandal, N. (2003). The dvd-vs.-divx standard war: Empirical evidence of network effects and preannouncement effects. *Journal of Economics & Management Strategy*, 12(3):363–386.

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

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review*, 103(5):1598–1628.

Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072.

Farrell, J. and Saloner, G. (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *The American economic review*, pages 940–955.

Farronato, C., Fong, J., and Fradkin, A. (2020). Dog eat dog: Measuring network effects using a digital platform merger. Technical report, National Bureau of Economic Research.

Ferrari, S., Verboven, F., and Degryse, H. (2010). Investment and usage of new technologies: evidence from a shared atm network. *American Economic Review*, 100(3):1046–79.

Gandal, N., Kende, M., and Rob, R. (2000). The dynamics of technological adoption in hardware/software systems: the case of compact disc players. *The Rand Journal of Economics*, pages 43–61.

Hagiu, A. and Hałaburda, H. (2014). Information and two-sided platform profits. *International Journal of Industrial Organization*, 34:25–35.

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

Jeziorski, P. (2014). Effects of mergers in two-sided markets: The us radio industry. *American Economic Journal: Microeconomics*, 6(4):35–73.

Kaiser, U. and Wright, J. (2006). Price structure in two-sided markets: Evidence from the magazine industry. *International Journal of Industrial Organization*, 24(1):1–28.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, 75(3):424–440.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7):2960–3000.

Liebowitz, S. J. and Margolis, S. E. (1994). Network externality: An uncommon tragedy. *Journal of economic perspectives*, 8(2):133–150.

Nair, H. (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.

Nair, H., Chintagunta, P., and Dubé, J.-P. (2004). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics*, 2(1):23–58.

Nair, V. and Harris, W. (2004). A capacity factor as an alternative to soil test phosphorus in phosphorus risk assessment. *New Zealand Journal of Agricultural Research*, 47(4):491–497.

O'Neill, M., Vaziripour, E., Wu, J., and Zappala, D. (2016). Condensing steam: Distilling the diversity of gamer behavior. In *Proceedings of the 2016 internet measurement conference*, pages 81–95.

Reshef, O. (2019). Smaller slices of a growing pie: The effects of entry in platform markets. Technical report, Working Paper.

Rochet, J.-C. and Tirole, J. (2002). Cooperation among competitors: Some economics of payment card associations. *Rand Journal of economics*, pages 549–570.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029.

Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: a progress report. *The RAND journal of economics*, 37(3):645–667.

Rysman, M. (2004). Competition between networks: A study of the market for yellow pages. *The Review of Economic Studies*, 71(2):483–512.

Rysman, M. (2009). The economics of two-sided markets. *Journal of economic perspectives*, 23(3):125–43.

Shankar, V. and Bayus, B. L. (2003). Network effects and competition: An empirical analysis of the home video game industry. *Strategic management journal*, 24(4):375–384.

Sokullu, S. (2016). A semi-parametric analysis of two-sided markets: An application to the local daily newspapers in the usa. *Journal of Applied Econometrics*, 31(5):843–864.

Song, M. (2021). Estimating platform market power in two-sided markets with an application to magazine advertising. *American Economic Journal: Microeconomics*, 13(2):35–67.

Staiger, D. O. and Stock, J. H. (1994). Instrumental variables regression with weak instruments.

Zhou, Y. (2017). Bayesian estimation of a dynamic model of two-sided markets: application to the us video game industry. *Management Science*, 63(11):3874–3894.

A Appendix

A.1 Time Trend of All Employees in U.S. Computer and Software Industry

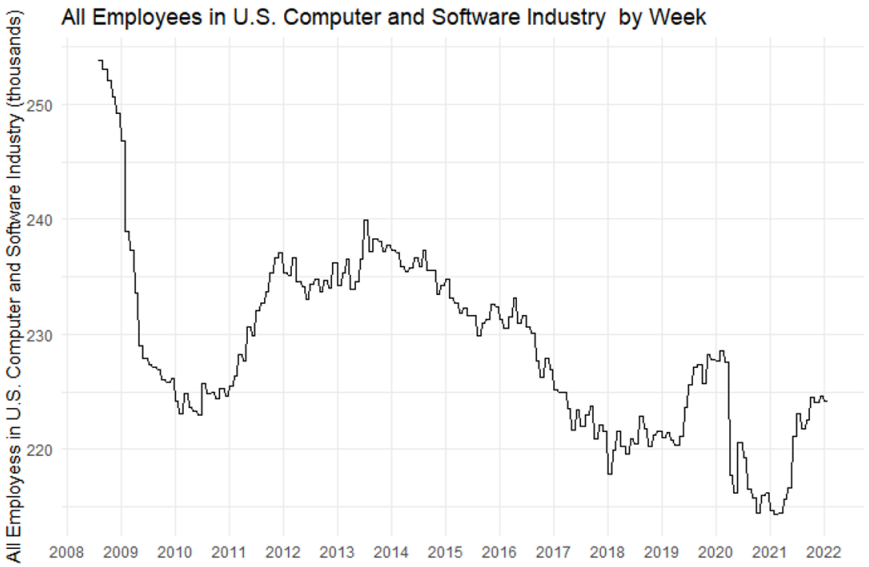


Figure 7: All Employees in U.S. Computer and Software Industry by Week

A.2 Time Trend of Global Weekly Confirmed Cases

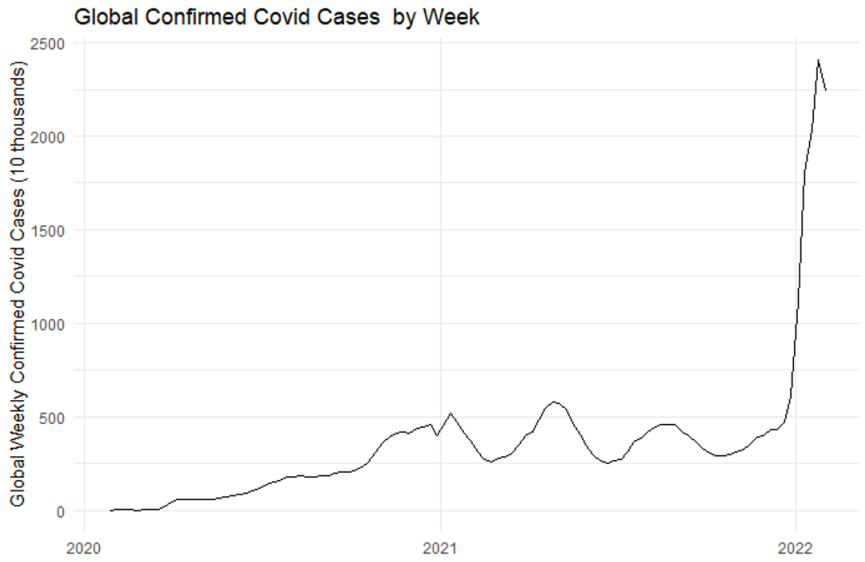


Figure 8: Global Confirmed Covid Cases by Week

A.3 An Informal Analysis of Direct Network Effects: CS:GO Players

Figure 9 shows the logarithm transformation of the weekly averaged time series of the number of Steam users (black line) and number of CS: GO players (red line). The initial data point of players has been removed because of time discontinuity. The time range of users is adjusted to merge the time series of players. The number of players experiences a rapid increase in 2012 and become more stable after 2016. Compare the two lines, I detect a coherent pattern of movement in the second half of the time range.

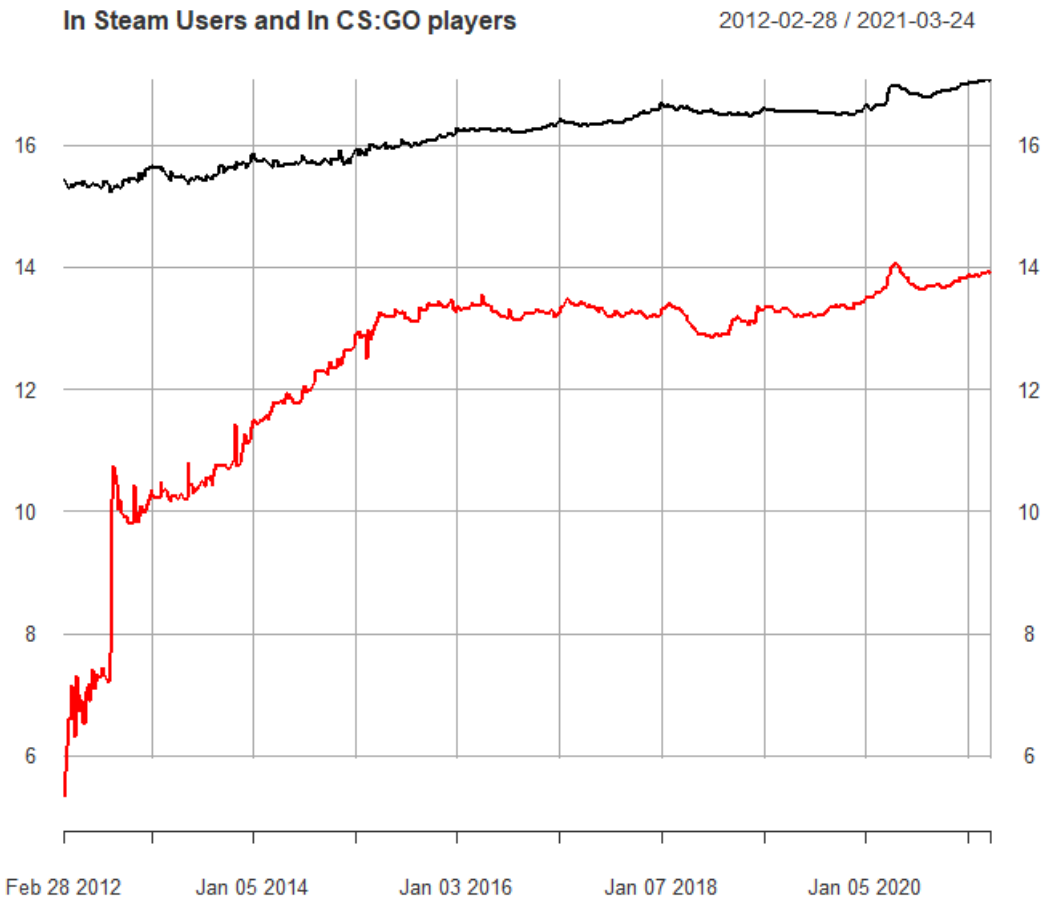


Figure 9: Log(# Steam Users - black) v.s. Log(# CS:GO Players - red)

Figure 10 presents the first difference of the logarithm transformation of the number of CS: GO players, i.e. the player growth rate. Though it is not stable at the beginning, the series of the player increase rate appear stationarity. Thus, the time series satisfies the weak stationarity conditions for an autoregression model. Before estimating the potential models, I conducted a unit root test for the player growth rate, and the existence of unit roots is rejected.

I first try an autoregression model with the order of 10, which is determined by the AIC. The dependent variable is the first difference in the number of players. Then I reduced the coefficients

of the model to make it a sparse autoregression model. The criteria used to reduce coefficients are the corresponding t-statistics. The first two columns of Table 8 show the results. Comparing the two models, the AIC decreases a little, so the second model behaves better. However, except for AR(3) coefficients, the coefficients are only significant after 7 lags.

Next, I use an ARMA(1,3) model to replace the AR(10) model. The last two columns of Table 1 present the results. In both ARMA(1,3) model and ARMA(1,3) model without MA(2) coefficient, the AR(1) coefficients are negative and significant with a significance level of 0.01. Since I expect the sign of AR(1) coefficients to be positive so they reflect the network effect, these models are failed to prove the existence of a positive network effect.

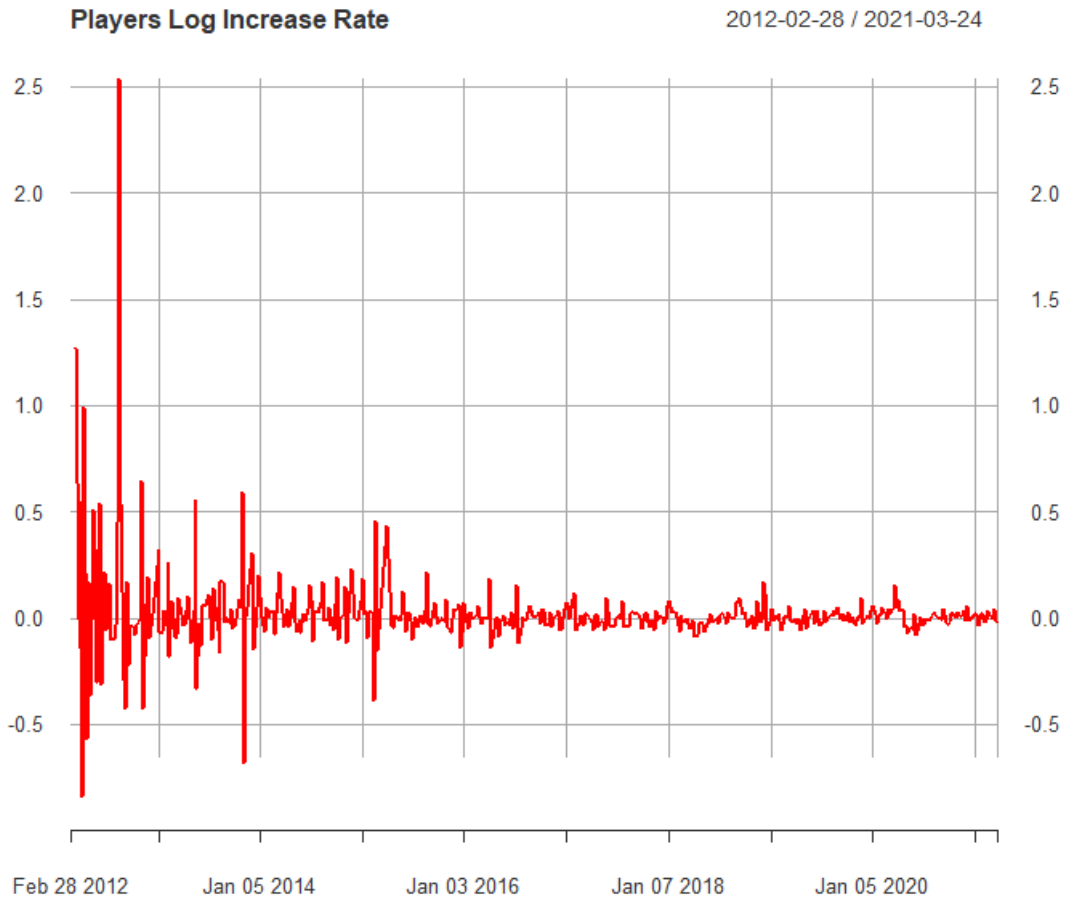


Figure 10: The First Difference of Log(# CS:GO Players)

Figure 11 depicts a fitted line of the number of players generated by an OLS model including a time trend and its squared term. The corresponding values of the coefficients are shown in the second column of Table 9. Notice the dependent variable is the number of players rather than its first difference. For this OLS model, all coefficients are statistically significant, and the model has an R^2 of 0.848. Although the OLS model must be biased since the residuals are sequentially

correlated, the number of players has a strong correlation with the time trend.

To overcome the issue of sequential-correlated residuals, I use AR(6) and ARMA(2,1) models to estimate the series of the number of players again. The orders are set by AIC and shapes of ACF and PACF. Time trend and its squared term are included in both models. The results are presented in the last two columns of Table 9. For AR(6), the AR(1) coefficient is statistically significant and positive. For ARMA(2,1), the AR(2) coefficient is statistically significant and positive. A possible explanation is that the MA(1) process makes the network effect lagged.

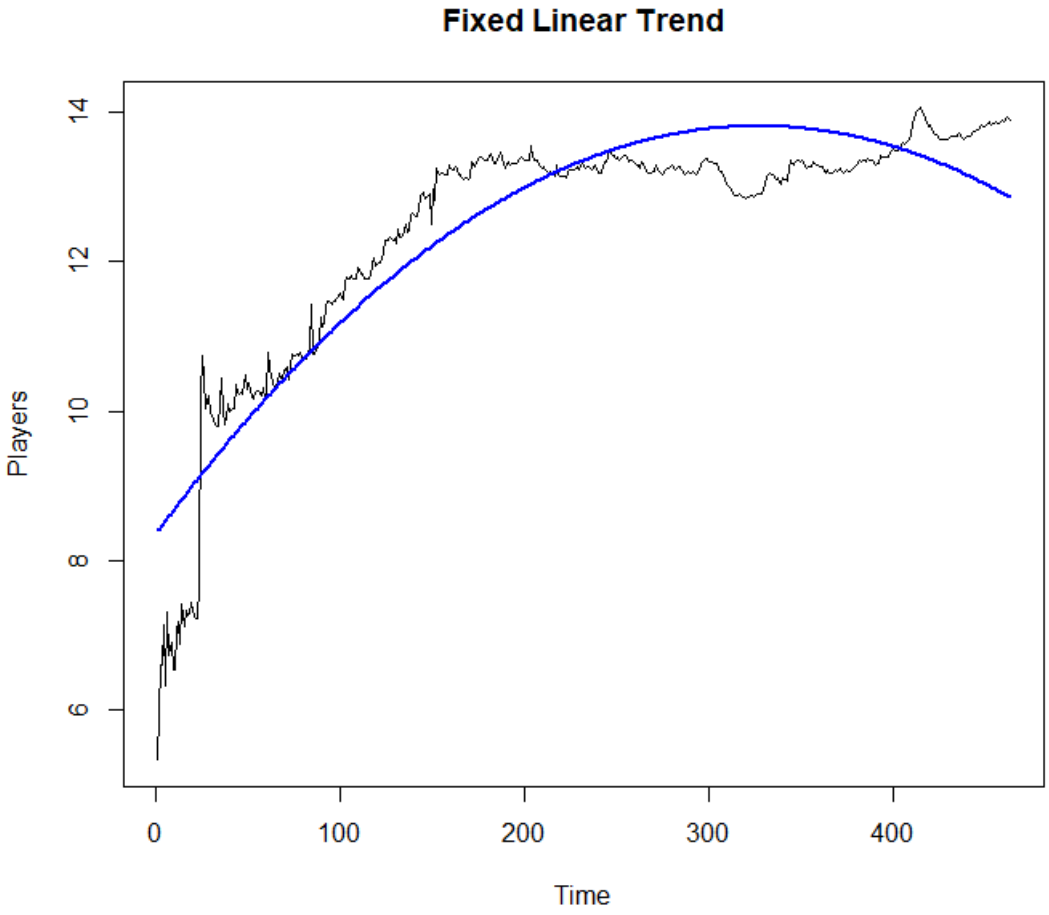


Figure 11: Log(# CS:GO Players) and OLS Fitted Line (Time Trend)

Figure 12 depicts a fitted line of the number of players generated by an OLS model including a lagged users term, a time trend, and the squared time trend. The last three columns show of table 10 an AR(5) model with lagged users, an AR(11) model with the three terms mentioned above, and an ARMA(1,3) model with the three terms. In all three models, the AR(1) coefficients are statistically significant and positive. These results may suggest the existence of a network effect for the single game. The lagged number of users is only significant when there is no time trend, which implies that the correlation of the number of players and number of users are both

correlated with the time trend rather than each other. Thus, I can not confirm the network effects of the platform on the games.

This simple time series analysis is just an informal analysis about the single good in the network effect and the single good could have a network effect alone. Here I choose the most-player game CS: GO. Similar analysis can be conducted on other titles. The coefficients could be different for different titles. Thus, it is necessary to raise a general method to estimate the overall network effect of games and the platform.

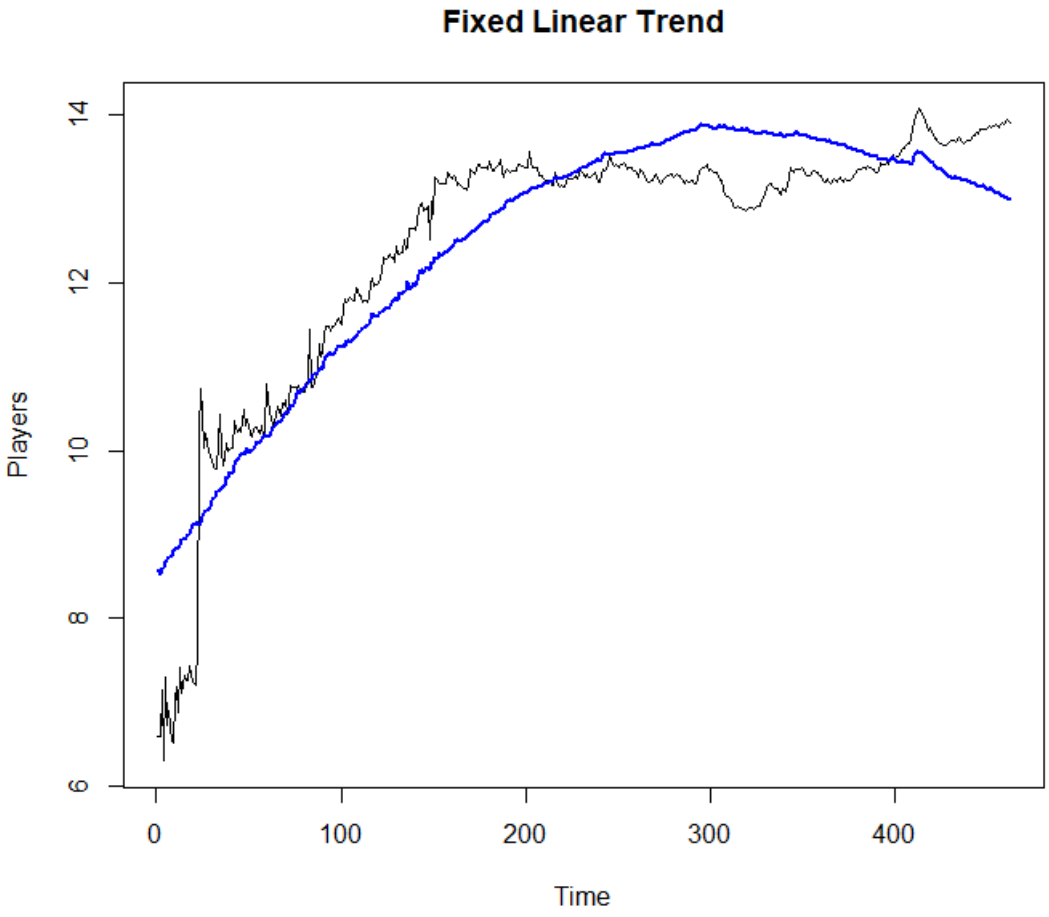


Figure 12: Log(# CS:GO Players) and OLS Fitted Line (Time Trend + lagged(Users))

Table 8: Regression Results of ARMA without time trend

	Dependent variable:			
	Δplayers			
	(1)	(2)	(3)	(4)
ar1	-0.035	0.000	-0.261*	-0.272*
	(0.048)		(0.136)	(0.141)
ar2	0.022	0.000		
	(0.047)			
ar3	-0.227***	-0.227***		
	(0.047)	(0.047)		
ar4	0.072	0.000		
	(0.049)			
ar5	-0.022	0.000		
	(0.050)			
ar6	-0.056	0.000		
	(0.051)			
ar7	-0.071	0.000		
	(0.051)			
ar8	0.110**	0.116**		
	(0.050)	(0.051)		
ar9	-0.118**	-0.110**		
	(0.051)	(0.050)		
ar10	0.171***	0.189***		
	(0.052)	(0.051)		
ma1			0.199	0.199
			(0.128)	(0.129)
ma2			0.034	0.000
			(0.047)	
ma3			-0.250***	-0.252***
			(0.054)	(0.054)
intercept	0.018**	0.018	0.018***	0.018
	(0.007)	(0.008)	(0.007)	(0.006)
Observations	462	462	462	462
Log Likelihood	143.530	139.583	132.732	132.470
σ^2	0.031	0.032	0.033	0.033
Akaike Inf. Crit.	-263.061	-267.166	-253.463	-254.940
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 9: Regression Results of OLS and ARMA with Time Trend

	Dependent variable:			
	players			
	OLS		*ARIMA*	
	(1)	(2)	(3)	(4)
ar1			0.907***	-0.007
			(0.048)	(0.020)
ar2			0.101	0.971***
			(0.064)	(0.019)
ar3			-0.255***	
			(0.064)	
ar4			0.294***	
			(0.066)	
ar5			-0.104	
			(0.071)	
ar6			0.043	
			(0.051)	
ma1				0.902***
				(0.031)
intercept			10.299***	10.452***
			(0.957)	(0.826)
t	0.010***	0.033***	0.015***	0.014***
	(0.0004)	(0.001)	(0.003)	(0.003)
t^2		-0.0001***	-7.015***	-6.811***
		(0.00000)	(1.763)	(1.720)
Constant	10.202***	8.357***		
	(0.098)	(0.091)		
Observations	463	463	463	463
R^2	0.605	0.848		
Adjusted R^2	0.604	0.847		
Log Likelihood			136.801	133.967
σ^2			0.032	0.033
Akaike Inf. Crit.			-253.602	-253.935
Residual Std. Error	1.048 (df = 461)	0.652 (df = 460)		
F Statistic	705.900*** (df = 1; 461)	1,278.851*** (df = 2; 460)		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Regression Results of ARMA with Time Trend and lag(# Users)

	Dependent variable:			
	players	lag(players)		
	(1)	(2)	(3)	(4)
ar1	0.946***	0.940***	0.943***	0.988***
	(0.048)	(0.047)	(0.047)	(0.008)
ar2	0.093	0.046	0.014	
	(0.067)	(0.065)	(0.064)	
ar3	-0.241***	-0.142**	-0.152**	
	(0.068)	(0.066)	(0.063)	
ar4	0.202***	0.142**	0.156**	
	(0.049)	(0.070)	(0.066)	
users	-0.388*			
	(0.219)			
ar5		0.011	0.043	
		(0.050)	(0.070)	
ar6			-0.106	
			(0.071)	
ar7			0.016	
			(0.071)	
ar8			0.180**	
			(0.071)	
ar9			-0.294***	
			(0.072)	
ar10			0.341***	
			(0.072)	
ar11			-0.159***	
			(0.050)	
ma1				-0.084*
				(0.047)
ma2				0.004
				(0.046)
ma3				-0.210***
				(0.054)
intercept			6.205**	8.149**
			(3.154)	(3.254)
lag(users)		0.602***	0.278	0.150
		(0.147)	(0.198)	(0.205)
t			0.012***	0.013***
			(0.003)	(0.003)
t^2			-5.983***	-6.047***
			(1.474)	(1.480)
Observations	463	462	462	462
Log Likelihood	129.737	141.317	167.118	153.810
σ^2	0.033	0.031	0.028	0.030
Akaike Inf. Crit.	-247.474	-268.634	-302.235	-289.621

Note: *p<0.1; **p<0.05; ***p<0.01