PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Luís Paulo Fernandes Bretanha Jorge

Product discovery in the PC games market

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia.

Advisor: Prof. Leonardo Bandeira Rezende

Rio de Janeiro
March 2017

Luís Paulo Fernandes Bretanha Jorge

Product discovery in the PC games market

Dissertation presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia. Approved by the undersigned Examination Committee.

Prof. Leonardo Bandeira Rezende
Advisor
Departamento de Economia – PUC-Rio

Prof. André Garcia de Oliveira Trindade
Escola de Pós-Graduação em Economia – FGV-Rio

Prof. Fábio Miessi Sanches
Departamento de Economia – PUC-Rio

Prof. Monica Herz
Coordenador Setorial do Centro de Ciências Sociais – PUC-Rio

Rio de Janeiro, March 27th, 2017

Luís Paulo Fernandes Bretanha Jorge

B.A., Economics, Faculdade de Economia e Administração da Universidade de São Paulo (FEA-USP), 2011.

Ficha Catalográfica

Jorge, Luís Paulo Fernandes Bretanha

Product discovery in the PC games market / Luís Paulo Fernandes Bretanha Jorge; orientador: Leonardo Bandeira Rezende. – 2017.

v., 65 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Informação Incompleta;. 3. Descoberta de Produto;. 4. Indústria de Jogos;. 5. Ciclo de Vida de Vendas;. 6. Bens Culturais;. 7. Modelo Estrutural de Demanda. I. Rezende, Leonardo Bandeira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 330

Acknowledgments

I am very thankful to my family for the support during these last two years, and to my colleagues for their friendship and for the times we had together. I would like to thank my advisor, Leonardo Rezende, with whom working has been a wonderful learning experience. I am also grateful to CAPES and PUC for their financial support.

Abstract

Jorge, Luís Paulo Fernandes Bretanha; Rezende, Leonardo Bandeira (Advisor). **Product discovery in the PC games market**. Rio de Janeiro, 2017. 65p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This paper investigates the role of product discovery in the demand for video games. We show that the lifetime sales patterns for video games vary widely, with some games selling most of their units in the first months after launch and others having longer tails. To understand these differences we propose a demand model in which consumers are periodically informed about the existence of a game and explore the lifetime sales patterns that this implies. We then take it to the data using price and sales figures from the Steam digital platform and web search figures from google trends. Our results imply that sales three months after launch are on average half of what they should be were consumers fully informed.

Keywords

Incomplete Information; Product Discovery; Gaming Industry; Sales Lifecycle; Cultural Goods; Structural Demand Model

Resumo

Jorge, Luís Paulo Fernandes Bretanha; Rezende, Leonardo Bandeira. **Descoberta de produtos no mercado de jogos para PC**. Rio de Janeiro, 2017. 65p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esse trabalho investiga o papel da descoberta de produto na demanda por jogos eletrônicos. Nós mostramos que o padrão de vendas para jogos eletrônicos é bastante variado, com alguns jogos tendo suas vendas imediatamente depois de seus lançamentos, e outros com caldas de venda maiores. Para entender essas diferenças nós propomos um modelo de demanda em que consumidores são periodicamente informados sobre a existência de um jogo e exploramos o padrão de vendas que isso implica. Em seguida aplicamos o modelo aos dados usando dados de preços e vendas da plataforma digital Steam e dados e busca online do google trends. Nossos resultados sugerem que vendas três meses depois do lançamento são em média metade do que seria caso consumidores tivessem informação completa.

Palavras-chave

Informação Incompleta; Descoberta de Produto; Indústria de Jogos; Ciclo de Vida de Vendas; Bens Culturais; Modelo Estrutural de Demanda

Table of contents

1

Introduction

In markets for cultural goods, such as music, movies and video games, consumers are faced with ever increasing choice sets as thousands of new products are released every year. Successful products are those that not only have an inherent quality but also are discovered by consumers, after all, for a product to be purchased, it must first be considered an option. Understanding product awareness is therefore paramount to understanding market outcomes.

Product awareness in this sense is a result of both the marketing that follows a product's release and social interactions from the consumer base, especially word-of-mouth, but directly measuring this awareness is difficult and comprehensive databases nonexistant. Though there exists marketing research that looks at how much a brand is known and how much of a reference it is to potential consumers, such as Hoyer and Brown (1990), these are of limited use when looking at individual product awareness and not brand awareness. Our strategy will be instead to infer product awareness by looking directly at the product's sales pattern.

In this paper we will focus on the PC games market. In this market, games can vary widely in terms of sales patterns: though most games sell mostly in their first few months, some games can have longer tails, maintaining consistent sales well into their second semester. Figures 1 and 2 show two examples of differing sales patterns for different games. Fallout 4, a title from a well-known franchise and well funded developer, has the greater part of its sales in the first two months, whereas Duck Game, an indie title with minimal budget, has more constant sales even outside of promotional sales (shaded areas). PC games, even more so than movies and songs, are single purchase goods, so repeat purchases are practically non-existant. Games, especially single player games, in which the number of online players is irrelevant to its enjoyment, have a set market to sell to, therefore a fall in current sales indicates either this market is being depleted or less consumers are discovering the game.

In this paper we will present a model in which product awareness is central to generating these different sales patterns, show how these sales patterns are present

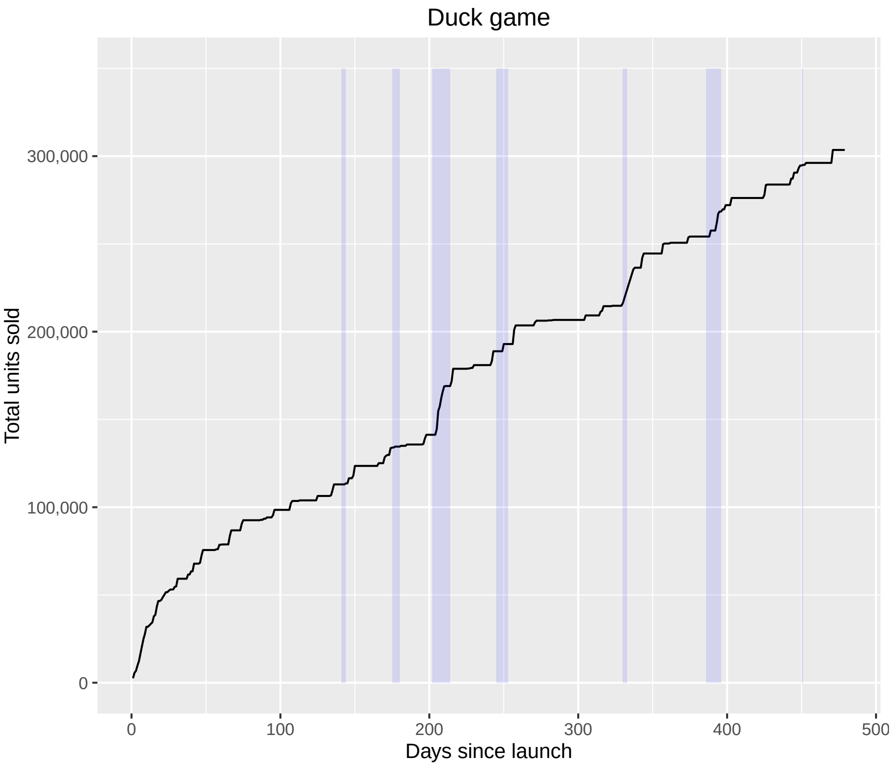

Figure 1: Estimated total sales for Duck Game over time

Notes: Shaded areas indicate periods with disconted price.

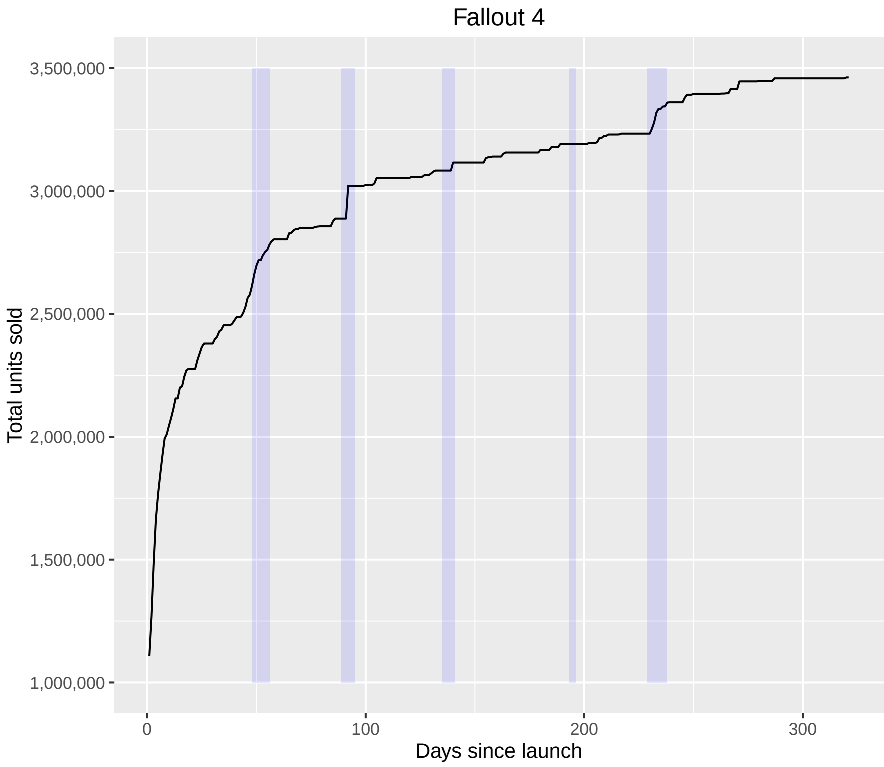

Figure 2: Estimated total sales for Fallout 4 over time

Note: Shaded areas indicate periods with disconted price.

in the data, and afterwards estimate our model. Estimation will be done using data from Valve Corporation's Steam digital games distribution platform, from which we have game price, sales and average playtime for all titles from April 2015 to September 2016. This data was not given directly by Valve but obtained conducting daily random samples of the population of Steam users using web scraping methods. We will also be using websearch data from Google trends to explore how much product awareness is affected by the search engine.

There is a vast literature on the role of information on consumer decision. Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) elaborate on the importance of herding and information cascades on determining model equilibrium. Salganik, Dodds and Watts (2006) investigate and confirm this experimentally, creating an artificial online music market and randomly assigning to users information on the choices of other users. Hendricks, Sorensen and Wiseman (2012) expand on this.

Much of this literature is focused on the effects of advertising on consumer learning. Ackerberg (2001, 2003) studies the impact of advertisement in a dynamic consumer learning model. Erdem and Keane (1996) look at the imporance of consumer learning, presenting a dynamic model in which consumers have uncertainty about product quality but learn through consumption and advertising. Moretti (2011) looks at how social learning about movie quality affects the pattern of box office sales given an initial quality expectation.

In broad terms, our paper contributes to the part of this literature concerned specifically with of the impact of lack of consumer information on their choice set. Goeree (2008) tackles this problem for the US PC industry by presenting a discrete-choice model in which the set of products from which consumers may purchase is influenced by spending on advertising, and shows that by ignoring informational assimetry in standard discrete choice models can lead to understimating product markups. Hendricks and Sorensen (2009) look at the music industry and use patterns of backward spillover, when the release of a new album augments sales of an older album, to identify sales that result from consumers learning about their choice set. In similar fashion, Kumar, Smith, and Telang (2014) use the effect of the timings of a movie's lifecycle (from theatrical release, to pay-per-view, to cable, to free-tv) on DVD sales to identify consumer learning.

The paper is organized as follows: in Section 2 we give a brief overview of the PC games market. In Section 3 we describe our data set. In Section 4 we present our product discovery model and in Section 5 we explain how we estimate the parameters in our model. In Section 6 we present and discuss our results and in section 7 we add our closing remarks.

2

PC Gaming Industry

Today the video game market stands as a significant part of the global entertainment industry, comparable in scope to the film and music industries. In the US market alone, non-hardware video game revenues have been estimated around 15 billion dollars, while box office revenue has been estimated around 11 billion and music sales around 7 billion.[1]

Another useful comparison is one of brand valuation: Activision's purchase of King Digital, known mainly for the CandyCrush franchise (US$ 5.9 billion) and Microsoft's purchase of Mojang, known mainly for the Minecraft game (US$ 2.5 billion) cost together roughly the same as Disney's purchase of Lucasfilm (US$ 4 billion) and Marvel (US$ 4 billion). Global revenues for the video game industry is estimated around US$90 billion, divided roughly equally between console, mobile and PC gaming, with PC gaming leading with US$32 billion. Of the PC market, half of the revenue comes from free to play games, one quarter from social gaming, and only one quarter from actual game sales.

Unlike in the console market, these PC games are predominantly sold through digital platforms, such as Steam, Origin, Amazon and GoG, each of these platforms carrying their own catalogue of games. Steam carries the largest catalog and deals about 75% of all PC game sales, though of note it does not carry many games of two large developers: Electronic Arts and Activision, both which sell their games through their own platforms. Steam's market dominant market position can be attributed to various factors:

- Convenient integration of sales, distribution, social networking and game access through the same platform;
- First-mover advantage: Steam was one of the first platforms on the market, establishing a large user base before others. The larger the user base, the more

[1]This does not take into account ancillary revenues for the film industry like DVD sales and digital subscriptions, and ancillary revenues for the music industry like concerts and shows, but still serves as a broad comparison.

developers are encouraged to participate, widening the catalogue and turning the platform more attractive;

- Pressure on developers to participate in seasonal sales;

- Regional pricing. Unlike the standard of setting price in US dollars and converting it to local currency at the prevalent exchange rate, prices are set in local currency and are largely invariant to the exchange rate. Because accounts are tied to the user's IP address, arbitrage is not possible.

3

Data

We have the price and sales histories of all games on the Steam catalog from April 2015 to September 2016. This data is not obtained directly from Valve nor from the developers themselves but from the website SteamSpy using a method we will describe below.

Though Valve does not make public information on video game sales made through their playform, they do allow user profiles to be accessed freely through an online API. This user profile includes which games the user has installed and how much play time he has on each one, both total and in the most recent 2 weeks. Each user has a 17-digit ID, so with some knowledge of how these IDs are generated it becomes possible to check every single profile, though this is not feasible because it would require going through millions of users on a daily basis. The alternative is to take a randomised daily poll of the population of Steam users, allowing a daily estimate of how many users have a certain game and thus have an estimate of how many total units have been sold up until that date, along with estimates on other things such as the number of hours played. From this polling we can obtain general statistics about the number of users, average game price, average units sold, etc for all paid games:

- Total active users: 150,000,000
- Recently active users (two weeks): 50,000,000
- Total games in the catalog: 9,388
- Average sales per game: 156,929
- Median sales per game: 14,575
- Average game price: $10.26
- Median game price: $8.99
- Average time spent per game: 7.05 hours

- Median time spent per game: 3.42 hours

The difference between average and median sales per game highlights the skewness in the market: there are a few games that sell well and a large number of games that sell poorly. We also note a large difference in user dedication to games, with average time spent per game at twice the median.

For every game in our database, besides the daily estimate of total units sold and average and median hours played, both recently and in total, we also have qualitative data such as the developer, the producer, game genre, etc. In Table 1 we have this for the game Mad Max, and in Figure 3 the daily statistics.

Table 1: Mad Max

Developer	Avalanche Studios
Producer	Warner Bros. Interactive Entertainment
Genre	Action, Adventure
Available languages	English, French, Italian, ...
Tags	Action (432), Open World (413), ...
Launch date	2015-09-01
User score	91%
Metacritic score	73%
Units sold	1,046,438
Price at launch	$59.99
Concurrent users	2,247

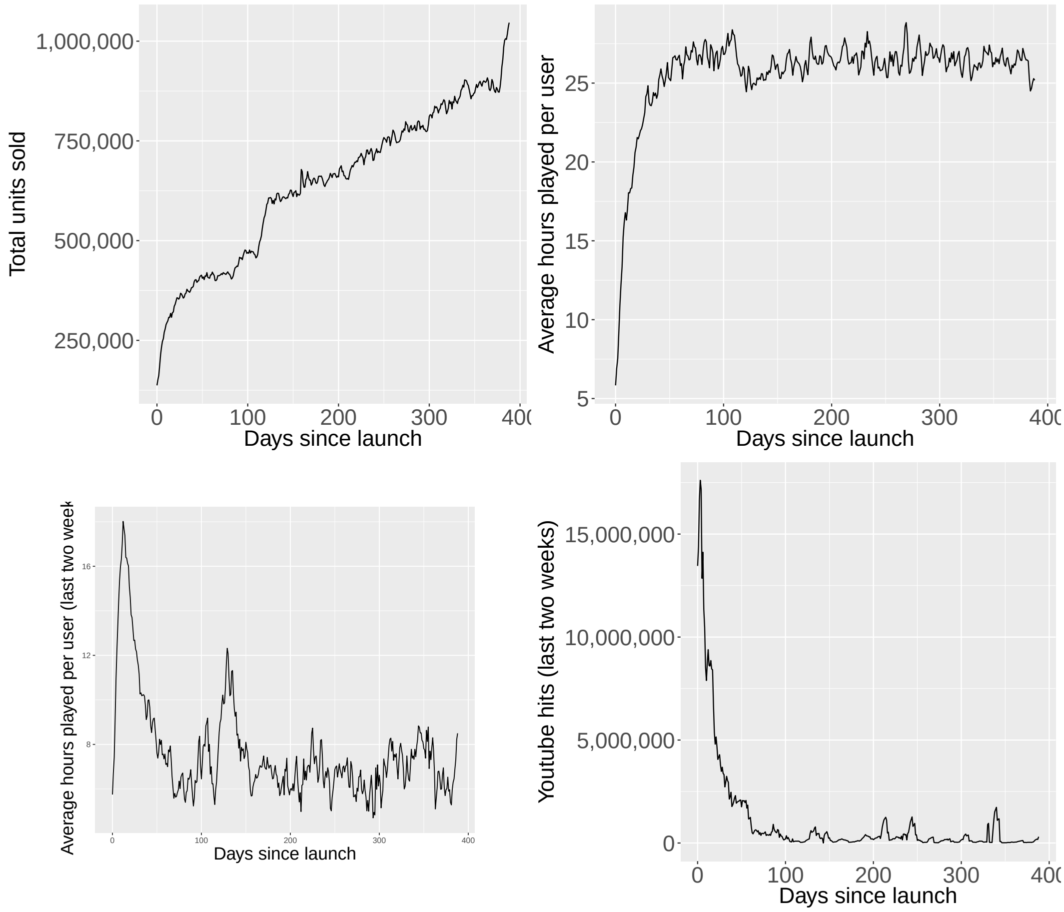

Figure 3: Available data: Mad Max

Our interest is in understanding the impact of consumer information on the game's lifetime sales patterns, so naturally we will only look at those games released after April 2015, when we begin to have data. We have 4727 games released in our database, which is about half of the total games in the Steam catalog. This means rate at which games are being launched is higher than in previous years, but a large part of these games are small projects, indicated by low sales and price tag, as we show in Figures 4 and 5. In fact, of these nearly five thousand games, only 591 of them have total sales above 50,000 units, and most have less than 10,000.

For most of these games, we note a general pattern in terms of pricing. The game is either launched at full price or at a small discount lasting the first week, followed by a two or three months at full price, after which the game remains at full price but is occasionally sold at a discount (Figure 6 shows this for the game Mad Max). More expensive games, priced between 40 and 60 dollars, tend to remain a longer time without discounts, whereas cheaper games often offer discounts a

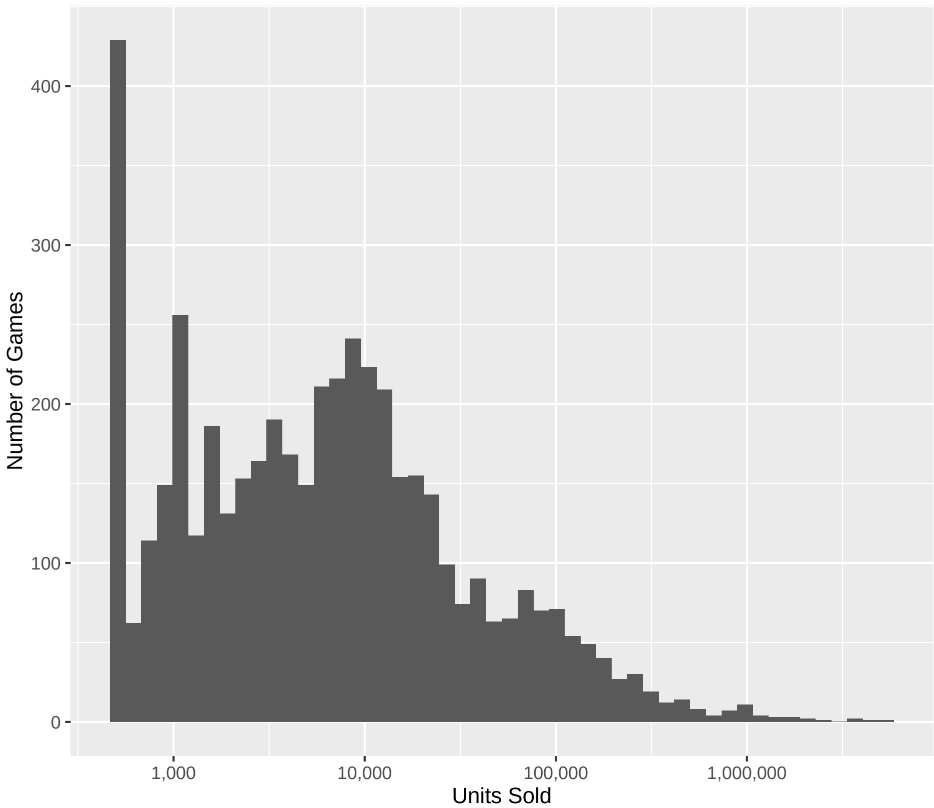

Figure 4: Distribution of total game sales

Note: Logarithmic scale. 4727 games release between April 2015 and September 2016

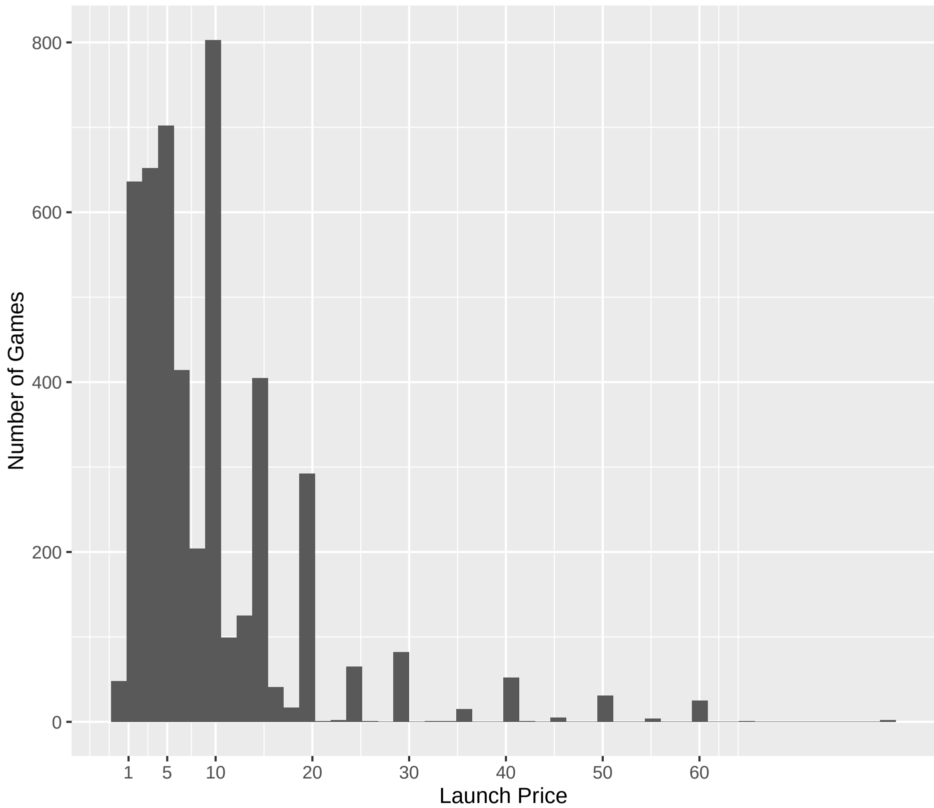

Figure 5: Distribution of game prices

Note: 4727 games release between April 2015 and September 2016 relatively short time after launch. This means prices are on average decreasing in time, though the timing of discounts vary widely game by game. By normalizing each game's price by the full price at which it was released, we show in Figure 7 how the average normalized price changes in time for all games release between April 2015 and September 2016.

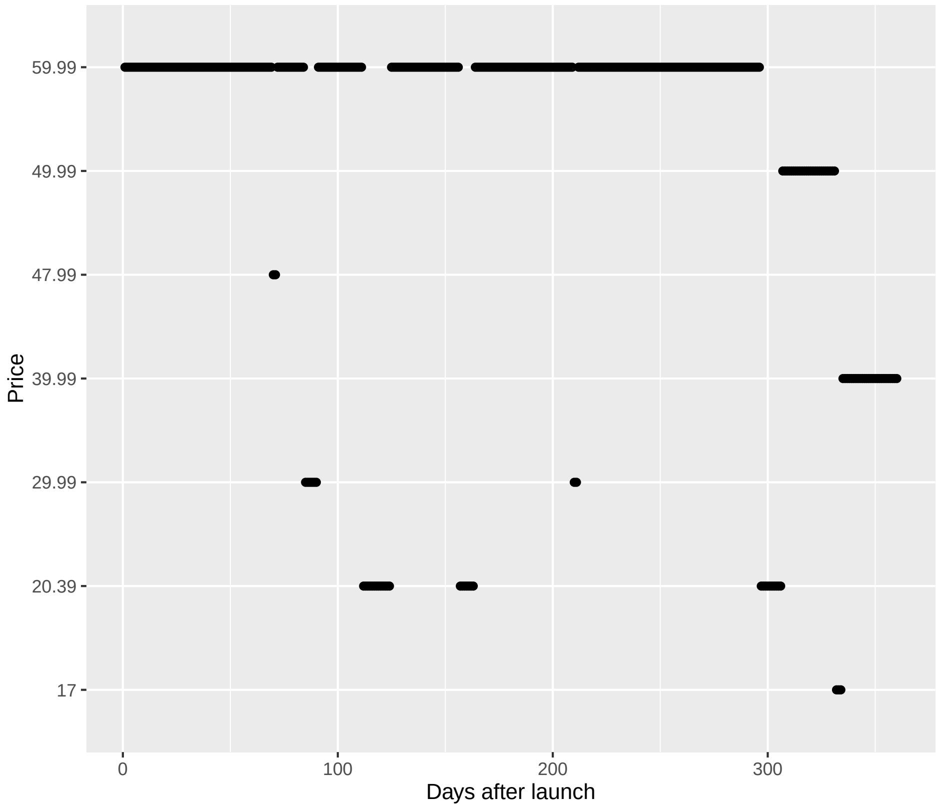

Figure 6: Mad Max prices

In Figure 8 we show the same average normalized price, but only for more successful games[2], which we will effectively use in our model. Note that for these, a much smaller part of these games offer discounts at launch and these discounts tend to be smaller as well.

As we showed for Mad Max, we have a daily estimate of average total hours played per user. Figure 9 presents a histogram of these average total hours played for these recently published games. For most of them, this average stabilizes a few months after launch, as the number of new users with no hours played become small in relation to the total number of users. Some games, usually games with important multiplayer components or management games, see average hours continue to rise

[2]In this case we are considering only paid games with more than 150,000 total sales

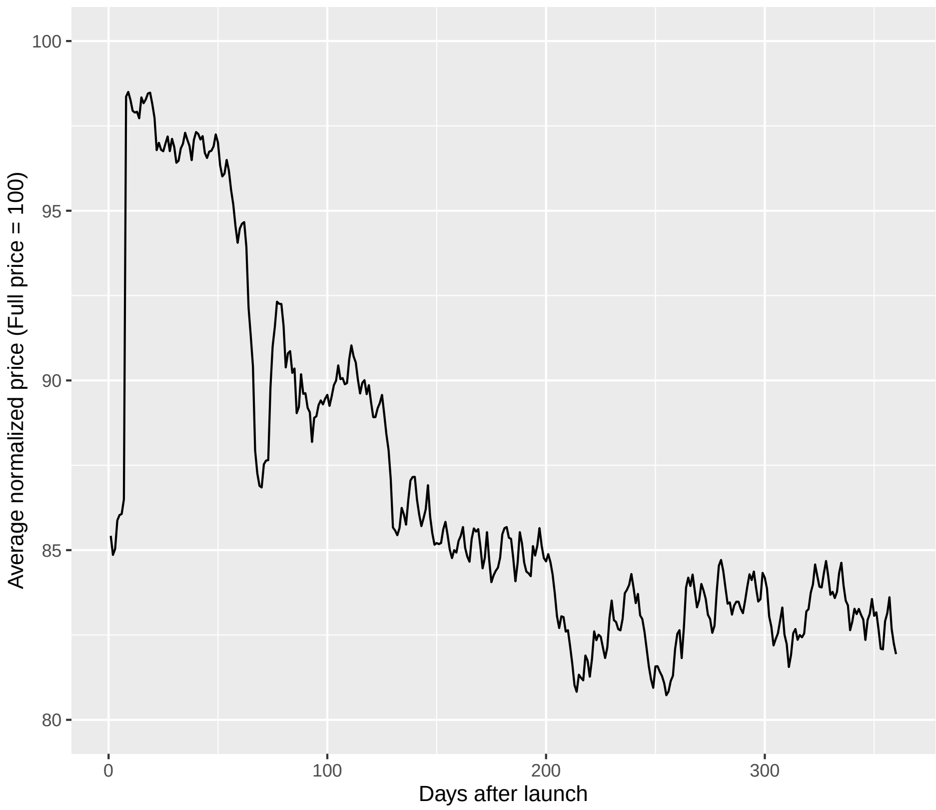

Figure 7: Average normalized price pattern

Note: All games release between April 2015 and September 2016

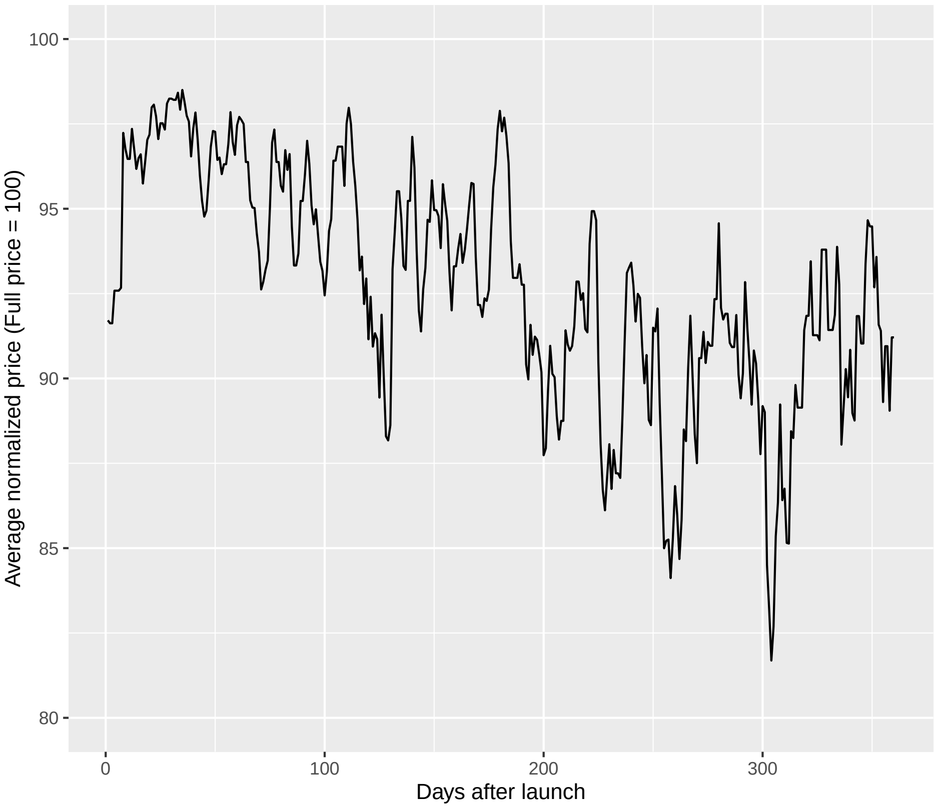

Figure 8: Average normalized price pattern

Note: Games in our effective sample

almost indefinitely. By collecting this average at the end of our sample, we have a measure of many hours of gameplay these games offer. We note this varies a lot on a game by game basis, with the average of 6 hours per game, but a median below 5 hours. There are also large outliers, again games with large replay values.

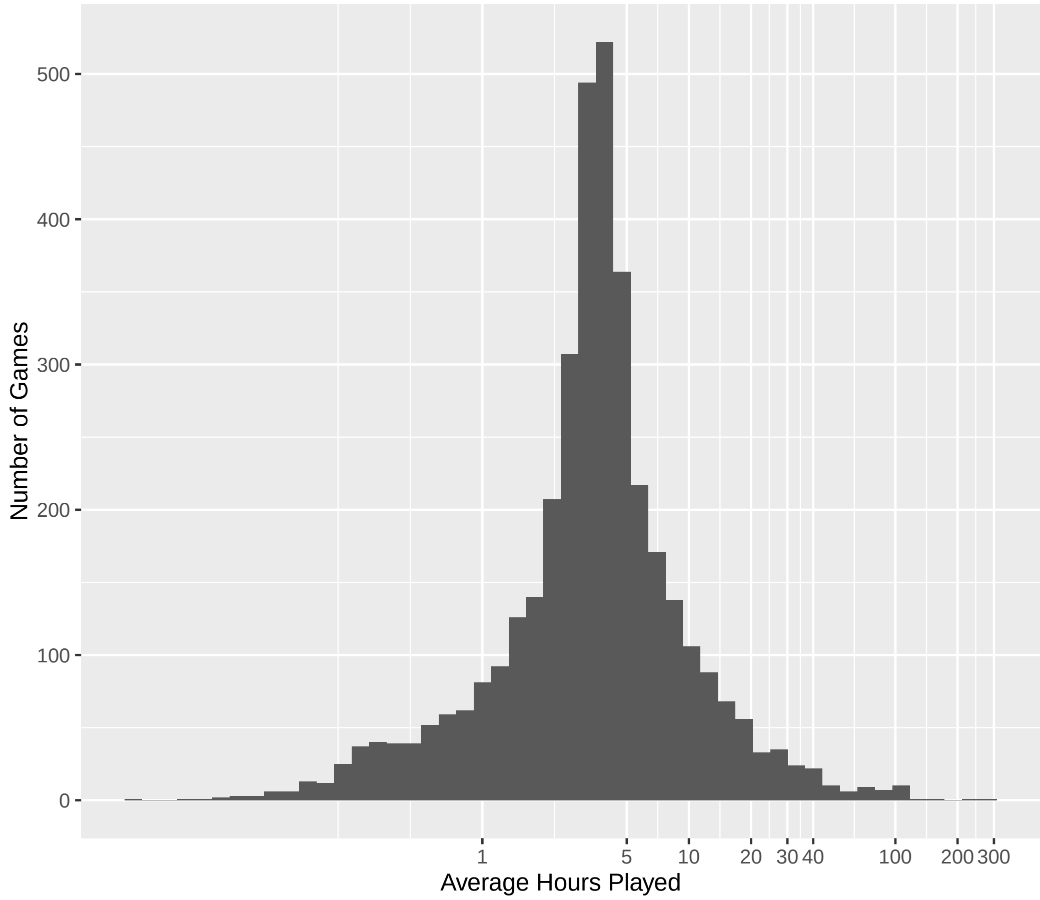

Figure 9: Distribution of average hours of gameplay

Note: Logarithmic scale

Among more successful games, we note a negative but almost insignificant relationship between price and units sold after 90 days (Figure 10). This is in part due to quality heterogeneity between games, as higher priced games are generally of higher quality, impacting sales. If we look at average hour played against sales after 90 days we notice a positive relationship between them (Figure 11), though in this case we are not accounting for different prices. If instead we consider price per hour of play (Figure 12) instead of price, we have a negative relationship as one would expect from a demand curve, though this relationship isn't perfect as games can still differ in the quality of the hours of play.

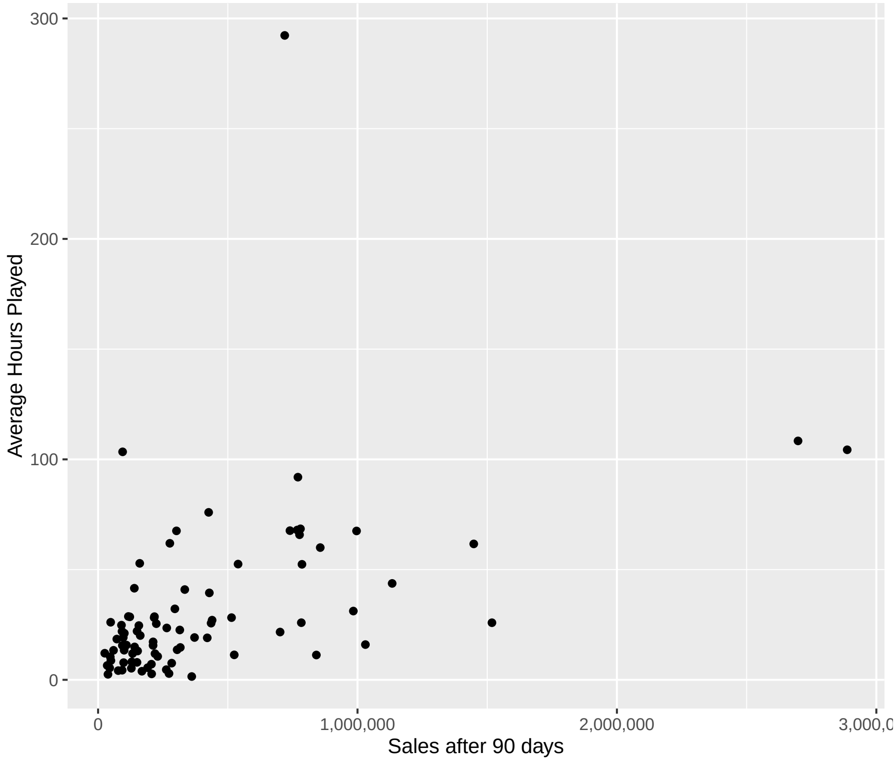

Figure 10: Average hours played x Game sales

Note: Games released after April 2015 with more than 150,000 sales

Because our sales histories are obtained through daily random sampling users, there are some specific issues to deal with. Games that have participated in large giveaways or bundled sales (such as Humble Bundle) show large increases in ownership when either no sales have taken place or they have been made at a price lower than the one registered for that day, so these games will be dropped from the sample.

Another problem is that because of random sampling, it often happens that the estimated total units sold is below what it was in the previous day, which should not be possible given that all sales are final. This problem is especially bad for games with few sales, as estimates for number of units sold vary widely, so we will also drop all games with less than 150.000 units sold from the sample.

To help identify when sales have actually taken place, we will apply an isotonic filter to the data. An isotonic filter involves finding the sequence of points (x_1, x_2, ..., x_T) that most closely follows the original sequence of points (y_1, y_2, ..., y_T) subject to being monotonic, which we know must be true for the games in our database.

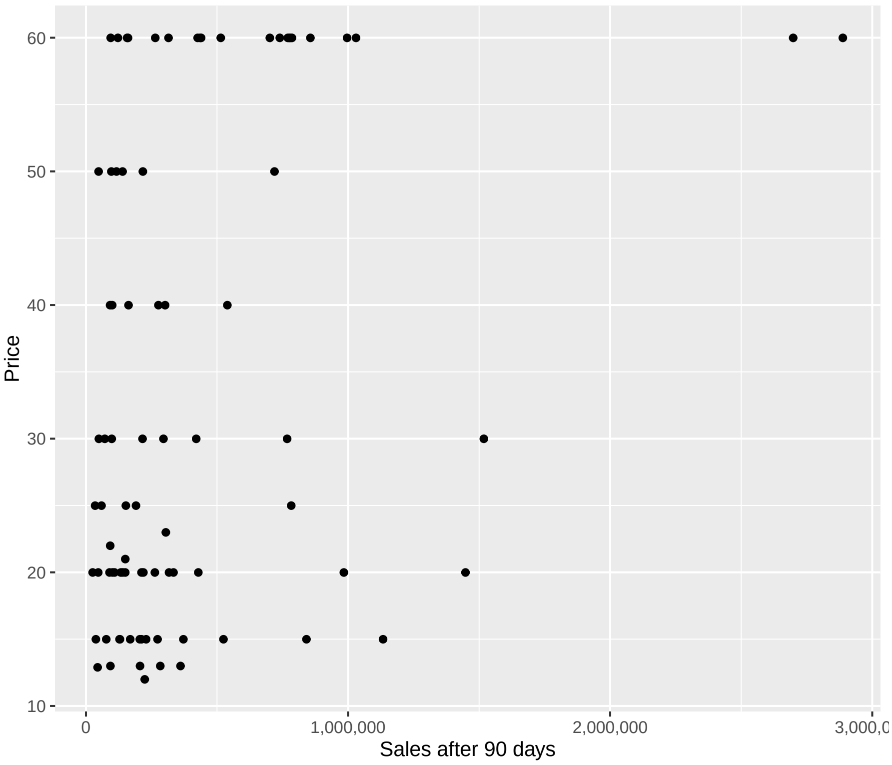

Figure 11: Game price x Game sales

Note: Games released after April 2015 with more than 150,000 sales

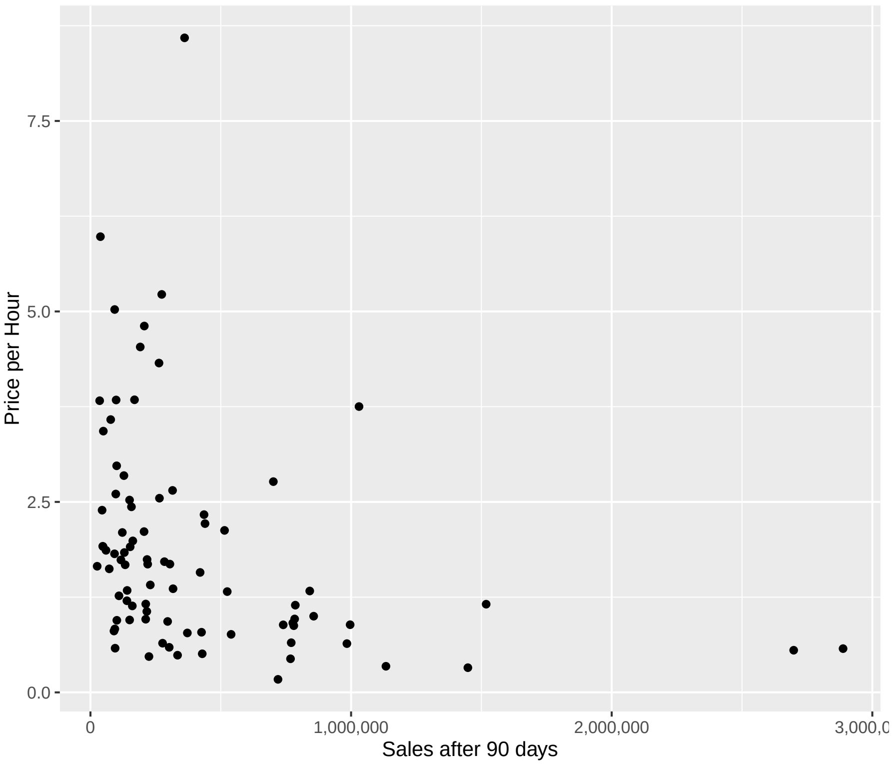

Figure 12: Game price per hour x Game sales

Note: Games released after April 2015 with more than 150,000 sales

More specifically, the sequence $(x_1, x_2, ..., x_T)$ solves the following problem:

$$\min_{(x_1,x_2,\ldots,x_T)} \quad \sum_{t=1}^{T}(y_t - x_t)^2$$
$$\text{s.t.} \qquad x_t \leq x_{t+1} \qquad \forall t$$

Applying this to the daily estimate of units sold gives us a new sales time series that is considerably smoother, as we show in Figure 13. From this new total sales time series we derive daily sales by subtrating from the estimated total units sold each day the total units sold the previous day.

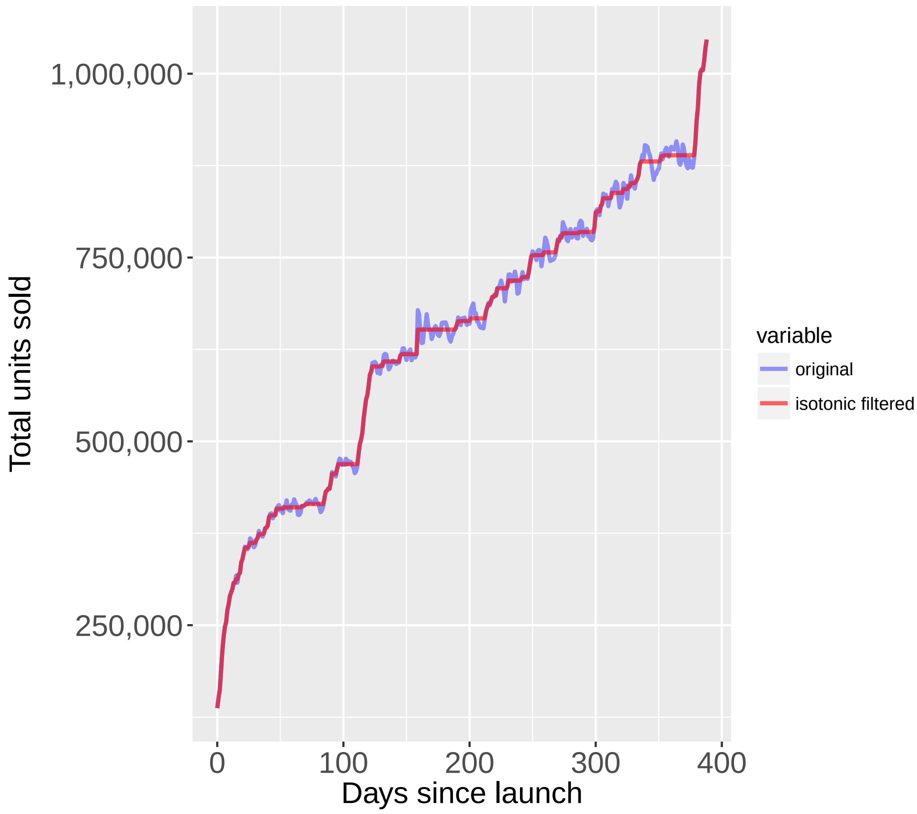

Figure 13: Isotonic filter applied to Mad Max sales

We will also be using web search data from Google, obtained directly from the Google trends website by searching for the full game name [3]. This web search data is weekly and only shows relative value (normalized to have its maximum at 100), and is created from a random sample of Google search data. Because our sales data is daily, we generated a daily google trends data set through interpolation. In Figure 14 we present these web search data for Mad Max.

[3]For those game titles clearly related to something other than the game (for example, the game Verdun), we added the suffix ”game” to our search

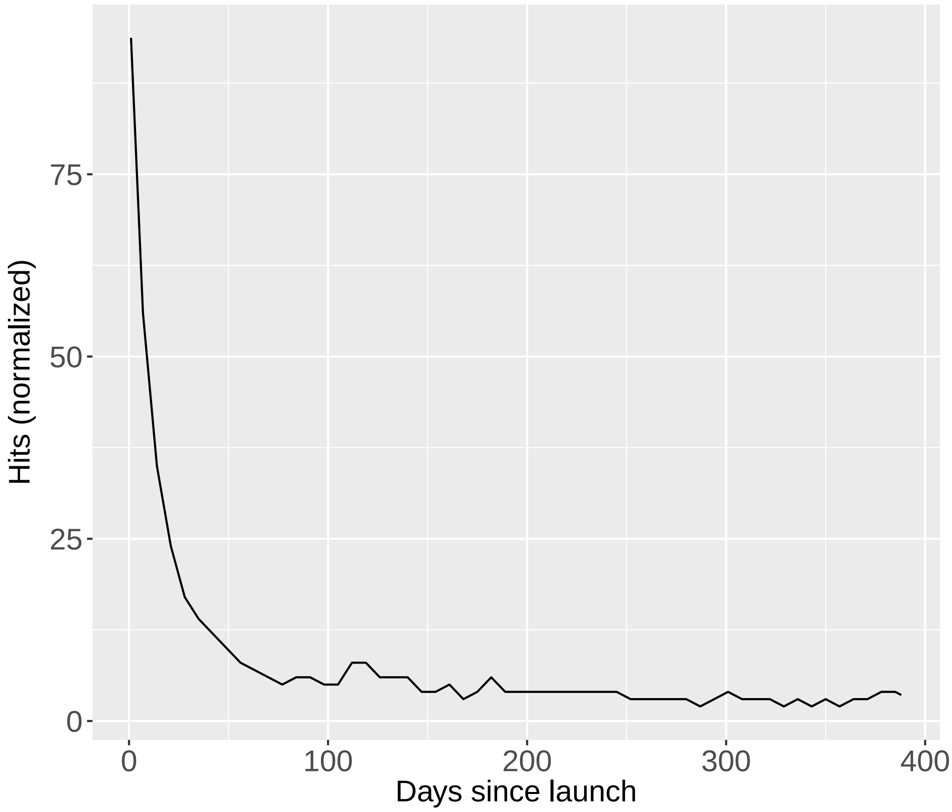

Figure 14: Daily google searches for ”Mad Max game” over time

For all but the least known games, this is the usual pattern: a large number of hits at launch, decreasing from that point forward but with occasional spikes when the game receives substantial changes or when there are price promotions.

4

Product Discovery Model

In this Section we present a model of how product discovery can impact the shape of a game's lifetime sales. The model is based on the notion that a consumer must both know the game exists and value it sufficiently high for a purchase to be made.

We will assume that preferences are fixed in the time frame, and we will not consider the effects of competition between games, treating each game in isolation. This assumption of non-substitutability is imposed for feasibility, but since video games are more akin to disposable media goods, this is not so problematic. This is also consistent with the literature specific to the video games industry and is used in other studies such as Lee (2013).

Our market consists of N_{gt} potential consumers, in other words N_{gt} is the number of consumers that would be interested in learning about game g, and is assumed to be constant over time. Each period, a fraction of the consumer base q_{gt} is randomly informed of the game, searches for information about its intrinsic quality, then proceeds to make a purchase decision, as shown in Figure 15. q_{gt} represents the speed with which product awareness grows over time, which we will call the discovery rate.

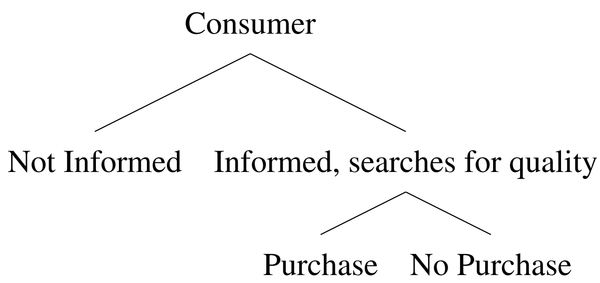

Figure 15: Product Discovery Tree

Those who make their purchase then exit the market, and those who either made no purchase or were not informed return next period to participate again in the draw.

In the next two Subsections we will discuss two different ways of determining how many consumers would make a purchase, given a certain price (what we will call the potential market at that price), if they were informed of the game.

4.1

Heterogenous consumer problem

To complete our model, we must describe how probable it is for a consumer to make a purchase at a certain price once he is informed about the game's quality. To do so, we will introduce a heterogenous consumer choice problem and arrive at solution for this probability, and consequently, at a solution for the number of consumers willing to purchase the game at each price point conditional on having been informed, what we will call the potential market at each price point. Our consumer will solve this static problem every period they are informed until they leave the market.

First, let us describe the consumer utility of consumer i as a function of the number of hours h_g the game provides of entertainment, its quality per hour of play and its price:

$$u_{gi} = \beta_{gi} * h_g - p_g^k$$

Because consumer tastes are heterogenous, we divide β_{gi} into its average and an error term:

$$\beta_{gi} = \beta_g + \xi_{gi},\ \xi_{gi} \sim N(0, \sigma_g)$$

The probability of any given consumer to purchase a game at price p_g^k is

$$P(purchase|p_g^k) = P(u_{gi} > 0) = 1 - \Phi(\frac{p_g^k}{h\sigma g} - \frac{\beta g}{\sigma g})$$

This way the initial potential market at each price point p_g^k is given by:

$$M_{g0} = N_g * (1 - \Phi(\frac{p_g^k}{h\sigma g} - \frac{\beta g}{\sigma g}))$$

By assuming the error term follows a Normal distribution we are assuming the game's demand function has a convex shape. We believe this assumption fails for only a very small subset of games; we will come back to this when discussing results.

Nair (2007) argues for the importance of the dynamic aspects of consumer decision making, such as postponing a purchase in order to wait for a sale. In this model we ignore these dynamic effects, which means that our estimates of consumer demand are conditional on expectations of future price changes, which we don't observe .

4.2

Flexible potential markets

In the previous section, we made assumptions of consumer behaviour in order to determine the potential market at each price point. Alternatively, we can allow the potential markets at each price point to be flexible.

For every game we define all the price points observed in the data from highest to lowest $(p_g^0, p_g^1 ... p_g^k)$. We then define m_{gt}^0 as the pool of consumers who value game g above p_g^0, m_{gt}^1 as the consumers who value the game between p_g^0 and p_g^1 (including p_g^1), and so on until m_{gt}^k. The potential market at price p_g^k is defined as the sum of these values up to k:

$$M_{gt} = \sum_0^k m_{gt}^k$$

5

Estimation

In this Section we will explain how we take the model to the data and discuss identification.

5.1

Maximum Likelihood Estimator

Because consumers are informed randomly with probability q_{gt}, sales V_{gt} follow a Binomial distribution:

$$V_{gt} \sim Binom(q_{gt}N_{gt}, \frac{M_{gt}}{N_{gt}})$$

In which M_{gt} is the potential market at the current price. We approximate this to a normal function:

$$V_{gt} = \frac{1}{\sigma_{gt}\sqrt{2\pi}} e^{-\frac{(V_{gt}-q_{gt}M_{gt})^2}{2\sigma_{gt}^2}}$$

Thus we write the likelihood function for our whole sample:

$$L(V) = \prod_{t=0}^{t=T} \frac{1}{\sigma_{gt}\sqrt{2\pi}} e^{-\frac{(V_{gt}-q_{gt}M_{gt})^2}{2\sigma_{gt}^2}}$$
$$\sigma_{gt} = q_{gt} * M_{gt} * (1 - \frac{M_{gt}}{N_{gt}})$$

M_{gt} is given by the initial potential market at each price level m_{g0}^0, m_{g0}^1, ..., m_{g0}^k, which are parameters to be estimated, and the sequence of sales up to that point. Period by period we subtract the actual sales from the initial potential market, as a proportion of how many potential markets are included in the period's price. [4],

It is possible to estimate N_{g0}, but its identification would be given by the variance in sales, as it does not influence expected sales. Because of how our data set was composed, we believe this identification is tenuous at best, so we will instead calibrate our model using a total consumer base of 50 million for every game, which is roughly the number of unique active Steam users every two weeks.

[4] For example, if the game is currently sold at \$20, we observe 30 sales and we estimate a potential market of 200 units between \$20 and \$30 and 100 units at \$30 or more, our algorithm subtracts 20 units from the first potential market and 10 from the second

We cannot estimate and identify q_{gt} for every period, but we can estimate it as a function of the google trends data. Our assumption is that after discovering the game, our consumer will search for information to determine its quality, so our probability of discovery is proportional to web searches. Because google isn't the only means of obtaining information on quality, we also include a fixed probability of discovery. Note that because our samples are of relatively short periods (from 6 to 18 months), when sales are nearly constant we lose identification. This generally happens for very small games, which is why we will only consider games with more than 150,000 units sold.

$$q_{gt} = \alpha_{g0} + \alpha_{g1} * GT_{gt}$$

In this equation, α_{g0} and α_{g1} are the parameters to be estimated.

For the parametric version of our model, estimation follows exactly as before, except instead of estimating each price point separately, we are estimating the two parameters β_g and σ_g.

There are a couple of issues that we have to deal with in order to be able to estimate our model. The first is the issue of pre-sales. Often, but not always, games are sold before their launch date, and this is done sometimes at a small discount. When a game has pre-sales, these show up on our database as day 1 sales, despite the sales having actually happened beforehand. Figure 16 shows this for the game Fallout 4.

Because the timing of sales is crucial to identification, this is likely to bias our results, so we will ignore pre-sales by estimating our model from the second day onwards.

We run four specifications for our model. In specifications 1 and 2, we do not use google trends data, that is, we assume the discovery rate q_{gt} is constant throught time, while in specifications 3 and 4 we allow the discovery rate q_{gt} to vary throughout each game's lifetime as a function of web search data from google trends GT_{gt}:

$$q_{gt} = \alpha_{g0} + \alpha_{g1} * GT_{gt}$$

In specifications 1 and 3 we estimate market sizes non-parametrically, that is, estimate the potential market size at each price point individually. In specifications

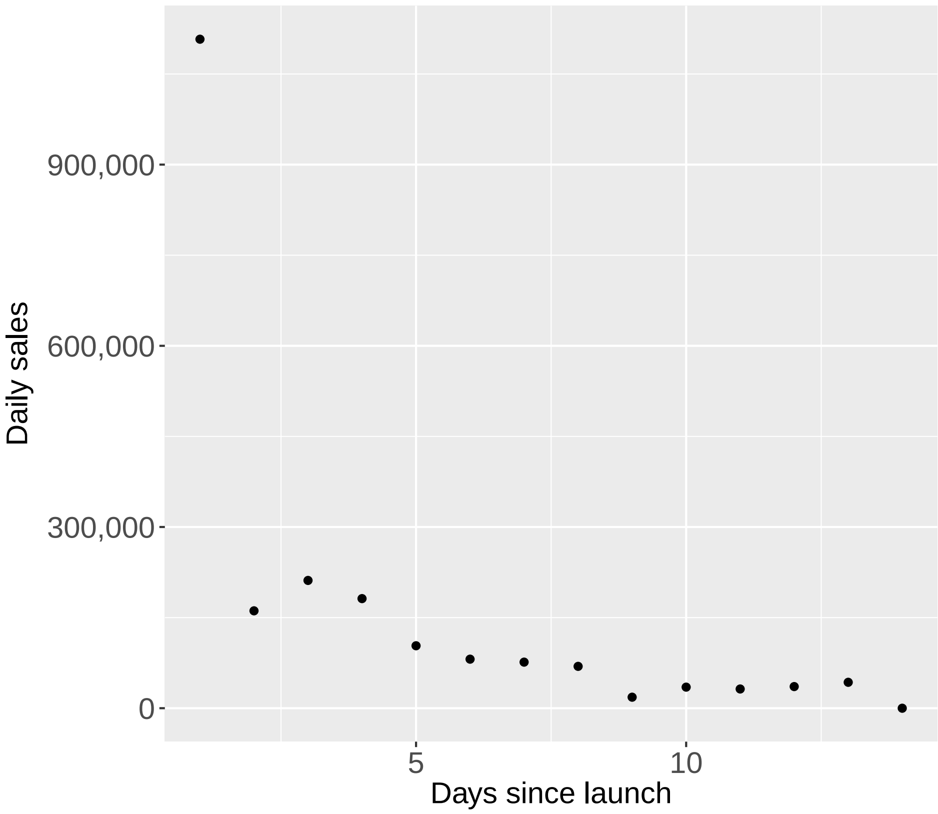

Figure 16: Estimated total sales for Fallout 4 over time

Note: Pre-sales show up in our database as day 1 sales.

2 and 4 we estimate the potential market at each price point according to the consumer problem presented in Subsection . This will yield two estimated parameters: β_g, which is the average (among all consumers) utility per hour of play, and σ_g, a measure of the heterogeneity in the idiossincratic utilities. These specifications are summed up in Table 2.

Table 2: Specifications

	Fixed q_{gt}	Variable q_{gt}	Non-Parametric Estimates	Parametric Estimates
1	Yes		Yes	
2	Yes			Yes
3		Yes	Yes	
4		Yes		Yes

5.2

Identification

Identification for the variables tied to the discovery rate q_{gt} (α_{g0} and α_{g1}) in all specifications comes primarily from the shape of the sales data. Games with strong early sales present large q_{gt} while games with weak early sales present a low q_{gt}. To showcase this, we simulate three sales patterns using for a game with a single price and differing but constant q_{gt}. For each simulation we have a total population of 1 million, of which 100,000 would buy the game conditional on having discovered it. Figure 17 presents these simulated sales patterns.

In the top simulation (q_g = 0.02), sales are strong in the first few months but drops off quickly and the daily sales graph takes a convex appearance. In the middle simulation (q_g = 0.002), sales also decrease, but more slowly, and the daily sales graph takes a more linear decreasing appearance. In the bottom simulation (q_g = 0.0002), sales remain constant over the time frame and the daily sales graph appears to have no trend. Note that in the first simulation by the end of the year nearly everyone knows about the game and nearly all possible units have been sold. In

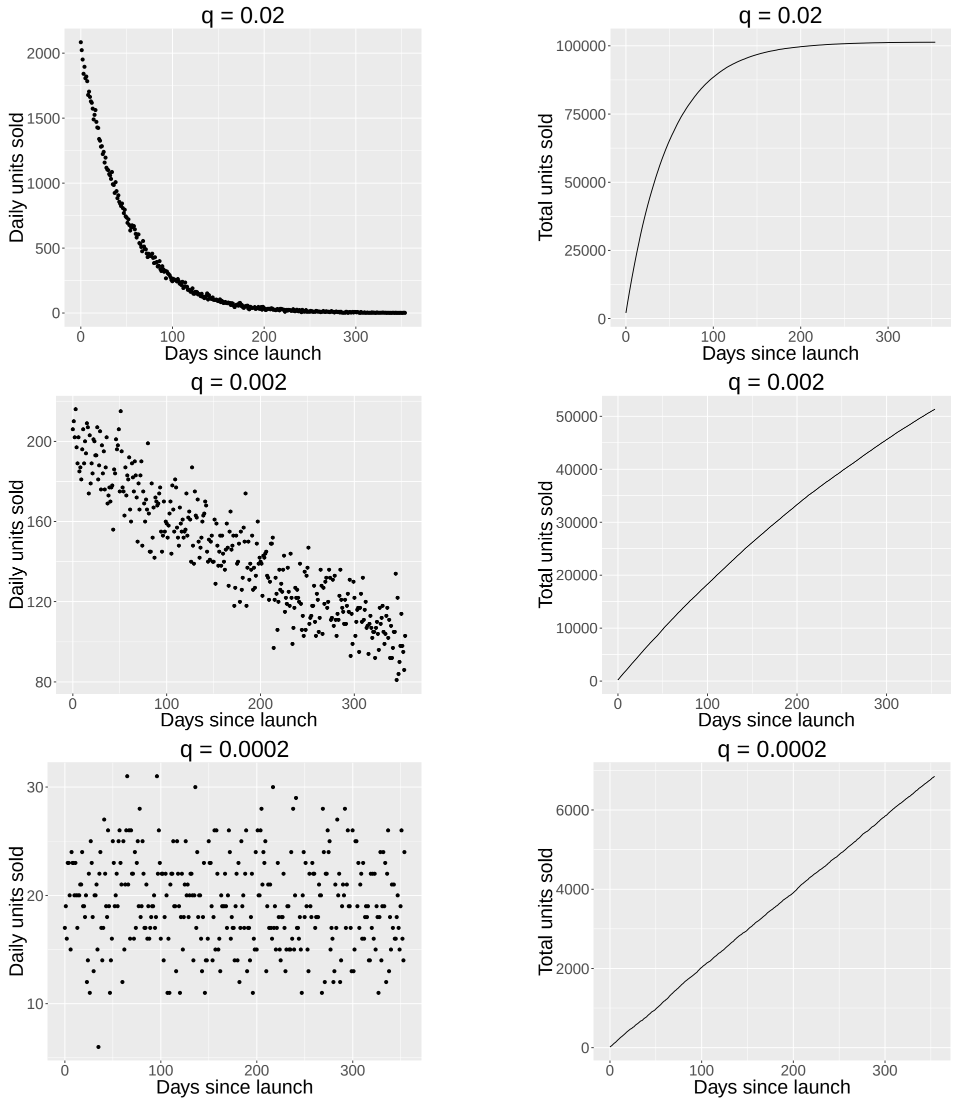

Figure 17: Model simulation for different q's

Note: N_{g0}=1,000,000, M_{g0}=100,000 the second simulation, only 50% of consumers know of the game and in the third simulation only 6%.

In Figure 18 and 19 we have the results for specification 1 for the games presented in the introduction.

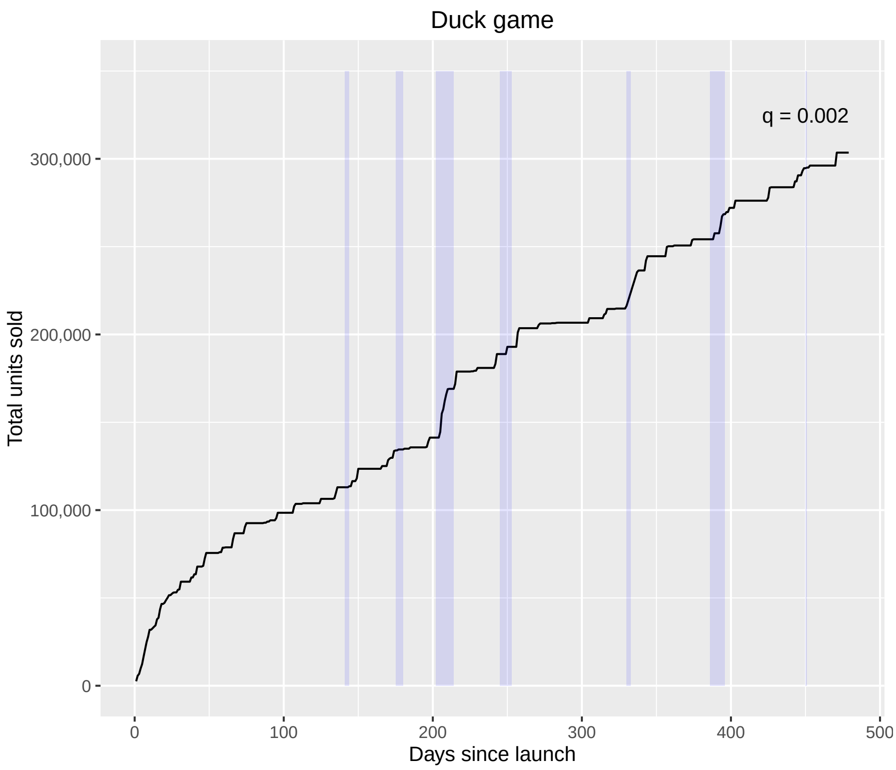

Figure 18: Estimated total sales for Duck Game over time

Note: Shaded areas indicate periods with discounted price

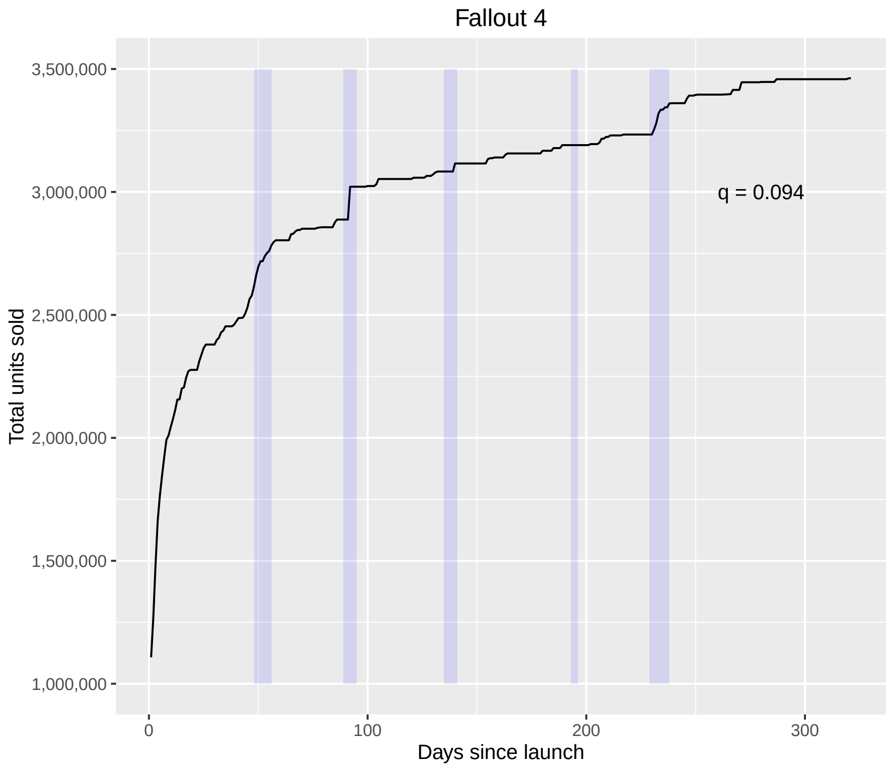

Figure 19: Estimated total sales for Fallout 4 over time

Note: Shaded areas indicate periods with discounted price

Identification for potential market sizes in specifications 1 and 3 comes primarily from the level of sales, as games with large sales naturally present larger estimates for potential markets than games with smaller sales with similar discovery rates.

Identification for β_g and σ_g comes from the effect that lower prices have on sales. Large increases in sales imply a low σ_g, and small increases imply a bigger σ_g. Intuitively, this is because in the former case there is less disagreement about the game's quality, so a drop in price is all it takes to bring in a large amount of new consumers. In the latter case, there is more disagreement and game is more niche, so a change in price affects less sales. We note that for games sold only at a single price, there are a multitude of combinations of β_g and σ_g that define the same potential market size at that price, therefore β_g and σ_g are not separately identified, though the potential market they generate is. In table 3 we show two games, both originally priced at \$60, with similar average playtimes and similar days spent on sale:

Table 3: Identification for β_g and σ_g

	Call of Duty Black Ops 3	Dark Souls 3
Release Date	2015-11-5	2016-04-11
Sales at $60	806,652	1,040,951
Sales at $45	131,247	46,861
Sales at $45 per day	7,720	4,260
$\hat{\beta}$	-1.68	-9.54
$\hat{\sigma}$	1.32	5.28

6

Results

We will run our model individually for games with over 150,000 total sales and price of over $10. Figures 20 through 23 are the results for the game The Witcher 3 for each of the 4 specifications. The graph to the left presents in dark blue the actual sales and in light blue the estimated potential market at each price point.

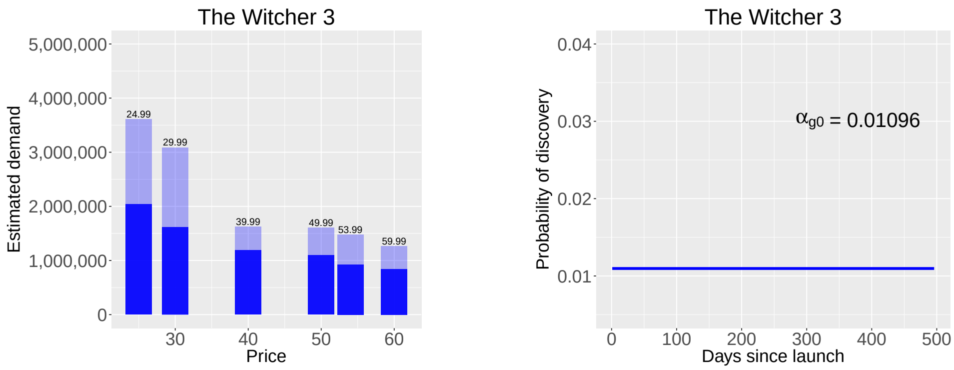

Figure 20: Estimation results for The Witcher 3, Specification 1

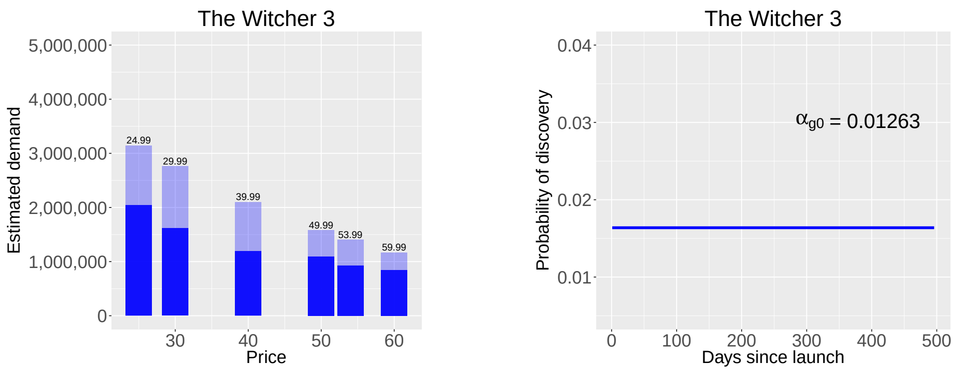

Figure 21: Estimation results for The Witcher 3, Specification 2

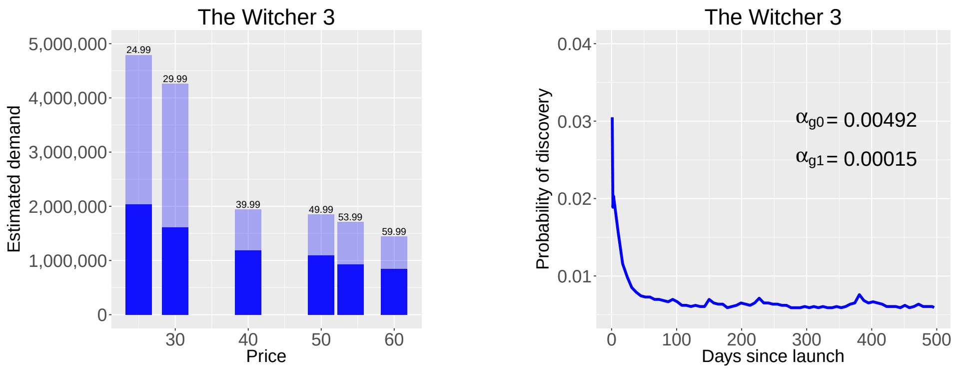

Figure 22: Estimation results for The Witcher 3, Specification 3

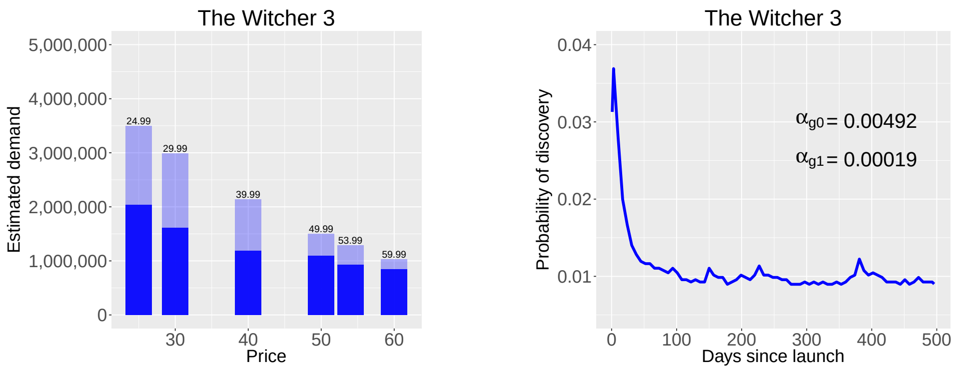

Figure 23: Estimation results for The Witcher 3, Specification 4

In Figure 24 we have the distribution of estimates for α_{g0}, our game specific fixed discovery rate, and in Table 4 some descriptive statistics [5]. In specifications 3 and 4, in which we use google trends data, we note these estimates are drastically lower and very close to 0 for many games, which would be expected if websearching is related to product discovery. On a game by game basis, note specifications 1 and 2 generally show similar estimates, as do specifications 3 and 4.

[5]Full results are in the appendix

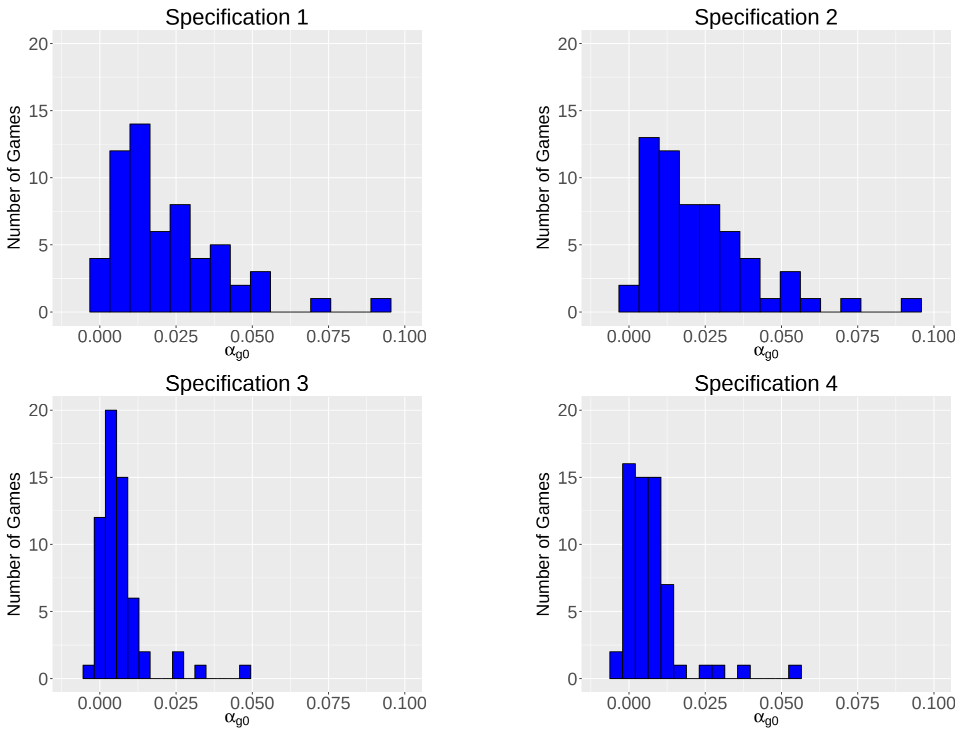

Figure 24: Distribution of α_{g0} estimates

Table 4: Estimated α_{g0}

	Specification 1	Specification 2	Specification 3	Specification 4
Mean	0.022	0.023	0.007	0.007
SD	0.018	0.018	0.008	0.010
Minimum	0.002	0.002	-0.005	-0.005
1st Quartile	0.009	0.010	0.002	0.002
Median	0.016	0.019	0.005	0.005
3rd Quartile	0.030	0.032	0.008	0.008
Maximum	0.094	0.094	0.046	0.053

In Figure 25 we have the distribution of our α_{g1} estimates, and in Table 5 some descriptive statistics. Because our web search data is normalized, these estimates are not actually comparable, though we note that almost all estimated α_{g1} are positive, as we would expect with a positive relationship between web searches and product discovery.

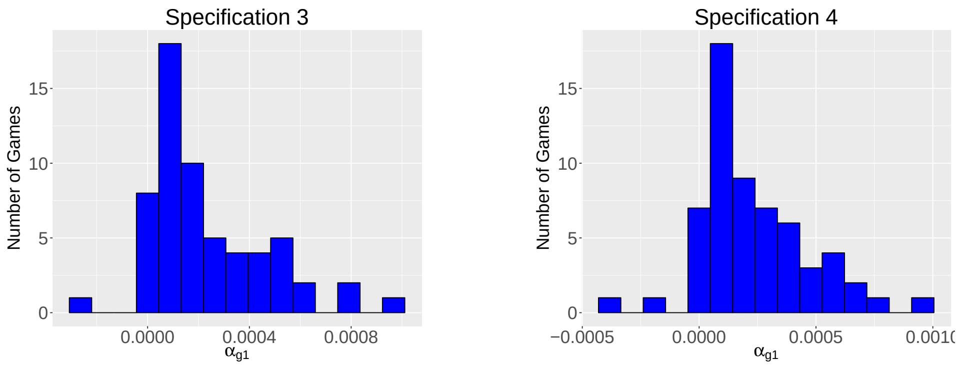

Figure 25: Distribution of α_{g1} estimates

Table 5: Estimated α_{g1}

	Specification 3	Specification 4
Mean	0.00023	0.00024
SD	0.00023	0.00023
Minimum	-0.00028	-0.00039
1st Quartile	0.00008	0.00008
Median	0.00015	0.00019
3rd Quartile	0.00034	0.00038
Maximum	0.00095	0.00095

Because α_{g0} and α_{g1} are both positive in most games, we can look at how much each one contributes to the total discovery rate q_{gt}. Naturally, because game exposure is highest right after launch, $\alpha_{g1} * GT_{gt}$ is also the highest at this point, being on average responsible for 75% of q_{gt} at launch, 43% on average and 25% at the lowest point. These estimates are very similar for both specifications 3 and 4.

Specifications 1 and 2 are nested in specifications 3 and 4, so we can run a likelihood ratio test to see for which games we can reject $\alpha_{g1} = 0$. The likelihood differences are reported in Figure 26. We reject $\alpha_{g1} = 0$ for all but one game in specification 4.

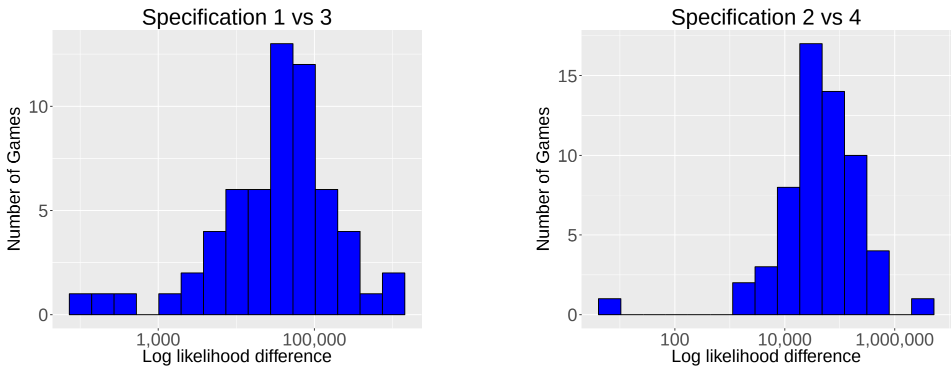

Figure 26: Distribution of Log likelihood differences between models

Note: Logarithmic scale

In Table 6 we present descriptive statistics of the estimated potential market at full price, and in Figure 27 we put the estimates side by side, ordered by the the specification 4 estimate.

Table 6: Estimated potential markets at full price

	Specification 1	Specification 2	Specification 3	Specification 4
Mean	626,667	622,384	751,920	731,330
SD	706,689	672,334	753,387	690,473
Minimum	114,841	134,585	137,695	134,036
1st Quartile	232,692	225,932	360,274	360,031
Median	390,302	424,287	554,244	557,905
3rd Quartile	738,224	797,831	848,056	868,835
Maximum	4,468,139	4,212,486	4,986,261	4,422,369

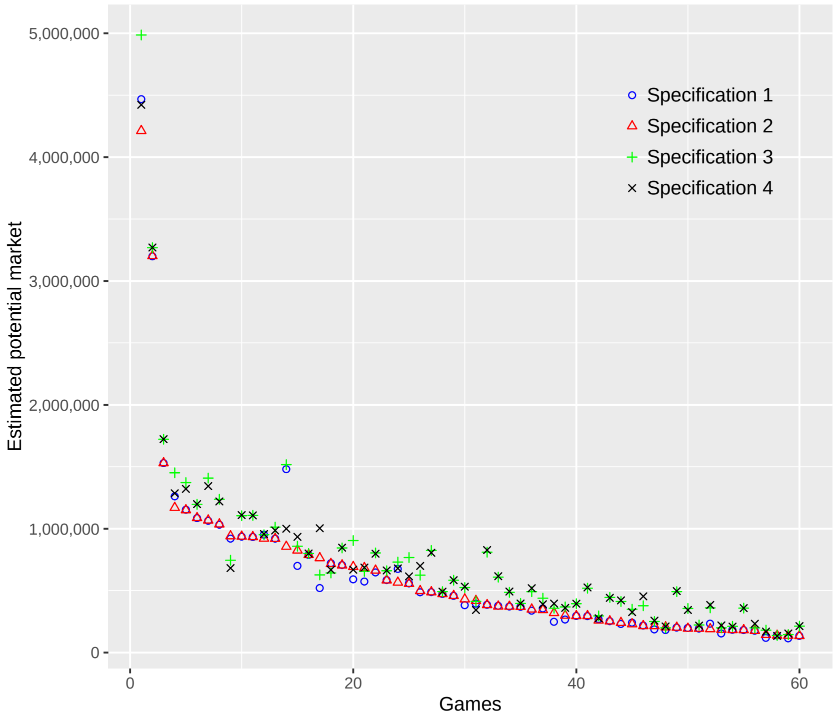

Figure 27: Estimated potential markets at full price for different model specifications

Our potential market estimates are similar for every game on a specification by specification basis, though the specifications that allow the discovery rate q_{gt} to change over time are higher on average.

Looking at the difference between the estimated market and actual sales, we have an estimate of how much sales have been lost (or have yet to be obtained) at each price point due to incomplete information. In Figure 28 we present these estimates for specification 4. The blue points are the games' effective sales at full price after 90 days, and the red points are the estimated markets at full price.

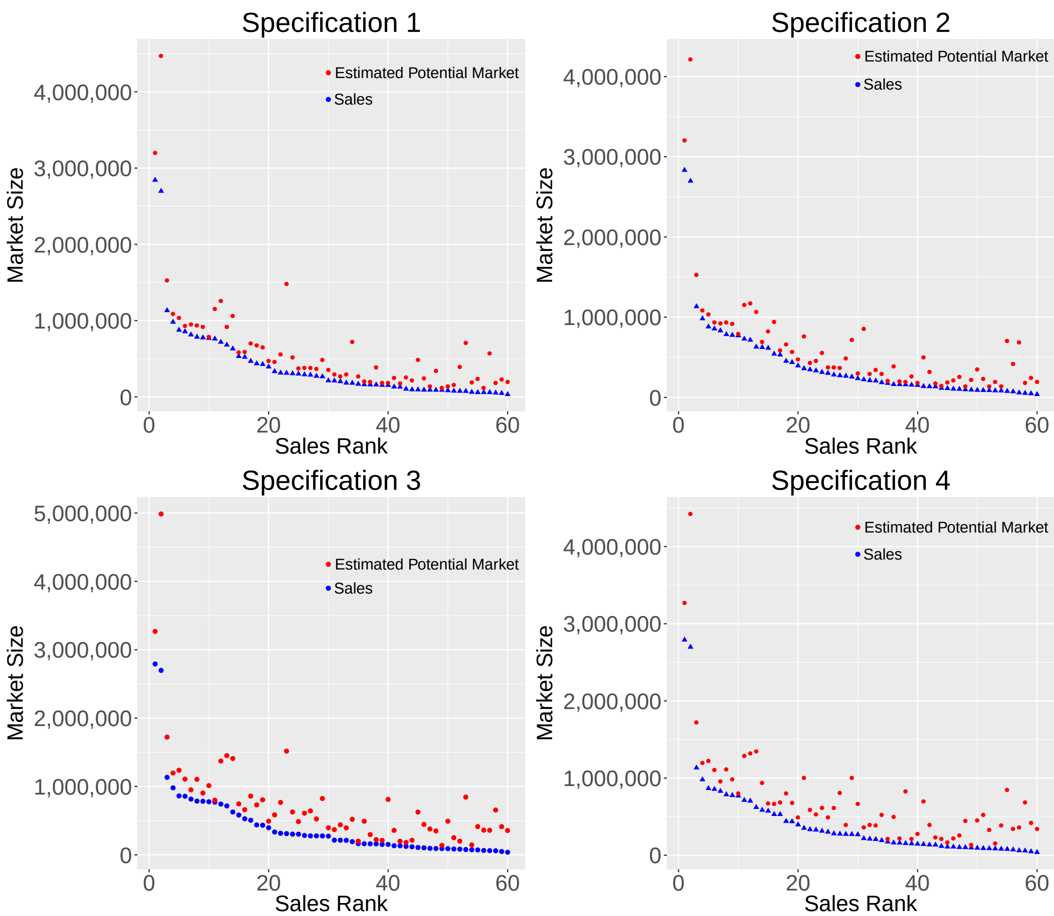

Figure 28: Total sales and potential market at full price

Even for our sample of relatively successful games, there is a wide variance between the potential market and effective sales after 90 days. The ratio between effective sales and potential market, which we will refer to as "market filled", holds a positive relationship with effective sales, though this relationship is mechanical in nature. By comparing potential market, which is directly related to intrinsic quality, and market filled, which a is measure of how much a game is known, we find no such relationship (Figure 29), no matter the specification.

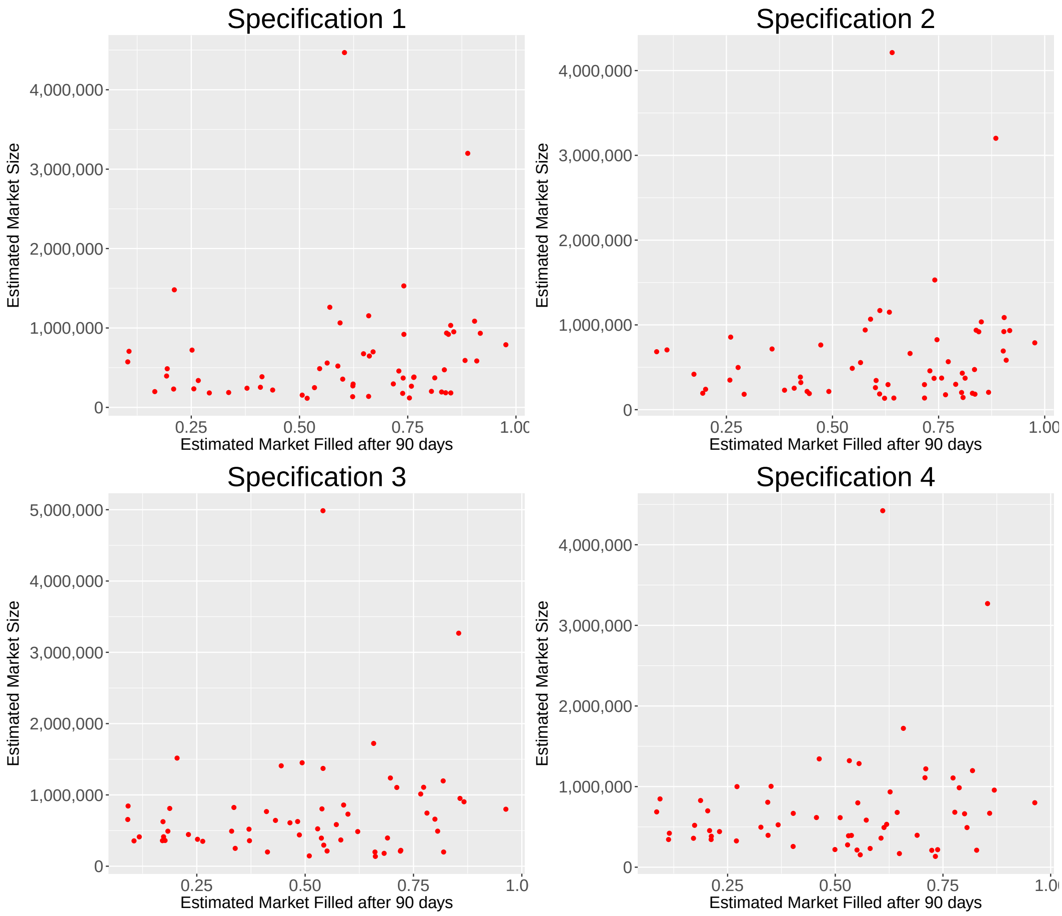

Figure 29: Estimated potential market x estimated market filled at full price

We might expect games that have sold very well to be more well known because of word-of-mouth, but this is not consistently so. For example, among these we estimate that Grand Theft Auto V, with 2.69 million units sold after 90 days, had only sold to 54% of its potential market, in contrast to Fallout 4, with 2.79 million units sold, that had sold to 85% of its potential market (specification 4).

This contrasts with Hendricks and Sorensen's (2009) results for the music industry, that find that albums at the very top of sales charts are almost universally known. We believe this is related to a difference in search costs. In the period in which they study the music industry, product discovery happened primarily through radio play, therefore discovering a new artist and assessing quality happens simultaneously and at no extra cost. With PC games, on the other hand, search costs for assessing game quality are not negligible, so despite public proeminence of a game, a substantial part of the consumer base may still remain relatively uninformed about its quality.

Overall, we find that on average games have sold to around 60% of their potential market after 90 days in specifications 1 and 2, and 50% in specifications 3 and 4, as shown in Figure 30 and Table 7.

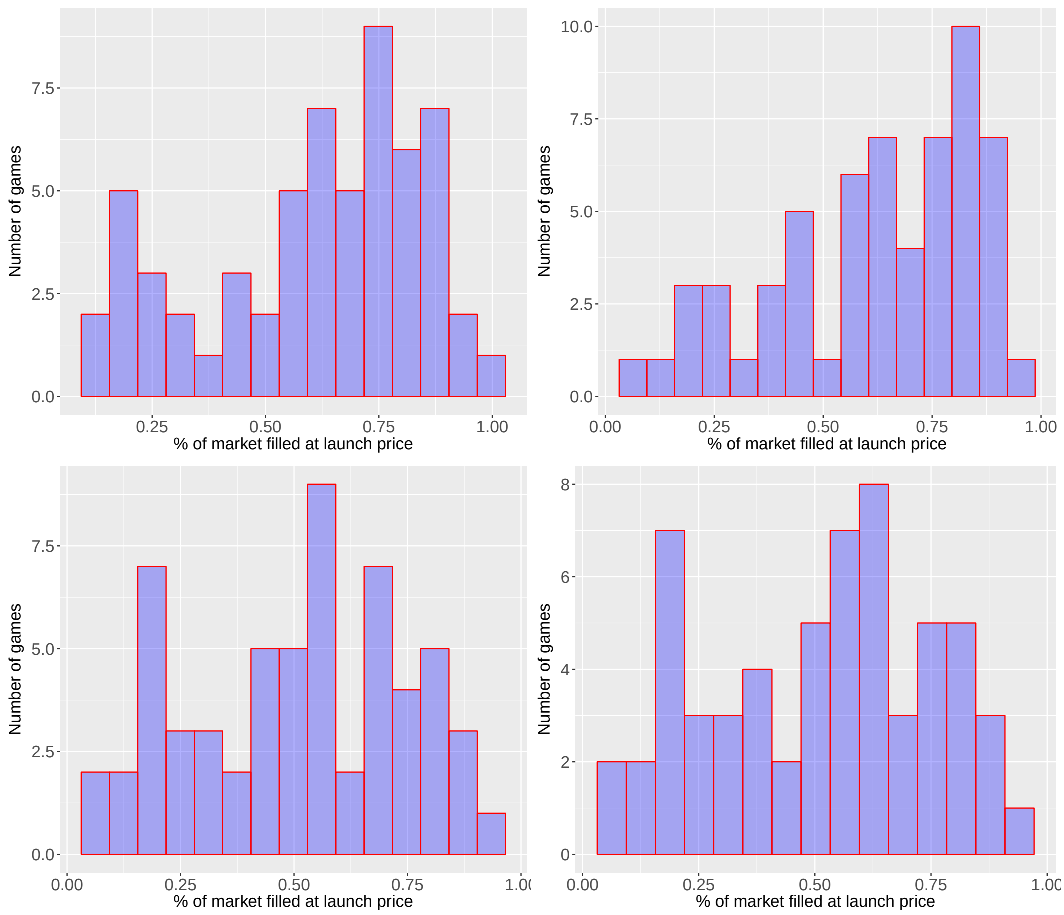

Figure 30: Distribution of estimated market filled at full price after 90 days

Specifications 2 and 4 gives us an estimate of β_g and σ_g, that is, the average value per hour of play consumers give game g and the variance of tastes around that average. For the three games in our sample that only appear with a single price, β_g and σ_g are not identified separately, so we do not report those games here. For another three games in our database, estimates of β_g and σ_g were many orders of magnitude higher than the others, which is an indication that for those games our assumption of a normal distribution of idiossincratic tastes and, consequently, of a convex demand curve is wrong. For the remaining games, in Figure 31 we present the estimated pairs.

Table 7: Market filled at full price

	Specification 1	Specification 2	Specification 3	Specification 4
Mean	0.51	0.52	0.60	0.62
SD	0.24	0.24	0.24	0.23
Minimum	0.09	0.09	0.10	0.09
1st Quartile	0.33	0.34	0.43	0.44
Median	0.54	0.55	0.65	0.64
3rd Quartile	0.69	0.71	0.81	0.81
Maximum	0.96	0.96	0.98	0.98

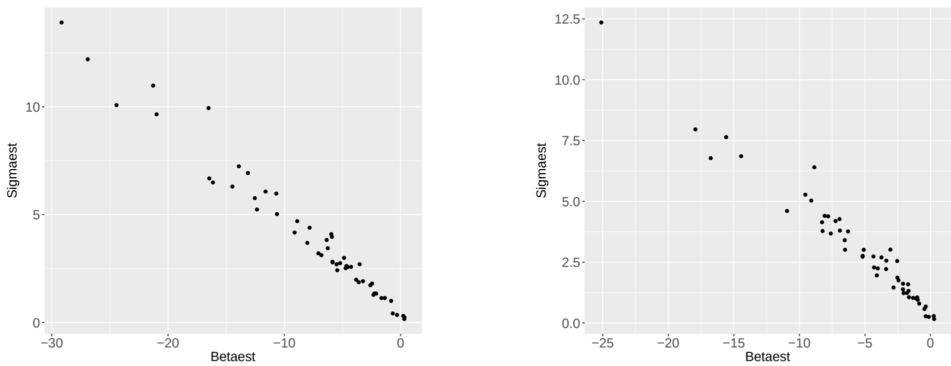

Figure 31: β_g and σ_g pair estimates

Nearly all estimates of β_g are negative. This tells us that on average, the associated time costs with both installing and playing a game in order to enjoy it are greater than the utility derived from it, but because tastes are heterogenous, it still will be purchased by some. Games with lower β_g and higher σ_g are what we would consider more ”niche” titles, and are games for which a change in price will yield less new sales than games with higher β_g and lower σ_g. Instead of looking at the average value per hour, β_g, we can look at the average value per hour of play of those consumers who effectively purchased the game at full price. This, along with the average price per hour, gives us a notion of consumer surplus for each game, and we show this in Figure 32. Game cost per hour of play is presented in red and average willingness to pay at full price is presented in blue

We estimate consumer surplus for buyers at full price is on average 100% of the full price in specification 2 and 70% of the full price in specification 4.

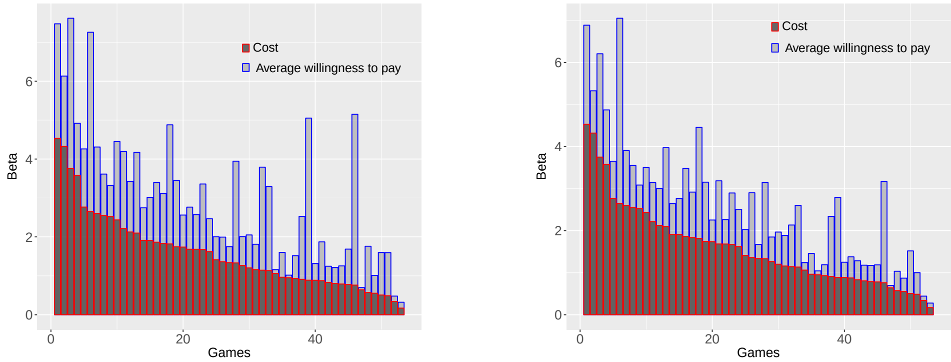

Figure 32: Willingness to pay and Cost per hour of gameplay

Furthermore, with β_g and σ_g in hand we can estimate the potential market size at prices other than the ones that have already been sold at. In Figure 33 we present the distribution of estimated elasticities.

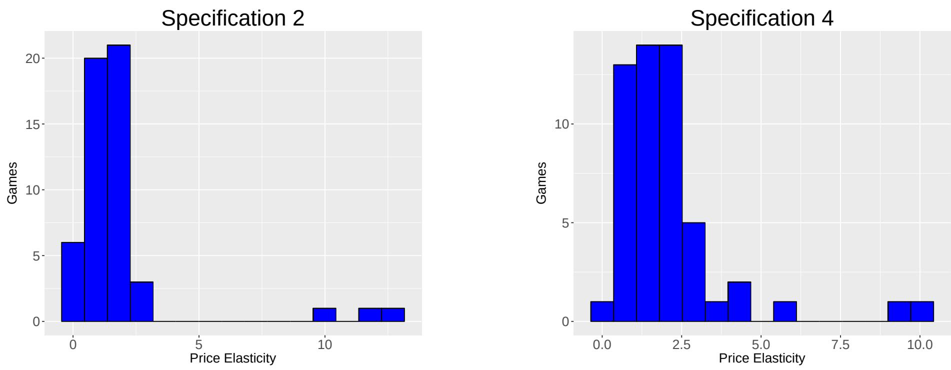

Figure 33: Estimated demand elasticities

In both specifications (but more pronounced in specification 2), we have three games that appear as outliers to the rest. We note these three games, contrary to most others, launched at a discount and not at full price, which may explain why our estimates for them are far away from others. Excluding these 3, we have an average price elasticity of -1.32 for specification 2 and -1.98 for specification 4. These are the elasticities of the potential market for each game, calculated at full price. We find that by calculating these elasticities in a simpler way, by looking at the number of units sold per day at each price point, ends up massively overstating elasticities, with an average estimated elasticity of -13.29, and sometimes finds positive elasticities. High average estimated elasticity happens because average sales per day can be very high if the game was only discounted for a few days, and positive elasticities can appear if a game doesn't have a large potential market or isn't well marketed after being discounted, so average daily sales end up lower than at full price. We show these estimated elasticities in Figure 34.

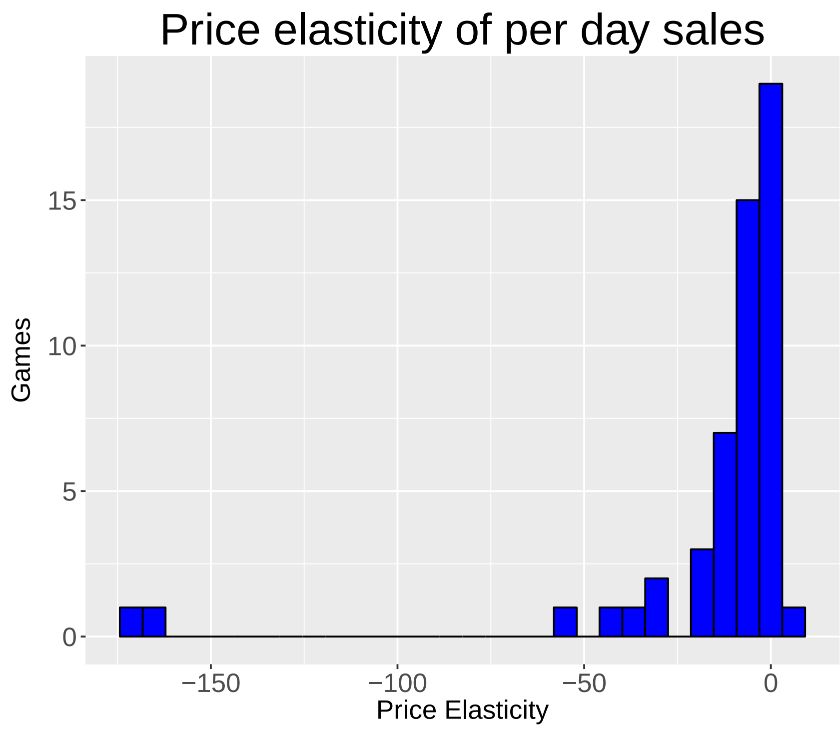

Figure 34: Estimated elasticities of per day sales

7

Conclusion

In this paper we have argued for the importance of product awareness in markets with large and ever increasing choice sets. Product awareness is an inherently difficult object to measure, so instead of testing it directly we presented a structural demand model in which product awareness is central and showed how we can replicate different sales patterns present in the PC games market, from a convex decline in sales with sales primarily in the first few months, to a linear decline in sales, to a constant pattern of sales.

We took this model to the data using PC game sales data from the Steam distribution platform, and websearch data from google trends. In testing different specifications for product dicovery in our model we have shown the importance of websearching in determining product awareness, and proposed a quantitative measure of it.

By using the average number of hours played for each game we were able to partially account for differences in game quality and estimate the distribution of consumer opinions about each game's value per hour of play, allowing us to estimate consumer surplus on average from 70% to 100% of each game's full price depending on the specification and estimate average price elasticity on verage from -1.98 to -1.32 depending on specification.

Furthermore, we have shown that for relatively successful games, 90 days after they are launched they have sold an average to only 50% of their potential markets in some model specifications and 60% in others, but that there is much variation around this average, and we find no relationship in how much of the potential market has been sold to and the product's inherent quality. These findings indicate that the distribution of sales in these markets are substantially different than it would be in a world in which consumers have complete information.

8

References

ACKERBERG, A.; Empirically Distinguishing Informative and Prestige Effects of Advertising. **The RAND Journal of Economics**, Vol. 32, No. 2, pp. 316-333, Summer 2001

ACKERBERG, A.; Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. **International Economic Review**, Vol. 44, No. 3, p. 1007-1040, Aug. 2003

BANERJEE, A.V.; A Simple Model of Herd Behaviour. **The Quarterly Journal of Economics**, Vol. 107, No 3, p. 797-817, Aug. 1992

BIKHCHANDANI, S.; HIRSHLEIFER, D.; WELSH, I. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. **Journal of Political Economy**, Vol. 100, No 2, p. 992-1026, Oct. 1992

ERDEM, T.; KEANE, M.P.; Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. **Marketing Science**, Vol. 15, No. 1, p. 1-20, 1996

GOEREE, M.S.; Limited Information and Advertising in the U.S. Personal Computer Industry. **Econometrica**, Vol. 76, No. 5, p. 1017-1074, Sep. 2008

HENDRICKS, K.; SORENSEN, A.; Information and the Skewness of Music Sales. **Journal of Political Economy**, Vol. 117, No. 2, p. 324-369, Apr. 2009

HENDRICKS, K; SORENSEN, A.; WISEMAN, T. Observational Learning and Demand for Search Goods. **American Economic Journal: Microeconomics**, Vol. 4, No 1, p. 1-31, Feb 2012

HOYER, W.D.; BROWN, S.P.; Effects of Brand Awareness on Choice for a Common, Repeat-Purchase Product. **Journal of Consumer Research**, Vol. 17, No 2, p. 141-148, 1990

KUMAR, A.; SMITH, M.D.; TELANG, R.; Information Discovery and the Long Tail of Motion Picture Content. **Management Information Systems Quarterly**, Vol. 38, Issue 4, p. 1057-1078, 2014

LEE, R.; Vertical Integration and Exclusivity in Platform and Two-Sided Markets. **American Economic Review**, Vol. 103, No 7, p. 2960-3000, 2013

MORETTI, E.; Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. **Review of Economic Studies**, Vol. 78, p.356-393, 2011

NAIR, H.; Intertemporal Price Discrimination with Forward-Looking Consumers: Application to the US Market for Console Video-Games. **Quantitative Marketing and Economics**, Vol. 5, Issue 3, p.239-292, Sep 2011

SALGANIK, M.J.; Dodds, P.S., Watts, D.J.; Experimental study of inequality and unpredictability in an artificial cultural market. **Science**, Vol. 311, Issue 5762, p.854-856, Feb. 2006

8

Appendix

Table 8: Estimates for α_{g0}

	S1	S2	S3	S4
Tabletop Simulator	0.0023	0.0052	-0.0007	-0.0022
Tom Clancy's Rainbow Six Siege	0.0123	0.0092	0.0058	0.0034
Crypt of the NecroDancer	0.0036	0.0042	0.0019	0.0014
Verdun	0.0029	0.0040	0.0015	0.0010
Fallout 4	0.0373	0.0378	0.0071	0.0073
Enter the Gungeon	0.0299	0.0341	0.0067	0.0080
Tom Clancy's The Division	0.0943	0.0943	0.0313	0.0313
Far Cry Primal	0.0429	0.0427	0.0079	0.0084
Battlefleet Gothic: Armada	0.0282	0.0281	0.0015	0.0014
Galactic Civilizations III	0.0103	0.0069	0.0038	0.0037
Dead by Daylight	0.0237	0.0148	0.0459	0.0533
Mortal Kombat X	0.0122	0.0126	0.0036	0.0071
Grand Theft Auto V	0.0100	0.0231	0.0013	0.0000
Total War: Warhammer	0.0507	0.0507	0.0073	0.0073
Rise of the Tomb Raider	0.0221	0.0283	0.0048	0.0027
Invisible, Inc.	0.0158	0.0189	0.0061	0.0072
Firewatch	0.0403	0.0440	0.0056	0.0064
Naruto Shippuden: Ultimate Ninja Storm 4	0.0106	0.0106	0.0029	0.0021
XCOM 2	0.0387	0.0385	0.0046	0.0045
Street Fighter V	0.0371	0.0372	0.0097	0.0114
Hyper Light Drifter	0.0456	0.0535	0.0041	0.0052
Move or Die	0.0044	0.0040	0.0049	0.0065
Call of Duty: Black Ops III	0.0217	0.0215	0.0046	0.0057
DARK SOULS III	0.0716	0.0714	0.0149	0.0149
American Truck Simulator	0.0296	0.0296	0.0044	0.0044
Hearts of Iron IV	0.0361	0.0361	0.0095	0.0096
Hitman	0.0273	0.0273	-0.0054	-0.0054
Metal Gear Solid V: The Phantom Pain	0.0370	0.0359	0.0085	0.0117
DOOM	0.0263	0.0265	0.0076	0.0083
Project CARS	0.0183	0.0189	0.0058	0.0081

Table 9: Estimates for α_{g0}

	S1	S2	S3	S4
DiRT Rally	0.0137	0.0136	0.0058	0.0072
Hyperdimension Neptunia Re;Birth2	0.0052	0.0058	0.0018	0.0018
Minecraft: Story Mode	0.0144	0.0173	0.0064	0.0027
Duck Game	0.0024	0.0024	0.0012	0.0011
Final Fantasy IX	0.0105	0.0105	0.0028	0.0028
Darkest Dungeon	0.0315	0.0336	0.0076	0.0143
Life is Feudal: Your Own	0.0089	0.0089	0.0002	0.0002
The Witcher 3: Wild Hunt	0.0110	0.0126	0.0049	0.0059
Stellaris	0.0545	0.0545	0.0118	0.0120
Plague Inc: Evolved	0.0129	0.0228	0.0124	0.0086
Just Cause 3	0.0153	0.0165	0.0055	0.0058
Lightning Returns: Final Fantasy XIII	0.0071	0.0134	0.0020	0.0006
Assassin's Creed Syndicate	0.0251	0.0310	0.0262	0.0393
Reign Of Kings	0.0158	0.0158	0.0112	0.0112
Stardew Valley	0.0205	0.0205	0.0081	0.0081
Football Manager 2016	0.0136	0.0138	0.0015	0.0025
Superhot	0.0527	0.0575	0.0270	0.0259
Helldivers	0.0166	0.0169	0.0029	0.0028
Batman: Arkham Knight	0.0089	0.0118	0.0025	0.0009
Rebel Galaxy	0.0181	0.0259	0.0072	0.0073
Audiosurf 2	0.0072	0.0089	0.0043	0.0049
Shadowrun: Hong Kong	0.0049	0.0046	0.0023	0.0021
Wolfenstein: The Old Blood	0.0096	0.0100	0.0015	0.0015
Dragon's Dogma: Dark Arisen	0.0301	0.0301	0.0015	0.0015
LEGO Jurassic World	0.0088	0.0093	0.0034	0.0036
Tales of Zestiria	0.0071	0.0071	0.0008	0.0009
Final Fantasy X/X-2 HD Remaster	0.0249	0.0249	0.0107	0.0107
One Piece Pirate Warriors 3	0.0055	0.0051	0.0018	0.0017
Devil May Cry 4 Special Edition	0.0022	0.0017	0.0019	0.0017
Victory: The Age of Racing	0.0270	0.0270	0.0138	0.0138

Table 10: Estimates for α_{g1}

	S3	S4
Tabletop Simulator	0.00008	0.00019
Tom Clancy's Rainbow Six Siege	0.00013	0.00011
Crypt of the NecroDancer	0.00003	0.00004
Verdun	0.00010	0.00012
Fallout 4	0.00051	0.00051
Enter the Gungeon	0.00016	0.00023
Tom Clancy's The Division	0.00095	0.00095
Far Cry Primal	0.00064	0.00064
Battlefleet Gothic: Armada	0.00012	0.00012
Galactic Civilizations III	0.00008	0.00006
Dead by Daylight	-0.00028	-0.00039
Mortal Kombat X	0.00011	0.00011
Grand Theft Auto V	0.00022	0.00022
Total War: Warhammer	0.00051	0.00051
Rise of the Tomb Raider	0.00028	0.00037
Invisible, Inc.	0.00017	0.00024
Firewatch	0.00060	0.00062
Naruto Shippuden: Ultimate Ninja Storm 4	0.00006	0.00005
XCOM 2	0.00051	0.00051
Street Fighter V	0.00043	0.00041
Hyper Light Drifter	0.00077	0.00067
Move or Die	-0.00003	-0.00004
Call of Duty: Black Ops III	0.00027	0.00025
Dark Souls III	0.00076	0.00076
American Truck Simulator	0.00035	0.00035
Hearts of Iron IV	0.00029	0.00029
Hitman	0.00056	0.00056
Metal Gear Solid V: The Phantom Pain	0.00042	0.00039
DOOM	0.00021	0.00020
Project CARS	0.00022	0.00023

Table 11: Estimates for α_{g1}

	S3	S4
DiRT Rally	0.00011	0.00009
Hyperdimension Neptunia Re;Birth2	0.00004	0.00004
Minecraft: Story Mode	0.00015	0.00018
Duck Game	0.00001	0.00002
Final Fantasy IX	0.00006	0.00006
Darkest Dungeon	0.00034	0.00024
Life is Feudal: Your Own	0.00010	0.00010
The Witcher 3: Wild Hunt	0.00015	0.00020
Stellaris	0.00054	0.00054
Plague Inc: Evolved	0.00002	0.00021
Just Cause 3	0.00013	0.00013
Lightning Returns: Final Fantasy XIII	0.00005	0.00008
Assassin's Creed Syndicate	-0.00003	-0.00019
Reign Of Kings	0.00030	0.00030
Stardew Valley	0.00013	0.00013
Football Manager 2016	0.00012	0.00011
Superhot	0.00048	0.00055
Helldivers	0.00013	0.00013
Batman: Arkham Knight	0.00020	0.00028
Rebel Galaxy	0.00032	0.00040
Audiosurf 2	0.00006	0.00006
Shadowrun: Hong Kong	0.00004	0.00004
Wolfenstein: The Old Blood	0.00011	0.00011
Dragon's Dogma: Dark Arisen	0.00034	0.00033
LEGO Jurassic World	0.00005	0.00006
Tales of Zestiria	0.00009	0.00008
Final Fantasy X/X-2 HD Remaster	0.00042	0.00042
One Piece Pirate Warriors 3	0.00005	0.00004
Devil May Cry 4 Special Edition	0.00000	0.00000
Victory: The Age of Racing	0.00019	0.00019

Table 12: Estimates for β_g

	S2	S4
Tabletop Simulator	-2.46	-7.24
Tom Clancy's Rainbow Six Siege	-2.11	-1.33
Crypt of the NecroDancer	-2.61	-2.44
Verdun	-4.87	-8.07
Fallout 4	-3.53	-1.02
Enter the Gungeon	-5.85	-2.05
Tom Clancy's The Division	-71.78	-69.62
Far Cry Primal	-14.47	-6.54
Battlefleet Gothic: Armada	-55393431.28	-2319870.03
Galactic Civilizations III	-5.51	-4.30
Dead by Daylight	0.31	0.28
Mortal Kombat X	-29.16	-25.11
Grand Theft Auto V	-0.82	-0.36
Total War: Warhammer	-5.82	-23.00
Rise of the Tomb Raider	-5.96	-2.54
Invisible, Inc.	-12.35	-8.23
Firewatch	-10.70	-3.06
Naruto Shippuden: Ultimate Ninja Storm 4	-12.54	-6.29
XCOM 2	-13.92	-4.35
Street Fighter V	-16.15	-10.94
Hyper Light Drifter	-4.57	-2.52
Move or Die	-5.92	-6.94
Call of Duty: Black Ops III	-4.27	-1.68
DARK SOULS III	-21.29	-9.55
American Truck Simulator	-7.07	-2.82
Hearts of Iron IV	-332871.58	-927142.29
HITMAN	-5.05	-9.37
Metal Gear Solid V: The Phantom Pain	-8.90	-6.91
DOOM	-16.52	-8.86
Project CARS	-10.63	-8.27

Table 13: Estimates for β_g

	S2	S4
DiRT Rally	-6.27	-5.09
Hyperdimension Neptunia Re;Birth2 Sisters Generation	-5.87	-5.17
Minecraft: Story Mode - A Telltale Games Series	-8.03	-5.18
Duck Game	-1.65	-1.79
Final Fantasy IX	0.32	0.25
Darkest Dungeon	-3.24	-2.10
Life is Feudal: Your Own	-39.76	-196565.62
The Witcher 3: Wild Hunt	-1.36	-0.97
Stellaris	-26.91	-14.44
Plague Inc: Evolved	-13.13	-9.10
Just Cause 3	-6.36	-3.37
Lightning Returns: Final Fantasy XIII	-2.25	-1.10
Assassin's Creed Syndicate	-16.46	-16.76
Reign Of Kings	-6.81	-6.52
Stardew Valley	-0.31	-0.12
Football Manager 2016	-0.68	-0.36
Superhot	-20.99	-15.59
Helldivers	-3.83	-4.02
Batman: Arkham Knight	-11.62	-3.74
Rebel Galaxy	-5.45	-4.09
Audiosurf 2	-9.12	-7.60
Shadowrun: Hong Kong	-2.34	-1.64
Wolfenstein: The Old Blood	-4.66	-1.71
Dragon's Dogma: Dark Arisen	-3.61	-0.86
LEGO Jurassic World	-24.44	-17.94
Tales of Zestiria	-4.73	-2.09
Final Fantasy X/X-2 HD Remaster	0.22	-0.46
One Piece Pirate Warriors 3	-5.21	-3.39
Devil May Cry 4 Special Edition	-7.83	-7.80
Victory: The Age of Racing	-3.22	4.78

Table 14: Estimates for σ_g

	S2	S4
Tabletop Simulator	1.80	4.19
Tom Clancy's Rainbow Six Siege	1.34	1.03
Crypt of the NecroDancer	1.73	1.75
Verdun	2.99	4.40
Fallout 4	2.70	1.05
Enter the Gungeon	2.78	1.23
Tom Clancy's The Division	33.69	32.77
Far Cry Primal	6.30	3.40
Battlefleet Gothic: Armada	20912542.11	995982.67
Galactic Civilizations III	2.70	2.28
Dead by Daylight	0.16	0.17
Mortal Kombat X	13.91	12.35
Grand Theft Auto V	1.00	0.68
Total War: Warhammer	3.28	11.93
Rise of the Tomb Raider	4.10	2.55
Invisible, Inc.	5.24	3.78
Firewatch	5.98	3.02
Naruto Shippuden: Ultimate Ninja Storm 4	5.76	3.76
XCOM 2	7.24	2.73
Street Fighter V	6.49	4.60
Hyper Light Drifter	2.57	1.86
Move or Die	3.97	4.27
Call of Duty: Black Ops III	2.58	1.32
Dark Souls III	10.98	5.28
American Truck Simulator	3.20	1.46
Hearts of Iron IV	136674.44	397591.01
Hitman	2.39	4.28
Metal Gear Solid V: The Phantom Pain	4.70	3.80
DOOM	9.94	6.40
Project CARS	5.03	4.14

Table 15: Estimates for σ_g

	S2	S4
DiRT Rally	3.44	3.01
Hyperdimension Neptunia Re;Birth2 Sisters Generation	2.81	2.76
Minecraft: Story Mode - A Telltale Games Series	3.69	2.73
Duck Game	1.13	1.23
Final Fantasy IX	0.23	0.29
Darkest Dungeon	1.91	1.38
Life is Feudal: Your Own	17.30	91786.01
The Witcher 3: Wild Hunt	1.13	0.95
Stellaris	12.20	6.85
Plague Inc: Evolved	6.93	5.03
Just Cause 3	3.83	2.56
Lightning Returns: Final Fantasy XIII	1.34	1.01
Assassin's Creed Syndicate	6.68	6.77
Reign Of Kings	3.12	3.01
Stardew Valley	0.35	0.25
Football Manager 2016	0.42	0.28
Superhot	9.65	7.64
Helldivers	1.98	2.24
Batman: Arkham Knight	6.07	2.70
Rebel Galaxy	2.42	1.96
Audiosurf 2	4.17	3.68
Shadowrun: Hong Kong	1.28	1.06
Wolfenstein: The Old Blood	2.62	1.59
Dragon's Dogma: Dark Arisen	1.86	0.80
LEGO Jurassic World	10.08	7.96
Tales of Zestiria	2.52	1.61
Final Fantasy X/X-2 HD Remaster	0.30	0.58
One Piece Pirate Warriors 3	2.75	2.22
Devil May Cry 4 Special Edition	4.39	4.39
Victory: The Age of Racing	3.46	0.18

Table 16: Estimate for market at full price

	S1	S3
Tabletop Simulator	1480533	1516668
Tom Clancy's Rainbow Six Siege	520696	627048
Crypt of the NecroDancer	486612	624973
Verdun	721398	642739
Fallout 4	3199110	3268283
Enter the Gungeon	382274	524385
Tom Clancy's The Division	788702	799754
Far Cry Primal	183806	211590
Battlefleet Gothic: Armada	201797	491601
Galactic Civilizations III	248436	357877
Dead by Daylight	919828	745332
Mortal Kombat X	558589	766839
Grand Theft Auto V	4468139	4986261
Total War: Warhammer	933284	1107279
Rise of the Tomb Raider	699583	858469
Invisible, Inc.	154268	199616
Firewatch	266676	368425
Naruto Shippuden: Ultimate Ninja Storm 4	218417	377771
XCOM 2	935938	1104581
Street Fighter V	193757	222634
Hyper Light Drifter	118819	180746
Move or Die	394838	413195
Call of Duty: Black Ops III	1153577	1371747
Dark Souls III	1085561	1196655
American Truck Simulator	457977	584103
Hearts of Iron IV	371777	486132
Hitman	295577	394169
Metal Gear Solid V: The Phantom Pain	919914	1012793
DOOM	1032409	1237613
Project CARS	355821	438918

Table 17: Estimate for market at full price

	S1	S3
DiRT Rally	270443	295838
Hyperdimension Neptunia Re;Birth2 Sisters Generation	198613	355520
Minecraft: Story Mode - A Telltale Games Series	176402	199227
Duck Game	705935	844584
Final Fantasy IX	181900	358804
Darkest Dungeon	591058	903949
Life is Feudal: Your Own	487717	823902
The Witcher 3: Wild Hunt	1260998	1451183
Stellaris	584816	660200
Plague Inc: Evolved	951804	951090
Just Cause 3	646766	804350
Lightning Returns: Final Fantasy XIII	233281	360763
Assassin's Creed Syndicate	138178	137695
Reign Of Kings	473347	491775
Stardew Valley	1529791	1722555
Football Manager 2016	1064020	1408779
Superhot	181850	199592
Helldivers	294042	519585
Batman: Arkham Knight	675189	730523
Rebel Galaxy	114841	144945
Audiosurf 2	187006	250512
Shadowrun: Hong Kong	338206	491306
Wolfenstein: The Old Blood	385765	810476
Dragon's Dogma: Dark Arisen	377104	609017
LEGO Jurassic World	241333	349487
Tales of Zestiria	253143	444741
Final Fantasy X/X-2 HD Remaster	134588	213251
One Piece Pirate Warriors 3	230926	411931
Devil May Cry 4 Special Edition	573274	655421
Victory: The Age of Racing	369590	396029