CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Second-hand Board Game Price Analysis

Bachelor's thesis

Author: Lukáš Novotný

Study program: Economics and Finance

Supervisor: Mgr. Petr Polák, MSc. Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 3, 2023

Lukáš Novotný

Abstract

This thesis investigates the determinants of price and value retention of pre-owned board games in the United States market. It is most likely the first thesis applying the hedonic pricing method to the board game industry. Two dependent variables, price and the portion of the original manufacturer's suggested retail price that remains, were modelled using various game characteristics. The analysis is performed on data obtained primarily from BoardGameGeek. Applying multiple linear regression and Ordinary Least Squares method on the cross-sectional sample of over 2000 observations, several factors were estimated to be significant. Condition, rating, age, complexity, duration and weight of the box turned out to be the most crucial board game value drivers. Moreover, in the case of regression on residual price share, the most significant predictors appeared to be condition, age and rating.

Keywords second-hand market, board game, hedonic pricing model, OLS regression

Title Second-hand Board Game Price Analysis

Abstrakt

Tato bakalářská práce zkoumá faktory ovlivňující cenu a uchování hodnoty deskových her z druhé ruky na trhu Spojených států amerických. Pravděpodobně se jedná o první tezi, která aplikuje hedonický cenový model pro odvětví deskových her. Dvě závislé proměné, cena a podíl z doporučené maloobchodní ceny, byly modelovány pomocí nejrůznějších herních charakteristik. Analýza byla provedena za pomoci dat získaných primárně z BoardGameGeek. Použitím vícenásobné lineární regrese a metody nejmenších čtverců na vzorek čítající přes 2000 pozorování bylo zjištěno několik významných proměných. Proměnné udávající stav, hodnocení, stáří, složitost jakož i váhu krabice se ukázaly jako nejdůležitější faktory určující hodnotu deskové hry. Navíc v případě regrese pro podíl zůstatkové ceny se jako nejvýznamnější jevily stáří, hodnocení a herní stav.

Klíčová slova bazar, desková hra, hedonický cenový model, metoda nejmenších čtverců

Název práce Analýza bazaru deskových her

Acknowledgments

I would like to express my gratitude to Mgr. Petr Polák, MSc. Ph.D.for his guidance, patience and valuable advice. In addition, many thanks belong to my parents for supporting me during my studies.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Novotný, Lukáš: *Second-hand Board Game Price Analysis.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 59. Advisor: Mgr. Petr Polák, MSc. Ph.D.

Contents

List of Tables viii

List of Figures ix

Acronyms x

1 Introduction 1

2 Theoretical background 3

2.1 General overview 3

2.2 Board game market 4

2.3 Board game crowdfunding 5

3 Literature Review 7

3.1 Hedonic Price Analysis 7

3.2 Manufacturer's Suggested Retail Price 8

4 Data 10

4.1 Data sources 10

4.2 Data extraction 11

4.3 Data cleaning 13

4.4 Variables description 15

4.4.1 Dependent variables 15

4.4.2 Condition 15

4.4.3 Weight and volume of the box 16

4.4.4 Number of players 17

4.4.5 Minimum age 17

4.4.6 Time duration 18

4.4.7 Published year 18

4.4.8 Rating 18

4.4.9 Complexity rating 19
4.4.10 Categorizations 19
4.4.11 Publishers 20
4.4.12 Variables not used 20
4.5 Data limitations 20

5 Methodology **25**
5.1 Model 25

6 Discussion Results **27**
6.1 Price Modelling 27
6.2 Residual Price Share Modelling 30
6.3 Robustness Check 31
6.4 Multiple Linear Regression (MLR) assumptions 32

7 Conclusion **35**

Bibliography **41**

A Figures and Tables of Descriptive Statistics **I**

B Residuals and Fitted Values Scatter Plots **IV**

C Robustness Check **VI**

List of Tables

4.1 Frequency by Top 5 Countries 12
4.2 Frequency by Conditions 16
4.3 Descriptive Statistics of Variables 22
4.4 Description of Quantitative Variables 23
4.5 Description of Qualitative Variables 23
4.6 Description of Qualitative Variables - Categories and Mechanics 24

6.1 Regression Output - Price 28
6.2 Regression Output - Residual Price Share 34

A.1 Frequency by Minimum Age III
A.2 Frequency by Age III

C.1 Robustness Check - Price Regression VII
C.2 Robustness Check - Residual Price Regression VIII

List of Figures

A.1 Histogram of Price I
A.2 Histogram of Residual Price Share II
A.3 Scatter Plot of Price and Weight II

B.1 Scatter Plot of Residuals and Fitted Values - Regression with Price IV
B.2 Scatter Plot of Residuals and Fitted Values - Regression with Price Retention V

Acronyms

API Application Programming Interface

BGG BoardGameGeek

FIDE The International Chess Federation

HPM Hedonic Pricing Model

MLR Multiple Linear Regression

MSRP Manufacturer's Suggested Retail Price

USD U.S. dollars

OLS Ordinary Least Squares

US United States

XML Extensible Markup Language

Chapter 1

Introduction

Board games have been popular over the world for thousands of years. They might actually be considered a social event as they bring people together to play face-to-face with other playmates in contrast with video games. The board game community loves them for their entertaining realization making players feel absorbed in the game as well as their ability to develop communication skills, logical thinking and spatial imagination. Just Monopoly with over 275 million sold copies (Karwatka 2016) has been played by more than 500 million people (guinnessworldrecords ca. 2021) since its introduction in 1935. Nowadays, the board game industry with billions of dollars in market value is predicted to follow its increasing trend as board games still constitute a frequent choice for leisure-time activity despite the competition in the digital world (Statista 2019a). Thousands of new titles are created yearly to satisfy board game lovers' needs and the older used ones are thus being sold and traded on second-hand markets. If someone is passionate about constantly seeking the latest masterpieces or even collecting them, it could easily become an expensive hobby. The solution for such people might be trading on used marketplaces. Nevertheless, is it worth buying pre-owned games with respect to their suggested retail price?

There are two main objectives of this thesis, to identify relevant factors explaining the price of second-hand board games and their value retention. The research question is thus whether such determinants can be identified. To our best knowledge, this is the first study conducting hedonic pricing of board games. The analysis is conducted using data from BoardGameGeek (BGG), a website for the board gaming community including a marketplace of used games, Geekmarket. Only listings from the United States (US) were taken into consideration, though anyone from anywhere can trade there. The findings might provide useful insight for consumers by mitigating the information asymmetry when seeking underrated listings.

The thesis is structured as follows. In the Chapter 2, a brief theoretical background of the board game industry is provided. Chapter 3 examines the existing literature closely linked to our topic. Chapter 4 is devoted to the introduction of data sources, the detailed process of data cleaning, explanation of variables choices and limitations of this study. Chapter 5 comments on applied methods. In the Chapter 6, results are discussed and interpreted. Besides, it contains sections devoted to robustness checks and verification of model assumptions. Chapter 7 summarizes our findings and suggests further research related to the topic of this work.

Chapter 2

Theoretical background

This chapter consists of three main sections. The first Section 2.1 concentrates on the general overview of board games. Continuing with Section 2.2 where interesting figures about the board game market are presented. Ultimately, in the Section 2.3, the crowdfunding of board games is mentioned.

2.1 General overview

To begin with, let's stress the distinction between tabletop and board games since this thesis occasionally refers to these terms. Nevertheless, people usually do not distinguish between them in reality.

According to Rossel (ca. 2020), a tabletop game is defined as "any game that is typically played on a flat surface, primarily a table." The most common categories of tabletop games are card games, usually containing just a deck of cards as the only component, and board games, generally played by moving pawns, meeples, or pieces on top of the game board. Both of these categories usually come in a box containing their components together with the rules. The most recent board games include a combination of all possible components.

Mijal (2014) states there are a lot of possible game categorizations according to the research questions the study is trying to answer. By way of illustration, these might include classification according to the theme, the number of players, characteristic playing time, recommended minimum age to play, published year, mechanisms, category, components, complexity, level of randomness, game popularity rating, etc.

Furthermore, board games can be divided into modern games in which the author and published year are usually known and classic/traditional games which are typically hundred or thousand years old (Tomášková 2009). Traditional games typically had simple rules and remarkably many of them prevailed to the present day. They also put the basis for modern more complex ones. This paper conducts analysis only on modern games.

The earliest board games were played even thousands of years ago. The oldest known ones come from ancient Egypt approximately 3,500 before the Christian Era (Tomášková 2009). Others were created for instance in China, Greece and Africa (Tomášková 2009). The most worthy of mention seems to be Senet and Go which can be found in some changed forms listed on second-hand markets even today. However, probably the most famous classic game is Chess coming from India approximately 1 500 years ago (Averbakh 2012). It is even considered a sport by The International Chess Federation (FIDE) being played on average more than 60 million times daily (FIDE ca. 2008).

Modern games tend to be more expensive as they often include more elaborate content offering a more strategic experience. Glassdoor (2022) states the average salary of a board game designer ranges approximately from 63,000 to 107,000 U.S. dollars (USD) per year. Furthermore, most games naturally have numerous different versions that might vary in language, published year, artist, publisher, box weight, and box sizes. Besides, the best-selling games commonly have multiple reimplements and expansions. Reimplements are reworks of the original game with some gameplay and rules changes, also having the link to the original in its name. On the other hand, expansions generally add other components to the game and cannot be played without the base game. They might also change the player count and usually cost less than their corresponding base games. There exist more than 27,506 known expansions (BGG ca. 2014a).

2.2 Board game market

The popularity of the board game industry has been steadily increasing in recent years. In addition to that, the number of published games each year is also growing. The most probable causes appear to be the constant development of games including more elegant mechanics, components and graphics, or the extension of the internet enabling the selling of games online. The internet also allows playing some not-so-complex board games through electronic devices, visiting web pages of famous games where all their details are presented or even introducing projects aiming to attract potential sponsors that would bring these projects to life.

According to Statista (2019a), a company providing insights and interesting figures from various markets, the value of the global market of cards and board games was estimated to be worth approximately 12 billion USD in 2018. What's more, it is predicted to reach a value of 21.56 billion USD by 2025. Its biggest part appears to constitute the US market. For the sake of comparison, its market value was estimated to be around 2.69 billion USD in 2019 and is forecast to reach 5.04 billion USD by 2025 (Statista 2019b). The other dominant markets holding the largest share of the board game industry are for example China, France, Germany and the United Kingdom (Technavio 2022).

The popularity increase in the board game industry may also be boosted by restrictions due to the outbreak of COVID-19 in December 2019 (Statista 2020a), since people had to stay at home and thus have more time for home leisure activities such as playing tabletop games. In addition, demographic data from the survey conducted in 2020 shows that the younger generations of Americans generally like playing board games more than the older ones (Statista 2020b). For example, 82% of Gen Z respondents (born in 2000 and later) found playing these games enjoyable. On the other hand, only 67% of respondents from the Silent Generation (born between 1928 and 1945) did so. Besides, another survey was conducted asking 1,003 respondents how much money they would be willing to pay for a new game (Statista 2019c). The findings demonstrate that only 18% of the respondents consider appropriate the price above 30 USD. While price not exceeding 10 USD is found suitable by 16% of the sample.

A big part of the board game market is concentrated on second-hand markets where used games are traded. Among the most famous belong GeekMarket (used as our data source later in empirical study), Noble Knight games, Amazon and Facebook. However, one must be cautious as people may be selling games for either close to or even above the retail price trying to use other people's ignorance since most used games should be priced below the price tag of a new one.

2.3 Board game crowdfunding

Crowdfunding constitutes a very popular way how to raise the money necessary for the creation of exciting projects. Probably the most dominant crowdfunding platform is Kickstarter where over 7 billion USD was raised on more than 238,070 successful projects since its creation in 2009 (Kickstarter 2023a). The funding campaigns might be of many various categories such as design or technology. However, the main branch constitutes the games which the platform further divides into tabletop games, video games, mobile games, gaming hardware, live games, playing cards and puzzles. The majority seems to be represented by tabletop and video games. Project creators launched 76,558 game campaigns with a 47.35% success rate. The successful projects collected funds in the amount of approximately 2 billion USD out of which 32,594 raised less than 100,000 USD and 3,351 raised more (Kickstarter 2023a). One of the most successful board game projects ever is Frosthaven which gathered nearly 13 million USD from 83,193 supporters (Kickstarter 2023b).

On Kickstarter, almost anyone can launch their raising campaign there, not only individuals but also companies in order to predict product sales. This indicates relatively low barriers to entry, however, developing quality games constitute time demanding process. Generally, the aim is to attract potential supporters called backers that might pledge the money necessary for the realization of the project. The attention grabbers might be the introduction of the proposed game using eye-catching images as well as attractive rewards depending on the amount of the pledge. Such rewards typically include Kickstarter exclusive editions that can be received only by backing the game during the campaign and thus cannot be purchased in retail stores, just in second-hand markets. These usually include some sort of special content like upgraded components or extra miniatures. Besides, campaign creators must set a funding goal and a deadline that must be fulfilled for the project to become successful and thus charge the pledges from backers (Kickstarter ca. 2014).

Chapter 3

Literature Review

Board games have been studied a lot from many perspectives. For example, Mijal (2014) studied game preferences among different cultures. Besides, there are plenty of studies revealing playing board games has a positive effect on cognitive functions, education and physical activity (Noda *et al.* 2019). Nevertheless, there is a lack of literature regarding the board games price.

This chapter focuses on the summary of the related literature to our study. The first Section 3.1 focuses on Hedonic Pricing Models (HPMs). Continuing with Section 3.2 where literature pertaining to Manufacturer's Suggested Retail Price (MSRP) briefly is presented.

3.1 Hedonic Price Analysis

In the literature, we can find a lot of HPMs of various goods. However, the hedonic pricing of board games is still missing. Therefore, the pricing of similar goods will be discussed and its findings will be related to our research.

HPMs have an extensive literature background of several decades. According to Lancaster (1966), consumers' utility is derived from goods' characteristics rather than from goods directly. This insight presents the introduction of a new approach to consumer theory and thus Lancaster's paper is usually cited by studies dealing with HPMs as it provided the basis for examining such utility-generating attributes (Malpezzi *et al.* 2003). In general, hedonic pricing constitutes the valuation of goods using their features.

The extensive amount of papers dedicated to HPMs is devoted to real estate. Among the frequently used mostly significant variables with positive effects on the value of the house are size, number of rooms such as bathrooms or bedrooms and other equipment including garage, swimming pool or fireplace (Sirmans *et al.* 2005). For comparison, the common significant variable with a negative sign is constituted by age. To relate the insight, that individual components of the house were found significant predictors of housing price, to the case of pricing board games, a similar price regression using its components might be applied. Nevertheless, tabletop games in general diverse a lot in content. Therefore, the attempts to estimate the price by the weight of the box or various component categories are performed instead in the empirical part of this thesis. Besides, Sirmans *et al.* (2005) warns that HPMs for real estate should be used on the local level as identical features may affect the price differently in various geographical regions. By way of example, a garage and swimming pool affecting the value of the house depending on the climate is proposed. However, we would expect these dissimilarities to be smaller for board games.

Nevertheless, the closest study to ours was conducted by Cox (2017), identifying several relevant factors influencing used video games price on the sample of 5,078 observations. The most significant among them were found to be, for example, genres, publishers, consoles on which the game can be played or dummies if the game is a special edition, is being sold bundled with an accessory or allows for online play via the internet.

Finally, regarding the functional form of HPM, Sirmans *et al.* (2005) declares that the most common model structure constitutes a semi-log form, having the dependent variable in the natural logarithm transformation. Malpezzi *et al.* (2003) states that semi-log specification was found to have several benefits over the linear form. These are the percentage change interpretation of estimated coefficients on price as well as the mitigation of the heteroscedasticity problem, just to mention a few.

3.2 Manufacturer's Suggested Retail Price

MSRP constitutes a non-binding price recommended by the manufacturer (Lubensky 2017), though it may vary from the market price which depends on demand and supply and is usually lower because games can be discounted when the demand is low. Among suggestions for retailers, it also serves as an indicator to consumers when deciding on purchases (Lubensky 2017). Some tabletop games may have MSRP printed on the box together with basic game features such as the number of players, estimated playing time and the minimum age.

Though, the retail prices may differ among different editions of the game and be in different currencies for games released first outside the US. In addition, it is problematic to be obtained MSRP for luxury editions, for example from crowdfunding campaigns.

Various goods may lose their value differently. For example, Majid & Russell (2015) and Majid & Russell (2019) found variances in value retention of used cars or Majid & Russell (2019) showed that some more expensive products lose their value more on the sample of sold iPads from eBay.

Chapter 4

Data

This chapter is dedicated to the comprehensive description of the data used for our research. It is divided into five main sections. To start with, in the Section 4.1 the data sources are introduced. The other Section 4.2 explains the process of data collection. Furthermore, in the Section 4.3 data cleaning is performed. Moreover, Section 4.4 provides the description of variables used later in the modelling part. On top of that, in the Section 4.5 data limitations are discussed.

4.1 Data sources

Data were primarily taken from BGG (boardgamegeek.com or its shorter form bgg.com). It is a website for board game enthusiasts founded in January 2000 with a lot of amazing content such as discussion forums or a huge online database including detailed information on more than 140,000 tabletop games (as of March 2023). Wachs & Vedres (2021) liken it to IMDB (Internet Movie Database) in the film industry. The database is continuously being updated by the board gamers. However, modifications must be approved by the site administrators before it gets reflected. Mijal (2014) points out these data must be tackled carefully due to the administrators' inability to properly check all the content. The other part of BGG is the Geekmarket, a second-hand P2P (peer-to-peer) market for pre-owned games having thousands of active listings as well as an extensive history of purchased items.

To be able to buy, sell, trade games, discuss on forums as well as contribute to the database by rating games or editing game facts, one must become a BGG user by creating an account. The creation is for free though one must be at least 18 years old.

When creating a new listing on Geekmarket, the seller must provide the item's name, location, condition, price, payment method and to which location they are willing to ship. In addition, they might optionally specify the item's version and additional notes (for example specify bundles). Last, they must agree to pay a 3% fee if the item will be sold via this ad.

BGG was chosen since it appears to be the best tabletop games data source with registered users exceeding two million (BGG ca. 2014b). What's more, it also has 2 different active Application Programming Interfaces (APIs).

In addition, the majority of data on MSRP of the games was extracted from boardgameatlas.com, the nifty price comparison site. The main alternative was Noble Knight Games, the market of used games, from which MSRP if available or a price for the new game was obtained. The last options were either listed retail price on BGG or a manual search on the official pages of publishers.

4.2 Data extraction

This section continues by presenting the way of the data-obtaining process. Data were extracted, cleaned and merged in the programming language Python in Jupyter Notebooks and subsequently analysed in Python and R.

As stated in the Section 4.1, BGG has 2 APIs. The newer one was chosen for downloading the data as its scope better fits the needs of this work. It contains information on active listings from the GeekMarket, the game's characteristics included in the database as well as statistics of BGG users (BGG ca. 2015a).

The API possesses details on various items including not just tabletop games but also their accessories and video games. Therefore, only figures on board games and board game expansions were collected, the rest was skipped together with the games having no listings on Geekmarket. Requests were called on almost all items in the database using their ids from 1 up to 380,000. Few of them were considered appropriate to be disregarded as there are 381,212 available items (as of April 2023) and the most recent ones tend to be the last added. Therefore, they usually do not have the entirety of the data such as the decent number of votes and active listings on GeekMarket. Moreover, the API allows requesting multiple items with a single query. This enabled us to be downloading details on 10 various games at once. Furthermore, the raw data were scraped in the format of Extensible Markup Language (XML) and subsequently cleaned using BeautifulSoup, an XML parser. Thereupon, they were transformed into a data frame. The whole procedure of downloading takes approximately 60 hours.

For our purpose, 245,112 unique listings of 41,578 unique games were extracted in accordance with its terms of use (BGG ca. 2015b) within the period range of 18th and 20th of February 2023. These constitute listings active during the data extraction process. Thus, the final data frame consists of listings details and game characteristics. Listings details constitute the date of insertion, price, condition, currency and item notes. While game characteristics are the weight of the box, dimensions of the box, year published, popularity rating, complexity rating, number of players, playing time, minimum age, language dependence, mechanics, categories, and families. Each of them will be discussed later in the study.

In addition, ParseHub, a free web scraping tool was utilized to extract 2 other variables, the location and edition of the game because these variables are not available in any of the APIs. However, the variable on the game's edition has a lot of missing values as it is an optional detail in the listings. Both were extracted in the period of the 19th and 20th of February 2023. Afterwards, the data sets were merged and missing values dropped reducing our data to 242,627 observations from 96 different states. The Table 4.1 shows that most of them are from the US, Germany, Belgium, Great Britain and the Netherlands.

Table 4.1: Frequency by Top 5 Countries

country	frequency
United States	87648
Germany	46794
Belgium	17392
Great Britain	16654
Netherlands	15340

Source: Author's calculations

Finally, the majority of data on MSRPs was extracted from the API of boardgameatlas.com in the format of JSON (JavaScript Object Notation). Firstly, the requests were called on top-ranked games as this allows the extraction of multiple games in one request. However, the site limits users to get only 1,100 such titles. Therefore, in addition to that, other requests were called directly on names of games for which MSRPs were needed. Nevertheless, there was a problem as the name of the identical game may differ among BGG and boardgameatlas.com. Also, various games could have identical names. Hence, games not matching were manually checked and their MSRPs were filled from one of the sources mentioned in Section 4.1. Utterly, data sets were merged.

4.3 Data cleaning

This section is devoted to the general cleaning process. Further transformations of individual variables will be discussed in Section 4.4. Considering the fact that data were obtained using web scraping techniques, proper data cleaning was applied.

Initially, a filter on listings from the US was applied. There are several reasons for this filter. Firstly, the US is probably the biggest board game market plus having the most listings on GeekMarket. Secondly, this allows us to narrow games mostly to English editions which is beneficial as it facilitates estimating the publisher of the game. Also, retail prices are easier to gather as they might differ among various editions and be in different currencies across countries. Besides, HPMs are challenging to generalize across multiple different regions as factors affecting price may differ among them.

Furthermore, some listings might be a few years old (the oldest one is from 2003). Therefore, only those with dates of insertion after the 1st of January 2021 were taken into consideration. There was a trade-off between losing too many observations and considering expired listings.

Moreover, inasmuch as this paper intends to study only modern games, listings with the published years 2004 and earlier were dropped. The newer games tend to be more complex with more elaborate designs so it would be challenging to compare games from different time horizons plus the portion of older games is smaller in comparison with more recent ones. Besides, games whose published year was missing were dropped, too since these were either some sort of accessory or really unknown ones.

Furthermore, we had 8 listings with other currency than USD. These were simply dropped as the majority would be filtered out later in the cleaning process anyway. As the next step, the price boundaries were set such that all listings with prices exceeding 200 USD or being lower than 10 USD were removed. This allowed us to filter out overpriced listings as well as suspiciously cheap ones that could encompass accessories, promotion cards and mini-expansions.

In addition, a lot of effort was done to drop special editions which often have specific characteristics, e.g. painted miniatures and may not lose value when entering the used marketplace (Majid & Russell 2019). The reason why they were not studied instead is explained in Section 4.5. If the game is a special edition, it is most often specified either in the name or edition specification. Firstly, all games having one of the keywords "deluxe", "special", "limited", "luxury", "premium", "kickstarter", "collector's" or "expansion" (as there might be a game having an edition including more expansion) in edition specification were filtered out. Even though the variable game edition is often missing, sellers would rather be expected to share this information with potential buyers if their item would be a luxury edition as it would most likely increase its value in the eyes of customers. In addition, games having similar keywords in names were dropped as well. Nevertheless, some special editions may remain undetected. It would be highly time demanding or almost impossible to filter them all based on item notes. Therefore, more effort to remove them will be introduced later on.

Moreover, games are sometimes sold bundled with various items varying in value such as expansions or accessories, leading to unrealistic overpricing for the given game. If it is the case, the seller usually specifies so in the item notes. To do our best to tackle this, we observed the most used keywords occurring when this happens. Therefore, listings having a combination of words such as "expansion", "bundle", "include", "kit" and "collection" in the item descriptions were dropped. Alternatively, bundles might be studied. For instance, Cox (2017) studied its effect on the price of video games. However, they will be dropped for the sake of this work as their value could markedly differ depending on what it presents - is it only some little accessory or a couple of expansions?

As the next step, the filter on games that have at least 8 listings was applied since it allows the restriction of the most frequently traded games as well as dropping unpopular ones. Next, advertisements having pre-owned prices higher than their MSRP were dropped. This enables us to dismiss other overpriced listings that would otherwise skew our results as well as the rest of the unidentified special editions and bundles. Finally, the median price was used for advertisements that differ in no details, just in price. This might happen due to the fact that many identical games are typically advertised more than once for various prices. Kalita *et al.* (2004) states that 2 brands with identical combinations of product attributes/features must charge identical prices in a perfectly competitive market. Thus, our data set was reduced to 2,184 unique games after this data-cleaning process.

4.4 Variables description

The following section is dedicated to shedding light on the collected features of the games. Note that all variables were checked for abnormal values and their descriptive statistics can be seen in Table 4.3. Besides, the description of qualitative variables contains Table 4.5 and Table 4.6. Moreover, quantitative variables are described in Table 4.4. Some of the descriptions were used from the BGG.

4.4.1 Dependent variables

This thesis aims to model two dependent variables. The first one constitutes the natural logarithm of the price of the game being listed on the second-hand market. The other stands for the residual price/value share of the board game - pre-owned price divided by the MSRP, thus the proportion of the original value that remains. It was additionally multiplied by 100 to be in per cent. In the Figure A.1, one can see the distribution of the logarithm of price. Before the utilization of the logarithmic transformation, the average price was equal to 33.48. After the transformation, one can notice that the histogram is slightly skewed to the right. Moreover, a histogram of price proportion may be found in Figure A.2. This one appears to be close to the normal distribution.

4.4.2 Condition

Condition is the primary factor of our interest that informs consumers about the depreciation of the game. Worse-state games are expected to be cheaper and retain less value. Below, are stated the exact descriptions of possible conditions that occur in the marketplace (BGG ca. 2015c).

- “New - A brand new unused, unopened, and undamaged game in perfect condition. The original packaging and all materials are in brand new condition.”
- “Like New - Game just removed from shrink wrap. No wear and tear, all facets of the game are intact.”

- "Very Good - Very minimal wear and tear. All game materials are present. You would give this item to a friend as a gift."
- "Good - Minor damage to the box and/or its contents. All game materials are present. Game maybe played once or twice."
- "Acceptable - Some damage to the box, but the game is still intact. Possible split corner(s) on the box. Maybe missing a non-crucial game piece. Possibly missing rules/instructions, but are available on the web. Scuffing on the game board."
- "Unacceptable - Major damage to box and its parts. Possibly missing several important pieces. Broken or missing board/box. No rules/instructions, and they are not readily available."

Unacceptable games are prohibited from being offered on the market. In the Table 4.2, the frequency of particular conditions is exhibited. Regarding the figures, acceptable might be grouped with good as it represents only 2.5 % of the sample. However, it was decided not to, as the condition is the key factor of our interest.

Table 4.2: Frequency by Conditions

condition	frequency
new	658
like new	616
very good	605
good	242
acceptable	54

Source: Author's calculations

4.4.3 Weight and volume of the box

The variable of weight is rather an estimate as the weight and dimensions (length, width, depth) of the box may vary among different versions of the game. The differences are usually not too big, however, there might be a tendency of higher weight for special editions. As we do not have the entirety of data on game editions, we calculate the median of available values of weight and sizes of the box of all possible versions. Thereupon, the volume of the box is calculated by the multiplication of its dimensions. This action reveals that volume has 37 missing values and its units are cubic inches. On the contrary, weight is measured in pounds and has 226 missing values which represents approximately 10.4% of the sample. Games above 16 pounds and 1,650 cubic inches are considered outliers and dropped. To fill in missing values for weight, the variable volume was utilized as these two are highly correlated (0.75). Therefore, volume was cut into three bins small, medium and large which correspond to below 200, in between 200 and 500 and above 500, respectively. Then it was filled by the median values of the weight in these groups. Finally, the rest was filled by the median of the weight.

Looking at Table 4.3, the average weight is approximately 3.6 pounds which correspond to about 1.65 kilograms. Besides, weight is expected to have a positive effect on the price as heavier games are more likely to have more content with more components. The more pieces the game contains, the more it costs to produce. The scatter plot is displayed in the Figure A.3.

4.4.4 Number of players

The publisher typically provides the minimum and maximum number of players for which the game is intended to be played in accordance with the game rules. Three dummy variables were created for the purpose of our models. The first one demonstrates whether the game can be played by more than three players, the second informs if the game is intended just for two players and the last one stands for the possibility of a solo game. However, the second mentioned was not used as it is highly correlated with the first one. Moreover, variable identifying games that are able to be played by just one player was not used due to a small portion of such games.

4.4.5 Minimum age

This variable provides information on the publisher's recommended minimum age for the game. Table A.1 presents the frequencies of the variable. Regarding the figures, just one dummy variable was decided to be created: 1 if the recommended minimum age is at least 12 years; 0 otherwise. Besides, 25 missing values were put into the category above 12 as this group has more observations and thus could be considered a mode.

4.4.6 Time duration

This variable constitutes the approximate time duration in minutes for gameplay provided by the publisher. It might be either in the form of an estimate or a range of minimum and maximum playing time. Therefore, either the estimation if available or the average of minimum and maximum duration was taken into consideration. Afterwards, games that take longer than 5 hours were replaced with 5 hours. Games with higher playing time are expected to have a higher price as these may include more content than shorter ones. In the Table 4.3, one can note standard deviation, minimum and maximum values being equal to around 56, 10 and 300, respectively, indicating the presence of a diversity of the time requirement of the game in our data set.

4.4.7 Published year

This variable was transformed as the number of years since the base year that was chosen 2023 as the year of data collection. Therefore, it demonstrates the age of the games. It could be interesting to see whether older or newer games retain value more when being traded in bazaars. The published year may differ between distinct versions of the game. However, we are using the published year of the first version. As games older than 2004 were already filtered out, the average is approximately 8 years. Table A.2 demonstrates the frequency.

For the purpose of the residual price share modelling, several dummies were created and included in the model as the interpretation seems to be more intuitive in that form. In the model explaining price, the variable was used in the quantitative form.

4.4.8 Rating

This variable was decided to include as Cox (2017) demonstrates that it constitutes a significant determinant of the pre-owned price of video game software as consumers possess this information prior to purchase. Two different metrics might be used for this variable. Either the average of votes from BGG users using a 10-point scale or the BGG rating which also considers the number of votes aiming to prevent games with relatively few votes from getting among top-rated ones (BGG ca 2014d). A lower rating means worse. Mijal (2014) maintains more complex games with a lower level of randomness are favoured since BGG users prefer them. However, BGG as well as the used markets are primarily used by board game lovers. Besides, the correlation between rating and complexity is around 0.19. Therefore, it might even better fit the reality. For the sake of this study, BGG rating was decided to use.

Every game included in our data set was rated by at least 54 BGG users. The median number of votes is 4,791 with a standard deviation of around 15,864. In addition, the reader can register that the minimum value of the rating is equal to 5.52 after the filter on frequent-traded games was applied.

4.4.9 Complexity rating

That variable demonstrates the complicatedness of the game typically influenced by the complexity of rules, the level of analytical skills required to optimize the best game strategy, the portion of luck included in the game and the number of games played before one gets fully familiarised with the game (BGG ca 2014c). It is also measured by BGG users' ratings, this time using a 5-point weight scale where 1 means light and 5 means heavy. The minimum and maximum number of votes on the complexity of games in our data set are 3 and 6,220, respectively, with the median being equal to 203 votes.

We believe the higher the complexity is, the higher the price since complex games are likely to be more difficult to produce and also might need more components. One might also note the relatively high correlation between the duration of play and the complexity of the game (0.58).

4.4.10 Categorizations

There are 84 categories, 191 mechanics and 2454 possible families on BGG (as of March 2023) (BGG ca 2014e). One game typically has numerous features from these.

The mechanics or mechanisms basically describe the functionality of the game. Mijal (2014) declares new mechanics are gradually being added as for new games more classification might be needed. While families include very detailed information on games. One can find details such as what component categories are included in the game - for example, if the game contains miniatures or wooden components, or if the game is related to the given country, city, film, book, comics and many more.

Given the huge number of potential independent variables in our models, only those considered significant for our study and having at least 5% of zeroes or nulls were explored deeper. The selected variables belonging to categories constitute card, war, dice, miniatures, animals, economic, horror, fighting and adventure. Apart from that, modular_board, cooperation, storytelling, scenarios, team and luck are mechanics. All of them were encoded as dummies afterwards and their description can be seen in the Table 4.6. Lastly, crowdfunding_KS and dice_with_icons are classified as families on BGG and their description is provided in the Table 4.5 together with the rest of the qualitative variables.

4.4.11 Publishers

Different editions may have different publishers plus one edition might have multiple publishers. We do not have the exact data on whom published which listed games, just the list of all publishers that published any game's editions. Nevertheless, we believe it might have an effect on the price. Cox (2017) finds out that publishers' versioning may influence the value of pre-owned video games. We could argue that by applying a filter on games located in the US and taking the median price for games that may differ in the version, English editions prevail. Therefore, we are using 3 relatively big publishers that predominantly publish English editions in our price model. These are Fantasy Flight Games, Rio Grande Games and Z-Man Games.

4.4.12 Variables not used

To begin with, variable language dependence, demonstrating whether the game can be easily played without knowing the language of the game, was omitted because some games have just a few votes from BGG users. For instance, 227 observations have less than 3 votes. Besides, the variable that says if the game is an expansion was not used because we had only 1.8% of such games.

4.5 Data limitations

Regrettably, there are plenty of data limitations which would boost our study. Apart from the ones already mentioned previously in this Chapter, the other significant ones will be discussed. One such example is that using historical purchases instead of active listings would be more realistic to consider. Unfortunately, the history of sold listings is not available in the API. Though, it is unrestricted on the Geekmarket so it ought to be feasible to get this data using web scraping. Nevertheless, this procedure would be extremely time demanding.

Secondly, the data on exact game components is not at our disposal, just general categorical classification (e.g. miniatures, wooden components). It is partly due to the fact that board games often seriously differ in content. If we had this data, we could apply filters on games with similar content and apply multiple regressions for instance.

In addition, there are a lot of missing values on game editions. If we had this data we could try to identify special editions and study their effect on price. However, there would probably be huge price differences among them as they could be considered special for many reasons. What's more, there would likely be a very low portion of them like in a study conducted by Cox (2017) showing it has a hugely positive effect on the video games' price.

Last, if we had details on the game publisher of the listed game, studying price versioning among different publishers might be easier. For the analysis of this research, rather estimates of which listing has which publisher is used as one game version might have multiple publishers and thus the most famous are studied.

Table 4.3: Descriptive Statistics of Variables

variable name	mean	std	min	max
complexity	2.764142	0.728718	1.028600	4.704300
rating	6.748231	0.620496	5.526160	8.410230
price	33.475591	19.704341	10.000000	189.000000
residual price share	55.790155	16.838346	15.001500	99.980000
log(price)	3.366821	0.530805	2.302585	5.241747
weight	3.640999	1.676222	0.480610	12.850000
volume	355.221928	168.692447	18.526821	1415.680386
players_above_3	0.902989	0.296041	0.000000	1.000000
min_age_above_12	0.697471	0.459459	0.000000	1.000000
duration	84.632414	56.351617	10.000000	300.000000
age	7.871724	4.306731	1.000000	18.000000
age_1or2	0.084598	0.278346	0.000000	1.000000
age_3	0.081839	0.274182	0.000000	1.000000
age_4	0.091954	0.289027	0.000000	1.000000
age_5	0.078621	0.269208	0.000000	1.000000
card	0.195402	0.396601	0.000000	1.000000
war	0.106207	0.308173	0.000000	1.000000
dice	0.080000	0.271356	0.000000	1.000000
miniatures	0.096552	0.295414	0.000000	1.000000
animals	0.070345	0.255786	0.000000	1.000000
economic	0.199540	0.399747	0.000000	1.000000
horror	0.061609	0.240500	0.000000	1.000000
fighting	0.156782	0.363678	0.000000	1.000000
adventure	0.131494	0.338018	0.000000	1.000000
players_solo	0.235862	0.424634	0.000000	1.000000
modular_board	0.183448	0.387123	0.000000	1.000000
cooperation	0.199540	0.399747	0.000000	1.000000
storytelling	0.056552	0.231037	0.000000	1.000000
scenarios	0.090575	0.287069	0.000000	1.000000
team	0.065287	0.247089	0.000000	1.000000
luck	0.061149	0.239659	0.000000	1.000000
crowdfunding_KS	0.278621	0.448424	0.000000	1.000000
players_2	0.055632	0.229263	0.000000	1.000000
dice_with_icons	0.065287	0.247089	0.000000	1.000000
fantasy_flight	0.064368	0.245463	0.000000	1.000000
z_man	0.063448	0.243824	0.000000	1.000000
rio_grande	0.038621	0.192734	0.000000	1.000000
new	0.278161	0.448196	0.000000	1.000000
likenew	0.302529	0.459459	0.000000	1.000000
verygood	0.283218	0.450665	0.000000	1.000000
good	0.111264	0.314532	0.000000	1.000000
acceptable	0.024828	0.155635	0.000000	1.000000

Source: Author's calculations

Table 4.4: Description of Quantitative Variables

variable name	variable description
log price	natural logarithm of the price of the game in USD
price difference	the proportion between pre-owned price and MSRP
complexity	rating of the complexity of the game on a scale of 1 to 5
rating	BGG rating of the game on a scale of 1 to 10 also considering the number of votes
duration	time duration the game takes to play in minutes
weight	estimate of the weight of the box of the game in pounds
age	age of the game in years at the year of data collection

Source: BGG

Table 4.5: Description of Qualitative Variables/Dummies

variable name	variable description
new	1 if the condition of the game is new; 0 otherwise
likenew	1 if the condition of the game is likenew; 0 otherwise
verygood	1 if the condition of the game is verygood; 0 otherwise
good	1 if the condition of the game is good; 0 otherwise
players_solo	1 if "the game can be played by a single player against a Bot-player"; 0 otherwise
players_above_3	1 if the game can be played by more than three players; 0 otherwise
min_age_above_12	1 if the game is intended for people older than 12 years; 0 otherwise
age_1or2	1 if the age of the game is one or two years; 0 otherwise
age_3	1 if the age of the game is 3 years; 0 otherwise
age_4	1 if the age of the game is 4 years; 0 otherwise
age_5	1 if the age of the game is 5 years; 0 otherwise
crowdfunding_KS	1 if "the game was launched on the Kickstarter website"; 0 otherwise
dice_with_icons	1 if "the game contains dice which don't have numbers on them, but icons to represent actions or resources that you can take"; 0 otherwise

Source: BGG

Table 4.6: Description of Qualitative Variables/Dummies - Categories and Mechanics

variable name	**variable description**
card	1 if "the game uses cards as its sole or central component"; 0 otherwise
war	1 if "the game depicts military actions"; 0 otherwise
dice	1 if "the game often uses dice as its sole or principal component"; 0 otherwise
miniatures	1 if "miniatures are the key components in the game and are used to stage the game scenes"; 0 otherwise
animals	1 if "the game involves animals as a major component of the theme or gameplay"; 0 otherwise
economic	1 if "the game encourages players to manage a system of production, distribution, trade, and/or consumption of goods"; 0 otherwise
horror	1 if "the game often contains themes and imagery depicting morbid and supernatural elements"; 0 otherwise
fighting	1 if "the game encourages players to engage game characters in close quarter battles and hand-to-hand combat"; 0 otherwise
adventure	1 if "the game often has themes of heroism, exploration, and puzzle-solving" ; 0 otherwise
modular_board	1 if "the play occurs upon a modular board that is composed of multiple pieces, often tiles or cards" ; 0 otherwise
cooperation	1 if "players coordinate their actions to achieve a common win condition or conditions and they all win or lose the game together." ; 0 otherwise
storytelling	1 if "players are provided with conceptual, written, or pictorial stimuli which must be incorporated into a story of the players' creation" ; 0 otherwise
scenarios	1 if "the game has a system that can be applied to a variety of different maps, starting resources and positions, and even different win and loss conditions" ; 0 otherwise
team	1 if "in the game, teams of players compete with one another to obtain victory" ; 0 otherwise
luck	1 if "layers must decide between settling for existing gains, or risking them all for further rewards, in a game with some amount output randomness or luck" ; 0 otherwise

Source: BGG

Chapter 5

Methodology

This chapter introduces the methodology that will be used for the analysis. In its only Section 5.1, the model is presented.

5.1 Model

MLR with Ordinary Least Squares (OLS) were chosen for the analysis because of the nature of the collected data. The advantages of such a method are easy interpretation, popularity and simplicity. On the contrary, its results are unbiased if and only if all of the MLR assumptions are fulfilled (Wooldridge 2012).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i, \text{ i} = 1, \ldots, \text{n} \tag{5.1}$$

where:

- y is the dependent variable
- $x_1, \ldots, x_k$ are k independent variables
- β_0 is the intercept
- $\beta_1, \ldots, \beta_k$ are parameters associated with $x_1, \ldots, x_k$
- u_i is the error term

In our first model, y_i represents the natural logarithm transformation of the price of the game. Thus, a percentage change interpretation is feasible which fits our case better. In the other model, y_i stands for the pre-owned price divided by the MSRP, additionally multiplied by 100, so the variable is in per cents and the interpretation of the effect of independent variables will be in percentage points. Besides, $x_1, \ldots, x_k$ constitute our explanatory variables described in the Section 4.4. Both explained variables were regressed using 32 predictors. However, these predictors differed as the first model additionally controlled for publishers and for the sake of the second model, the dummies created from the age variable were utilized. Finally, n represents the number of observations.

Chapter 6

Discussion Results

This chapter comments on the results of our analysis. Two dependent variables, the logarithm of the price and the residual price share (the proportion of the suggested retail price that remains), were modelled by numerous independent variables using R studio. The results of regressions of both explained variables are presented and interpreted in Section 6.1 and Section 6.2, respectively. Continuing with Section 6.3, where the robustness check is performed. Finally, in the Section 6.4, the MLR assumptions are discussed.

6.1 Price Modelling

The regression results of the two different models are shown in the Table 6.1. Compared to Specification (1), Specification (2) includes extra the interaction term among the age of the game and rating, similarly as in the study conducted by (Cox 2017) where it was found statistically significant. Both models are log-level, meaning that the interpretation of parameter k will follow the given form: $\%\Delta y = (100\beta_k)\Delta x$ (Wooldridge 2012).

To start with, let's declare that this work measures model performance using adjusted R-squared since this metric also considers the number of independent variables included in the model in comparison with R-squared (Wooldridge 2012). Both models have adjusted R-squared above 0.54 which is close to the value of a study conducted by (Cox 2017). The interaction term was found to be positive and statistically significant even at a 99 % significance level, demonstrating the positive effect of rating on the price of the title is higher for older games. At the same time, the negative effect of the age of the game on price is diminished for better-rated games. In Specification (2), the inclusion of

Table 6.1: Regression output - Price

	Dependent variable:	
	log(price)	
	(1)	(2)
new	0.423*** (0.066)	0.412*** (0.064)
likenew	0.279*** (0.065)	0.267*** (0.064)
verygood	0.199*** (0.065)	0.186*** (0.064)
good	0.085 (0.069)	0.075 (0.067)
weight	0.099*** (0.007)	0.100*** (0.007)
age	−0.112*** (0.021)	−0.005** (0.002)
rating	0.047* (0.026)	0.162*** (0.015)
complexity	0.091*** (0.018)	0.102*** (0.018)
duration	0.002*** (0.0003)	0.002*** (0.0003)
players_above_3	0.005 (0.032)	0.001 (0.032)
players_solo	0.051*** (0.020)	0.044** (0.020)
min_age_above_12	0.036* (0.020)	0.035* (0.020)
card	−0.055** (0.022)	−0.054** (0.022)
war	0.101*** (0.038)	0.109*** (0.038)
dice	0.069*** (0.026)	0.072*** (0.026)
miniatures	0.203*** (0.036)	0.204*** (0.036)
animals	0.050* (0.028)	0.049* (0.028)
economic	0.015 (0.021)	0.020 (0.021)
horror	0.050 (0.033)	0.049 (0.033)
fighting	0.038 (0.027)	0.033 (0.028)
adventure	−0.011 (0.030)	−0.007 (0.030)
modular_board	0.010 (0.024)	0.010 (0.024)
cooperation	−0.082*** (0.026)	−0.075*** (0.026)
storytelling	−0.040 (0.037)	−0.041 (0.038)
scenarios	0.073** (0.032)	0.069** (0.032)
team	−0.001 (0.037)	0.004 (0.037)
luck	0.105*** (0.032)	0.106*** (0.032)
crowdfunding_KS	0.187*** (0.019)	0.184*** (0.019)
dice_with_icons	−0.020 (0.028)	−0.028 (0.028)
z_man	0.070** (0.033)	0.087*** (0.033)
fantasy_flight	0.034 (0.044)	0.041 (0.044)
rio_grande	0.034 (0.037)	0.038 (0.038)
I(age ∗rating)	0.016*** (0.003)	
Constant	1.893*** (0.186)	1.125*** (0.114)
Observations	2,175	2,175
R^2	0.548	0.542
Adjusted R^2	0.541	0.536
Residual Std. Error	0.360 (df = 2141)	0.362 (df = 2142)
F Statistic	78.673*** (df = 33; 2141)	79.343*** (df = 32; 2142)
Note:	*p<0.1; **p<0.05; ***p<0.01	
Note:	Robust standard errors in parentheses	

the interaction term alters the estimated coefficients (do not change signs) and thus interpretation of age and rating while estimates of non-interacted variables remain mostly consistent with values from Specification (1). Therefore, we will continue with the interpretation of estimates from Specification (2).

The signs of the majority of estimated coefficients are consistent with the theoretical expectations discussed in Chapter 4. Getting back to age and rating, both show significance at a 5% significance level. For the US market of pre-owned games, if the game is older by one year, the price is estimated to be lower by 0.5 % and games having a rating higher by one point are estimated to have higher prices by 16.2% on average, ceteris paribus.

To continue with quantitative variables, weight, complexity rating and duration were found to be statistically significant at a 99% significance level. Their impact on price resulted in an increase by 10%, 10% and 0.2% on average, c.p. if these variables increase by one pound, point and minute, respectively. All these estimates were anticipated as they are believed to have a tendency to include more components in the box which drags the price up. Also, the invention of such games may be more time-demanding.

As expected, the condition turned out to be one of the most significant predictors. Besides good, all condition variables turned out to be statistically significant even at a 99% significance level. The results demonstrate that the better the condition is, the higher value the game is. The condition of being new, like new and very good is associated respectively with a 41.2%, 26.7% and 18.6% increase on average, controlling for other attributes, in comparison to games in acceptable conditions.

Moreover, characteristics players_above_3 and min_age_above_12 were not found to be statistically significant. On the other hand, a variable indicating whether the game can be played solo was displayed to be statistically significant at a 95% significance level, increasing the game's value by 4.4% on average, c. p. This might be due to the fact of the need for additional costs to enable mode for one player.

Regarding game categorizations, only some games categories were identified to be statistically significant with p-values lower than 0.05, including card, war, dice, miniatures, cooperation, scenarios and luck. To begin with their interpretation, the fact that pre-owned games belong to the wargame category or include miniatures as the key components are estimated to be associated with an increase in value by around 11% and 20%, respectively, on average, c.p. This finding could indicate that miniatures might be a valuable game component. Slightly smaller effects on the price of pre-owned titles were shown by games using cards (-0.054) and dice (0.072) as their sole or central component. Apart from that, the coefficients of luck, scenarios and cooperation were estimated as around 0.1, 0.7 and -0.075, respectively. Utterly, the coefficient indicating whether games including dice with icons instead of numbers possess higher values was not detected to be statistically significant at any reasonable significance level.

Furthermore, out of the publishers, only Z-Man Games were found statistically significant with the estimated coefficient being equal to 0.087. Though, rather a negative effect would be expected as large publishers may benefit from economies of scale and thus set the price of new game lower. Controlling for this in our models was inspired by Cox (2017) who compared to our study found evidence of versioning in publishers that may influence the market value of used video games.

Utterly, games being launched on the Kickstarter website were found to have higher prices by 18.4% on average, c.p. A possible explanation could be that such games are published in smaller numbers and thus for higher prices as their publishers do not benefit from the economies of scale and need to cover the costs. However, it could also serve as an additional mitigator for unfiltered bundles and special editions as these games might have a higher chance of belonging to such categories. Therefore, one shall be careful with the interpretation of this variable.

6.2 Residual Price Share Modelling

In Table 6.2, the regression outputs are presented. By looking at the adjusted R-squared being equal to 0.245, one can notice that it is smaller compared to previous models explaining the price. Besides the expected importance of the condition, rating, some other game features and age were demonstrated to be significant.

For the US market of pre-owned games, predictably, residual price share is positively affected by the condition. All four conditions, new, like new, very good and good, were revealed to be statistically significant at a 95% significance level, corresponding to an increase of residual value share by approximately 22, 14.3, 10.7 and 6.1 percentage points on average, ceteris paribus, in comparison to games in acceptable conditions.

Regarding age, only games being old one or two years appeared to be statistically significant at a 95% significance level. This might be interpreted as, keeping other factors equal, the nature of the game being one or two years old increases the value retention on average by 7.2 percentage points in contrast to games older than 5 years. This finding seems reasonable as newer games are less likely to be played several times and thus more demanded.

In addition, rating and duration both have a positive significant effect. A one-point higher rating increases the residual price share by 8 percentage points and one additional minute of time duration increases it by 0.062 percentage points on average, c.p. This implies that well-rated and time-demanding games hold their value well. On the contrary, variables players_above_3 and complexity significantly negatively impact the residual value share. The nature of the game being able to be played with more than three players and one additional complexity rating point decreases the residual price share by 4.6 and 1.5 percentage points on average, c.p.

Besides, some categorization factors were found statistically significant at a 95% significance level. These include dice, adventure, storytelling, luck and dice_with_icons having estimated coefficients around 2.8, 2.5, -3.5, 4.5 and -2.9 respectively. Furthermore, if the game was launched on the Kickstarter platform, the value retention is estimated to be higher by 6.4 percentage points on average, c.p. This could be caused by the fact games initially launched on Kickstarter could have fewer printouts and thus are kind of more precious. However, it could also stand for the unidentified bundles and special editions as already discussed in Section 6.2.

6.3 Robustness Check

Robustness checks were performed to check how reliable the results are. Therefore, both dependent variables were modelled once again using only variables that were shown to be statistically significant at least at a 95% significance level, thus having a p-value lower than 0.05. The regression outputs can be seen in Table C.1 and Table C.2. In both regressions, by looking at the adjusted R-squared, we can say that the models' performance remained more or less the same.

Regarding outputs of the regression explaining price (Table C.1), the effects of conditions diminished. That's most likely due to the fact that condition good was omitted as it had a p-value higher than 0.05 and thus represents the base group together with the games with acceptable states. The majority of other estimated coefficients remained mostly comparable.

Moreover, by checking the results of the regression regarding the second dependent variable, one can note the diminished effect of variable age_1or2. That is probably caused by the omission of other insignificant dummies for ages. Besides, the negative effects of variables complexity, players_above_3 and storytelling increased. The rest estimated coefficients remained mostly similar.

6.4 MLR assumptions

The fulfilment of the MLR.1 (Linear in parameters) follows from the equation 5.1. MLR.2 (Random Sampling) is complicated not to violate at all. We chose one of the biggest board game platforms from which we downloaded all active listings, therefore we believe it might be a good representation. In addition, filters were applied to games intended to study.

MLR.3 (No Perfect Collinearity) requires none of the independent variables to be a constant and no existence of perfect linear relationships among the independent variables. Therefore, we did not use dummy variables having a too-small portion of zeroes or ones. In addition, a correlation matrix was depicted to identify highly correlated variables. If two variables had a correlation above 0.65 or below -0.65, only the more relevant one for our study was left in the model, the other was omitted. Moreover, to prevent a dummy variable trap, and thus the violation of MLR.3 assumption, some variables were omitted and set as a base group.

Let's continue with the assumption MLR.4 (Zero Conditional Mean). We believe this assumption holds as the intercept is included in the models. Nevertheless, to give an example of how we may be violating this assumption and thus have biased estimates is omitting a significant variable. To verify this assumption, scatter plots between residuals and fitted values of Table 6.1 and Table 6.2 were plotted (Figure B.1, Figure B.2). By looking at them, we may assume that MLR.4 holds as the residuals are mostly concentrated around zero. If all first four MLR assumptions were satisfied, OLS would be unbiased (Wooldridge 2012).

In addition, as this paper works with a cross-sectional data set, heteroskedasticity tests were performed to verify MLR.5 (Homoskedasticity). Both Breusch-Pagan and Lazy-white tests were applied. This was done since the Breusch-

Pagan test covers only linear dependence (Wooldridge 2012). The null hypothesis of the presence of homoskedasticity was rejected at a 99% significance level using the Breusch-Pagan test and at a 90% significance level using the Lazy-white test in both Table 6.1 and Table 6.2 despite the logarithmic transformation. Thus, robust standard errors were utilized as we may argue having a decent number of observations and therefore robust standard errors ought to be asymptotically valid for any form of heteroskedasticity (Wooldridge 2012).

Table 6.2: Regression output - Residual Price Sahre

	Dependent variable:
	price in the second-hand market/ MSRP (in per cents)
new	22.061*** (2.530)
likenew	14.382*** (2.516)
verygood	10.675*** (2.526)
good	6.119** (2.702)
weight	−0.335 (0.246)
age_1or2	7.227*** (1.228)
age_3	2.096* (1.173)
age_4	−0.110 (0.968)
age_5	0.662 (1.189)
rating	8.029*** (0.582)
complexity	−1.507** (0.714)
duration	0.062*** (0.011)
players_above_3	−4.581*** (1.303)
players_solo	0.754 (0.806)
min_age_above_12	−0.453 (0.811)
card	0.921 (0.908)
war	1.264 (1.522)
dice	2.779** (1.220)
miniatures	0.590 (1.413)
animals	2.178* (1.149)
economic	−0.596 (0.870)
horror	1.217 (1.468)
fighting	1.018 (1.088)
adventure	2.474** (1.255)
modular_board	0.044 (0.954)
cooperation	−1.496 (1.045)
storytelling	−3.461** (1.464)
scenarios	0.626 (1.259)
team	0.365 (1.361)
luck	4.555*** (1.347)
crowdfunding_KS	6.381*** (0.744)
dice_with_icons	−2.883** (1.304)
Constant	−11.652** (4.599)
Observations	2,175
R^2	0.257
Adjusted R^2	0.245
Residual Std. Error	14.626 (df = 2142)
F Statistic	23.105*** (df = 32; 2142)
Note:	*p<0.1; **p<0.05; ***p<0.01
Note:	Robust standard errors in parentheses

Chapter 7

Conclusion

Given the increasing board game market value, the importance of price analysis rises. This thesis aims to study the determinants of price and value retention of board games in the second-hand market in the US market. To the best of the author's knowledge, this study constitutes the first attempt to build a HPM on board games, thus could be considered a novel. Therefore, the main motivation for this thesis was to fill the literature gap and find out consumer price preferences.

Data used for the empirical evidence comes particularly from BGG, probably the biggest site for board games on the planet. Moreover, suggested retail prices of games were obtained from boardgameatlas.com. The data collection together with the data cleaning and manipulation processes constitute a challenging part of this research as data were acquired by the author using web scraping techniques, as no data set is publicly available. Data were collected in February 2023. As part of the data-cleaning process, several filters were applied to drop unwanted observations and prioritize the most frequently traded games. Thus, the final data set consists of a decent number of listings (above 2,000) enriched by detailed game characteristics.

Due to the lack of literature on pricing board games, another important contribution of this thesis might be seen in the applied feature selection of independent variables used later in the regressions. The selected ones constitute various game features including general ones namely rating, time duration or the weight of the box as well as the more specific such as dummy if the solo mode is available.

To study determinants' magnitudes and signs, applying MLR and OLS, two dependent variables, price and the portion of the original MSRP that remains, were modelled. Also, the verification of MLR assumptions was discussed. In addition, all models were subject to the robustness checks by running regressions again, this time only using significant variables. In general, by mentioned robustness check, estimates turned out to be consistent with previous results.

Regarding the results of price regression, variables that were discovered to have the most significant effects are condition, weight, rating, complexity and duration. As expected, all of them had positive estimated coefficients. One of the interesting insights to mention is that an additional increase of game weight by one pound is estimated to increase the price by 10% on average, keeping other factors fixed.

On the other hand, the regression of the residual price share reveals significant factors such as condition, age and rating. Continuing with the interpretation of more interesting insights, game collectors can retain the value of the games primarily by keeping them in perfect condition, ideally not unwrapping them at all. Our regression results indicate that on average, ceteris paribus, the value retention of a new game is higher by 22 percentage points compared with a game in an acceptable condition which corresponds to the worst state in which games can be traded on the second-hand market used for our analysis. Besides, keeping old standard titles is shown to be a bad strategy as games having an age of one or two years are estimated to retain value more by approximately 7.2 percentage points in comparison with games being older than 5 years on average, ceteris paribus.

The main findings of this work constitute the identification of some significant variables that aim to mitigate consumers' information asymmetry during the decision-making process of purchasing board games. Last but not least, the findings could also serve publishers and board game designers by shedding light on the variation in consumers' willingness to pay for the varied offer of titles (Cox 2017), even though listings instead of the actual historical sales are used for the analysis. Besides, in general, the conclusions may be beneficial to everyone seeking to assess a board game's market value.

Furthermore, the most limitation of this study is caused by the imperfection of the data. To mention a few, unfortunately, the author does not have access to historical sales which might help to make the research more realistic. Besides, more exact details for components of the games could definitely boost our analysis.

The discoveries of this research pawned the way for further analyses. Ideally, this paper could motivate other researchers to continue broadening this topic since more evidence on new data sets would definitely be useful. For example, obtaining data from one of the other second-hand markets or analyzing the factors affecting board game price and value retention in other regions, especially in European markets might bring interesting findings. Furthermore, uncovering other value drivers, such as expansion, special edition, if the game is bundled and from which publisher the game comes, would definitely be useful. Finally, factors influencing board game seekers' willingness to pay could also be examined.

Bibliography

Averbakh, Y. (2012): *A history of chess: from Chaturanga to the present day*. SCB Distributors.

BGG (ca. 2014a): "Expansion for base-game." https://boardgamegeek.com/boardgamecategory/1042/expansion-base-game. Accessed 2023-04-17.

BGG (ca. 2014b): "Welcome to boardgamegeek." https://boardgamegeek.com/wiki/page/Welcome_to_BoardGameGeek. Accessed 2023-03-13.

BGG (ca 2014c): "Weight." https://boardgamegeek.com/wiki/page/Weight. Accessed 2023-03-27.

BGG (ca 2014d): "Ratings." https://boardgamegeek.com/wiki/page/Ratings. Accessed 2023-04-13.

BGG (ca 2014e): "Board game categories." https://boardgamegeek.com/browse/boardgamecategory. Accessed 2023-04-13.

BGG (ca. 2015a): "Bgg xml api2." https://boardgamegeek.com/wiki/page/BGG_XML_API2. Accessed 2023-03-13.

BGG (ca. 2015b): "Xml api terms of use." https://boardgamegeek.com/wiki/page/XML_API_Terms_of_Use. Accessed 2023-03-13.

BGG (ca. 2015c): "Condition of a game." https://boardgamegeek.com/wiki/page/Condition_of_a_game. Accessed 2023-03-13.

Cox, J. (2017): "Play it again, sam? versioning in the market for second-hand video game software." *Managerial and Decision Economics* **38(4)**: pp. 526–533.

FIDE (ca. 2008): "About fide." https://www.fide.com/fide/about-fide. Accessed 2023-03-19.

GLASSDOOR (2022): "Board game designer salaries." https://www.glassdoor.com/Salaries/board-game-designer-salary-SRCH_KO0,19.htm. Accessed 2023-04-13.

GUINNESSWORLDRECORDS (ca. 2021): "Most popular board game." https://www.guinnessworldrecords.com/world-records/72473-most-popular-board-game. Accessed 2023-03-19.

KALITA, J. K., S. JAGPAL, & D. R. LEHMANN (2004): "Do high prices signal high quality? a theoretical model and empirical results." *Journal of Product & Brand Management* **13(4)**: pp. 279–288.

KARWATKA, D. (2016): "Elizabeth magie and the game of monopoly." *Tech Directions* **76(3)**: p. 10.

KICKSTARTER (2023a): "Stats." https://www.kickstarter.com/help/stats. Accessed 2022-04-24.

KICKSTARTER (2023b): "Frosthaven." https://www.kickstarter.com/projects/frosthaven/frosthaven. Accessed 2022-04-15.

KICKSTARTER (ca. 2014): "Kickstarter and taxes." https://www.kickstarter.com/help/taxes. Accessed 2022-04-24.

LANCASTER, K. J. (1966): "A new approach to consumer theory." *Journal of political economy* **74(2)**: pp. 132–157.

LUBENSKY, D. (2017): "A model of recommended retail prices." *The RAND Journal of Economics* **48(2)**: pp. 358–386.

MAJID, K. A. & C. A. RUSSELL (2015): "Giving green a second thought: Modeling the value retention of green products in the secondary market." *Journal of Business Research* **68(5)**: pp. 994–1002.

MAJID, K. A. & C. A. RUSSELL (2019): "Value dynamics in the secondary market: How pricing and product lines in the primary market affect value retention." *Journal of Business Research* **103**: pp. 89–99.

MALPEZZI, S. *et al.* (2003): "Hedonic pricing models: a selective and applied review." *Housing economics and public policy* **1**: pp. 67–89.

MIJAL, M. (2014): "Cultural determinants of board games preferences." *Homo Ludens* **1**: pp. 117–129.

Noda, S., K. Shirotsuki, & M. Nakao (2019): "The effectiveness of intervention with board games: a systematic review." *BioPsychoSocial medicine* **13(1)**: pp. 1–21.

Rossel, N. (ca. 2020): "What is a tabletop game? a guide with examples." https://agogegameacademy.com/what-is-a-tabletop-game-a-guide-with-examples. Accessed 2023-03-19.

Sirmans, S., D. Macpherson, & E. Zietz (2005): "The composition of hedonic pricing models." *Journal of real estate literature* **13(1)**: pp. 1–44.

Statista (2019a): "Global cards and board games market value from 2018 to 2025 (in billion u.s. dollars)." https://www-statista-com.ezproxy.is.cuni.cz/statistics/1072042/global-card-and-board-games-market-value. Accessed 2022-06-14.

Statista (2019b): "Cards and board games market value in the united states from 2015 to 2025." https://www.statista.com/statistics/1072762/us-card-and-board-games-market-value/. Accessed 2022-04-13.

Statista (2019c): "What price (in $) do you consider appropriate for a new board game?" https://www.statista.com/forecasts/862861/willingness-to-pay-for-a-new-board-game-in-the-us. Accessed 2022-05-2.

Statista (2020a): "Respondents that bought more board games than usual due to covid-19 in the united kingdom in 2020." https://www.statista.com/statistics/1195876/board-game-buying-trends-due-to-covid19-uk/. Accessed 2022-04-13.

Statista (2020b): "Level of enjoyment when playing board games in the united states as of may 2020, by generation." https://www.statista.com/statistics/1129932/us-board-game-enjoyment-by-generation. Accessed 2022-04-13.

Technavio (2022): "Board games market by distribution channel, product, and geography - forecast and analysis 2022-2026." https://www.technavio.com/report/board-games-market-industry-analysis. Accessed 2022-04-13.

Tomášková, Ž. (2009): "Moderní deskové a společenské hry a jejich postavení ve volném čase dnešní populace." *Diplomová práce. Masarykova univerzita. Vedoucí práce Mgr. Petr Soják, Ph. D* .

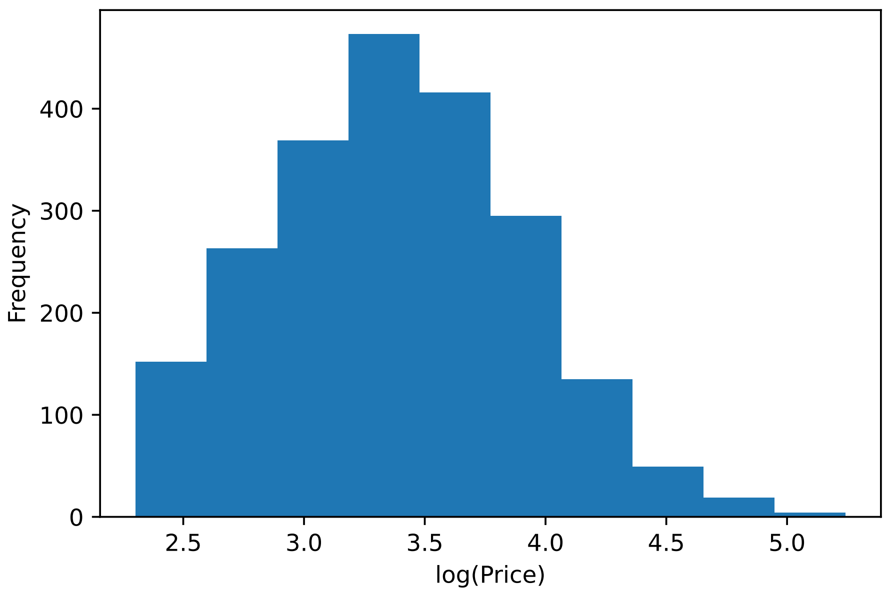
Wachs, J. & B. Vedres (2021): "Does crowdfunding really foster innovation? evidence from the board game industry." *Technological Forecasting and Social Change* **168**: p. 120747.

Wooldridge, J. M. (2012): *Introductory econometrics: A modern approach*, volume Fifth Edition. Michigan State University: South-Western Cengage learning.

Appendix A

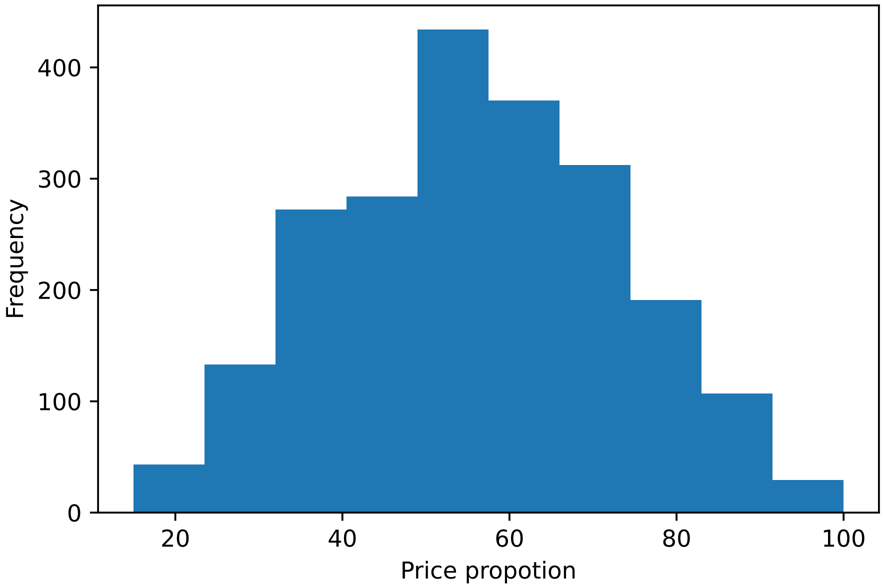
Figures and Tables of Descriptive Statistics

Figure A.1: Histogram of Price



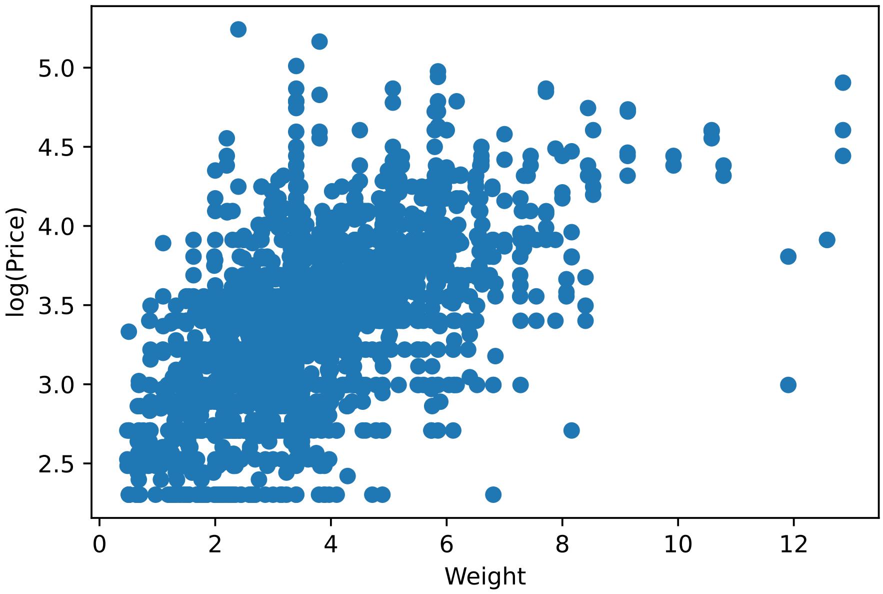
Source: Author's calculations.

Figure A.2: Histogram of Residual Price Share



Source: Author's calculations.

Figure A.3: Scatter Plot of Price and Weight



Source: Author's calculations.

Table A.1: Frequency by Minimum Age

min age	frequency
18	3
17	11
16	8
15	6
14	621
13	288
12	555
11	8
10	378
9	24
8	226
7	14
6	8
missing	25

Source: Author's calculations

Table A.2: Frequency by Age

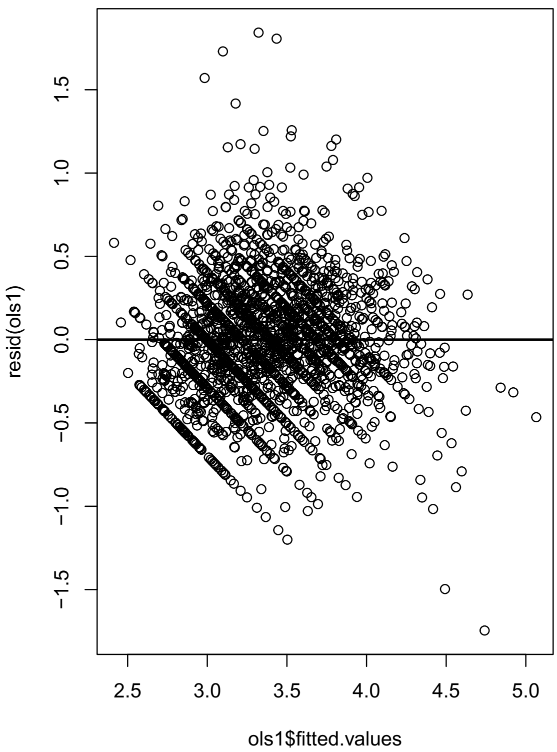
age	frequency
18	50
17	53
16	56
15	56
14	56
13	68
12	105
11	142
10	102
9	157
8	187
7	197
6	213
5	171
4	200
3	178
2	135
1	49

Source: Author's calculations

Appendix B

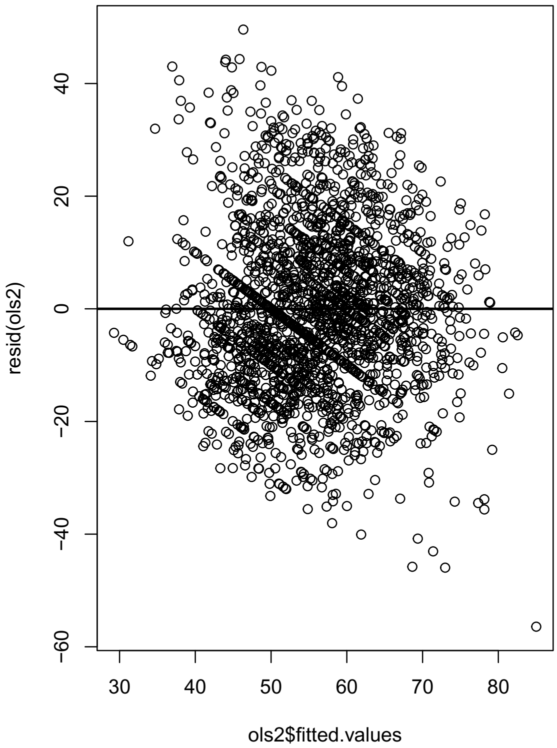
Residuals and Fitted Values Scatter Plots

Figure B.1: Scatter Plot of Residuals and Fitted Values - Regression with Price



Source: Author's calculations.

Figure B.2: Scatter Plot of Residuals and Fitted Values - Regression with Price Retention



Source: Author's calculations.

Appendix C

Robustness Check

Table C.1: Robustness Check - Price Regression

	Dependent variable:
	log(price)
new	0.353^{***} (0.029)
likenew	0.209^{***} (0.028)
verygood	0.129^{***} (0.028)
weight	0.100^{***} (0.007)
age	-0.110^{***} (0.021)
rating	0.049^{*} (0.026)
complexity	0.107^{***} (0.017)
duration	0.002^{***} (0.0003)
players_solo	0.056^{***} (0.019)
card	-0.049^{**} (0.022)
war	0.088^{**} (0.035)
dice	0.062^{**} (0.026)
miniatures	0.224^{***} (0.034)
cooperation	-0.079^{***} (0.022)
scenarios	0.065^{**} (0.032)
luck	0.097^{***} (0.030)
crowdfunding_KS	0.186^{***} (0.018)
z_man	0.062^{*} (0.033)
I(age *rating)	0.016^{***} (0.003)
Constant	1.936^{***} (0.177)
Observations	2,175
R^2	0.544
Adjusted R^2	0.540
Residual Std. Error	0.360 (df = 2155)
F Statistic	135.404^{***} (df = 19; 2155)
Note:	*p<0.1; **p<0.05; ***p<0.01
Note:	Robust standard errors in parentheses

Table C.2: Robustness Check - Residual Price Regression

	Dependent variable:
	price in the second-hand market/ MSRP (in per cents)
new	22.117*** (2.521)
likenew	14.425*** (2.507)
verygood	10.697*** (2.516)
good	6.056** (2.692)
age_1or2	6.901*** (1.151)
rating	8.099*** (0.561)
complexity	−1.935*** (0.632)
duration	0.060*** (0.008)
players_above_3	−5.615*** (1.188)
dice	2.583** (1.185)
adventure	2.623*** (0.997)
storytelling	−4.234*** (1.332)
luck	4.363*** (1.320)
crowdfunding_KS	6.415*** (0.744)
dice_with_icons	−2.730** (1.280)
Constant	−10.529** (4.456)
Observations	2,175
R^2	0.251
Adjusted R^2	0.246
Residual Std. Error	14.620 (df = 2159)
F Statistic	48.320*** (df = 15; 2159)
Note:	*p<0.1; **p<0.05; ***p<0.01
Note:	Robust standard errors in parentheses