CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Lego as investment

Bachelor's thesis

Author: Hana Tomanová
Study program: Economics and Finance
Supervisor: Mgr. Petr Polák, MSc. Ph.D.
Year of defense: 2024

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.

During the preparation of this thesis, the author used `ChatGPT` in order to `improve formulations of the text`. After using this tool/service, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

Prague, July 31, 2024

Hana Tomanova

Abstract

This thesis examines the characteristics of annual growth in value of LEGO sets retired from the primary market. Two dependent variables, annual growth in value of sealed sets and average annual growth in value of used sets, were analysed employing multiple sets specifications. The analysis is conducted on data retrieved from Brickeconomy.com. Utilizing multiple linear regression with ordinary least squares method on cross-sectional data having more than 11000 observations, many characteristics were found to be significant. Age, number of pieces and retail price turned out to have impact on the annual growth in value.

Keywords LEGO sets, investment, annual growth, OLS regression

Title Lego as investment

Abstrakt

Tato bakalářská práce zkoumá determinanty růstu hodnoty LEGO setů již nedostupných v běžném prodeji. Dvě závislé proměnné, roční růst hodnoty zapečetěných setů a průměrný roční rj̊ust hodnoty použitých setů byly analyzovány s použitím řady charekteristik setů. Analýza je provedena na datech získaných ze stránky Brickeconomy.com. Užitím vícenásobné lineární regrese s metodou nejmenších čtverců na datech čítajících přes 11000 pozorování bylo zjištěno několik významných prvků. Proměnné popisující věk setu, počet dílků a prodejní cena se projevily jako prvky ovlivňující roční růst hodnoty.

Klíčová slova sety LEGO, investice, ročnÍ růst, metoda nejmenších čtverců

Název práce Lego jako investice

Acknowledgments

I would like to express my gratefulness to Mgr. Petr Polák, MSc. Ph.D. for his guidance, support, patience and valuable insights. Special thanks belong to Jan Hlaváč for his help with finding a way of retrieving the data. Many thanks belong also to the ones who supported me throughout my studies.

Typeset in LaTeX using IES 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

Tomanová, Hana: *Lego as investment.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 44. Advisor: Mgr. Petr Polák, MSc. Ph.D.

Contents

List of Tables

List of Figures

Acronyms

CD Certificate of Deposit

ETF Exchange-Traded Fund

GMA Gauss-Markov Assumption

GVIF Generalized Variance Inflation Factor

MLR Multiple Linear Regression

OLS Ordinary Least Squares

VIF Variance Inflation Factor

Chapter 1

Introduction

LEGO being one of the most famous toy brand in the world was found to be an interesting well-performing part of the investors' portfolio (Dobrynskaya & Kishilova (2022).

The main ambition of this thesis is to identify significant characteristics of LEGO sets affecting their annual growth in value. Two groups of retired LEGO sets are examined, the sealed sets and used sets. Therefore, the research question is whether such relevant factors can be recognized. As far as we know, this is the first study conducted by using data from Brickeconomy, a website for LEGO enthusiasts, collectors and investors collecting all the possible information about LEGO sets, most importantly about their prices evolution in time. Only the records considering LEGO sets having retired status were taken into account, as it is reasonable to analyse the changes in their value since they are no longer available in the primary market. The conclusions might offer valuable insight for LEGO collectors and investors seeking for an item to enlarge their portfolio. The structure of this thesis has a following form. Chapter 2 provides a concise overview of theoretical background and existing literature discussing investing and LEGO toys. The Chapter 3 is dedicated to the presentation of the data repository, the description of the data cleaning process, introduction of variables chosen for the analysis and limitation of this study. In the Chapter 4, implemented methods and their assumptions are reviewed. Chapter 5 delivers an interpretation of the results and includes a section elaborating on the assessment of the model assumptions and robustness checks. Chapter 7 summarizes the outcomes and discusses potential avenues for further possible research related to the subject of this paper.

Chapter 2

Literature review

In this chapter, the existing literature and research about investing techniques and commodities are analysed and divided into two parts. Firstly, investing as a whole is discussed to broaden the topic's perspective and give a brief overview of investing possibilities. Secondly, essential insights into LEGO as an investment element and the current state of knowledge are presented.

2.1 Ways of investing

There are numerous ways and possibilities when considering investments. Two main approaches to investing can be outlined. They are discussed in the two following subsections.

2.1.1 Common types of investments

Firstly, common or traditional types of investments are especially (Geier (2023)):

- Stocks, also known as shares or equities: possibly the most well-known and simple way of investing, i.e. purchasing a stake in a publicly-traded company
- Bonds represent lending money to a business or a government entity in exchange for interest payments
- Mutual funds involve investing through fund companies into more entities to create a diversified portfolio, thereby reducing investment risk
- Exchange-Traded Funds (ETFs) Similarly to mutual funds: a set of investments bought and sold on stock markets

- Certificate of Deposit (CD) is studied as a very low-risk investment that involves providing a bank with a certain amount of money for an arranged period in exchange for interest on that money

While investing, people are rather looking for a secure long-lasting option that can yield as large profit as possible.

2.1.2 Alternative investments

Other categories of investing possibilities are so-called alternative investments such as fine wine, real estate (often indicated as the common type of investment as well), gold or silver (or precious metals and stones in general), cryptocurrencies, fine art and crowdfunding, among others. Collectables and luxuries have gained popularity as an alternative form of allocating funds (JordĂ *et al.* (2019)). A Barclays (2012) survey affirms this fact by showing that the average high-net-worth individual holds about 10% of their wealth in collectable assets.

The cause of fine wine representing a great alternative investment is its limited supply and big demand. Only about 1 percent of all wines produced worldwide are investment-worthy and supply decreases with every bought or opened bottle. Studies concerning shorter periods, such as 15 years or less (Masset & Weisskopf (2010); Kourtis *et al.* (2012); Lucey & Devine (2015)) came up with a conclusion of quite low net returns of wine investments, although the presence of wine in an investment portfolio helped its return/risk profile. Another study by Dimson *et al.* (2015) working with a long-term period between 1900 and 2012 predicted the real return of wine investments to equal 4.1 percent which outpaced bonds, art, or stamps but underperformed equities and precious metals.

Many studies concerning fine art items as objects of investment were conducted. Prior studies of the art market (e.g., Baumol (1986), Goetzmann (1993), Pesando (1993) have tested art performance in the 17th to the 20th centuries and reached disputed results. While Goetzmann's art index greatly exceeded both stocks and bonds from 1900 to 1986, Pesando (1993) on the other hand, claimed that modern prints failed to be as good as both stocks and bonds from 1977 to 1992 (Dobrynskaya & Kishilova (2022)). However, these studies have used short sample periods or bounded samples of paintings. Other studies concerning large samples of paintings and sales (Mei & Moses (2002) or Renneboog & Spaenjers (2013) during 1875–1999 and 1957–2007 respectively and estimated a 5 percent p.a. (4 percent p.a.) average real return to art. This value is comparable to corporate bond returns and makes art an interesting option for portfolio diversification Dobrynskaya & Kishilova (2022).

Precious metals and stones are also attractive investment options (Renneboog & Spaenjers (2012); Low *et al.* (2016)). Renneboog & Spaenjers (2012) constructed a hedonic price index for diamonds, finding that colored and white gems outrun the stock market in 1999–2010, getting a real return of 2.9 and 6.4 percent p.a. respectively. Auer & Schuhmacher (2013) confirmed that trend by comparing the data with the stock market performance from 2002 to 2012.

2.2 LEGO, a toy or investment?

LEGO Group is a worldwide well-known and considered to be the most favourite toy brand (Handley (2018)) producing 2.2 million bricks an hour which makes annually five times the Earth's total population (Ali (2021)). In compliance with a giant survey conducted among more than 3,000 adults in 2010, LEGO was labeled "the most favourite toy of all times". At the same time, the LEGO Group is considered to be the world's most valuable toy brand with its value of over 7.5 billion U.S. Dollars and according to Statista server (Statista (2024)), the company is making a net profit of more than 1 billion euros every year since 2015.

As it has been observed, the rising trend of prices in time is not the same size for every LEGO product or set. Substantial differences can be observed among individual items. As Dobrynskaya & Kishilova (2022) state, high collectible value, despite lower initial prices, is more probable to be observed by sets having any of the aspects given below.

1. rare parts or minifigures
2. licensed sets
3. large sets with > 1,000 pieces
4. sets with low price per piece ratio
5. sets with short production runs
6. limited edition sets
7. small sets and polybags
8. seasonal sets
9. sets, which were only sold at promotional events
10. unique sets

Dobrynskaya & Kishilova (2022) give a great example of such item, a minifigure of Mr. Gold. Only 5,000 exemplars were manufactured and the item was sold out in three months at original price of €2.49. This figure is demanded by many collectors and offered on the secondary market for approximately €6,500 for new and factory sealed Mr. Gold. Used set can be purchased for an amount in range of €3695 and €4525 depending on the condition of the figure (Brickeconomy.com (ca 2023b)).

2.2.1 Latest findings on LEGO investments

As not many studies or findings on investing into LEGO sets exist, current state of knowledge is discussed in following paragraphs in more detail.

The overall results of both studies show that LEGO sets can represent an interesting item in investor's portfolio and support diversification of it. Although, not every LEGO set is such an item. As discussed above, if some concrete set has any of the 10 given aspects, its collectible value and therefore even interest rate (i.e.price) is more probable to rise.

For example, as Dobrynskaya & Kishilova (2022) claim: small and huge sets, as well as sets based on popular movies or architectural buildings and seasonal ones, yield higher results. They also state that returns are higher in recent years as the second market deepens and more LEGO trading platforms exist. The first claim is partially supported by Shanaev *et al.* (2020) by arguing that smaller sets and sets with lower price-to-piece exhibit higher yields and also that older sets also represent higher returns, demonstrating a liquidity effect.

Dobrynskaya & Kishilova (2022) state that LEGO investments outperform large stocks, bonds, gold, and alternative investments, with an average return of at least 11 % per annum (8% in real terms) internationally in the sample period 1987-2015. Although, they also argue that the LEGO market does not outperform the stock market because of statistically insignificant values in their tests.

On the other hand, Shanaev *et al.* (2020) deal with a six-month sample (September 2018 - March 2019) where they synthesize more than 90,000 LEGO transactions on the secondary market. They found out that transactions within this period exceed €6.3 million. Shanaev *et al.* (2020) also constructed LEGO value-weighted index accounted for survivorship bias over 1966-2018 and state that the index enjoys 1.2% inflation-adjusted return p.a., distinctively below

5.54% for equities. Nevertheless, they claim that the defensive properties of LEGO are substantial and including 5%-25% of LEGO in diversified portfolio is favourable for investors, though, with varying levels of risk aversion.

Shanaev *et al.* (2020) also examine international dimensions of the LEGO secondary market and they claim the market is relatively internationalised, with investors from larger economies, countries with larger per capita incomes, and less income inequality are exposed to trade LEGO more actively. The researchers also detected that the buyers and sellers come from 90 and 65 countries, respectively. And an interesting attribute of the international LEGO market was found, i.e. the market is quite concentrated, with nearly 60% of trades performed by investors from five countries, concretely: USA (with 27.28%), Germany (12.64%), the Netherlands (10.88%), Canada (4.28%) and UK (4.21%).

To sum up all the findings, LEGO sets are considered to be an interesting and lucrative investment possibility almost all over the world. However, this statement does not hold for any LEGO set as some sets are special somehow and that brings a huge impact on gains. The more unique and if unsealed the set is, the higher expectations can its owner create. For getting faster returns, collectable sets are the best choice, as gradually increasing secondary market prices instantly after the release is observed (Dobrynskaya & Kishilova (2022)). As Koford & Tschoegl (1998) as well as Cameron & Sonnabend (2020) state, rarity is the crucial aspect making a toy valuable alternative investment, such as other collectables. If an investor wants to make a diversified portfolio, adding some LEGO investments is recommended.

This thesis aims to examine how the specific aspects of LEGO sets affect the annual growth of their value. Secondly, we want to show if the results of the existing studies concerning gains, rising interest rates, and overperforming of other investment possibilities apply also by taking a sample of all existing LEGO sets and using a different source of secondary market data, concretely BrickEconomy.com, than the prior studies.

Chapter 3

Data

This chapter is devoted to the comprehensive description of the data utilized in this research. It is divided into five sections. To begin with, the Section 3.1 introduces the source of the data. In the following Section 3.2 the procedure of the data collection is described.

3.1 Data source

All the data used in the thesis were taken from BrickEconomy (brickeconomy.com, in its full name BrickEconomy the economics of LEGO). Brickeconomy represents one of the websites for LEGO collectors, enthusiasts and investors collecting and presenting all the available data about existing LEGO sets and their trading.

Besides that, the website offers a lot of awesome content such as a detailed description of all the included sets, up to date containing a total amount of 19 078 sets (as of April 2024)Brickeconomy.com (2024a). Additionally, the webpage features analyses and graphs depicting the price development over time. Nevertheless, BrickEconomy uses Artificial Intelligence (AI) and Machine Learning (ML) "to provide an accurate estimated and predicted price of just every set and minifigure"(Brickeconomy.com (ca 2023a)). The website states that the showed current values (of sets and minifigs) and provided predictions of future prices are based on their machine learning models which are extremely accurate (Brickeconomy.com (ca 2023a)).

Brickeconomy was finally selected since it represents one of the sources not used yet by other authors in already existing works analysing LEGO sets.

3.2 Data retrieval

Data were retrieved and merged with the Python programming language and by running the code in the Windows Command Prompt, cleaned in Microsoft Excel and R and consequently analysed in R.

For the extraction of the data necessary for the following analysis, web-scraping methods by two Python codes were used. Since a distinct number (called a set number) uniquely identifies each LEGO set, the first code serves to download the set numbers of all the existing LEGO sets in the Brickeconomy database. The list of all the set numbers is saved as one text document. The second code downloads the details of each LEGO set listed in the document created by the first code (as described above) and stores all the data in a CSV file. As it would be too demanding for time and energy to download all the data in one run, the code is written so that it allows downloading the data by parts. More precisely, one can retrieve information about the specified number of sets starting with an arbitrary set chosen (defined by a number stating its position in the list of the existing LEGO sets). The entire proceeding of the data retrieval took about 8 hours.

3.3 Data cleaning

This section describes the data preprocessing. As the primary step, all the data retrieved was merged into one CSV file.

Though the output format was a CSV file, the dataset used for the analysis in this paper is present in the form of an Excel sheet as the retrieved data in the CSV file was divided into columns already (not the format with separators) and therefore it made sense to transform the file into an XLSX file directly.

Mainly, a filter on the sets with availability status matching to *'Retired'* was applied. The main reason for this filter is that examining the still retailed sets would not give us much interesting information about the price development at secondary LEGO markets as they are still available at the primary markets (i.e. in (e-)shops). Further, some of the downloaded values were not in numeric format as desired for the analysis. Therefore, those listings were corrected. Nevertheless, some of the values were downloaded as merged in one column and thus, they were separated and cleaned from redundant characters.

Various outlier removal or correction techniques such as the winsorizing method or deleting outliers using the value of interquartile range, were attempted; however, they did not produce the anticipated results. Thus, the outlying values were removed by the problematic variables directly according to the criterions stated below by the specific variables.

3.4 Variables

In this section, we present the key variables utilised in our study on LEGO revenues. These variables have been carefully selected based on their importance in addressing the central research question and contributing to the overall aims of our investigation. By delineating these variables, we aim to provide a comprehensive understanding of the factors under scrutiny and their implications for our research outcomes.

3.4.1 Response variables

The thesis intends to model two response variables described in the following paragraphs.

Annual Growth (of the Sealed Sets) Annual growth represents the average yearly increase in value of the unpacked set since the retirement up to the present day, calculated on an annualized basis. The annual growth ranges from 0 to 38.4%.

Values larger than 38.5% were considered outliers and dropped from the dataset.

Average Annual Growth of Used Sets (AAGUS) The Brickeconomy database offers just two variables concerning used sets, the estimated current value of a used set and the price range of a used set based on its condition. However, no information regarding the growth in value of used sets is provided.

Therefore, such variable was calculated by the author of this analysis using the following formula:

$$AAGUS = \frac{\left(\frac{Value\_used}{Retail\_price} - 1\right)}{(2024 - Year)} \tag{3.1}$$

where:

- $AAGUS$ is the annual average growth in value of the used sets (in %)
- $Value\_used$ represents the current value of a used set (in €)
- *Retail_price* the actual retail price the set was released for in the European Union (i.e. in €)
- $2024 - Year$ is the difference between the year 2024 and the calendar year when a specific set was issued, i.e. the age of the set

Values larger than 900% were considered outliers and removed.

Note: For a more transparent and clear interpretation of the results, both variables the *Annual growth* of the sealed sets as well as the *Average Annual Growth of the Used Sets* were multiplied by 100. Therefore, their values representing the percent average annual growth show the per cents as full numbers, not decimals as usual. This transformation of the values should lead even to better readability of the results in Chapter 5.

3.4.2 Explanatory variables

Value sealed The Value sealed is the estimated current value of the set in a new, factory-sealed and good packaging condition. The value is determined by several factors using the exclusive BrickEconomy algorithm which is designed to accurately estimate a set's value on the open market (Brickeconomy.com (ca 2023a)).

Year and Age

Year The Year variable represents the calendar year when a specific LEGO set was released. The value ranges from the year 1949 till today, i.e. 2024, making it 74 years of LEGO sets being part of the market. But the first known value of the annual growth is observed by a set from 1955. Below, a table of the seven most frequent and of the seven least frequent years is provided (Table 3.1). Overall, the number of released sets varies from 1 set in 1960 up to 503 sets in 2016.

Table 3.1: Frequency by 7 Top and 7 Worst Years

Year	Sets released	Year	Sets released
2016	503	1960	1
2015	496	1959	3
2017	481	1956	5
2018	476	1957	6
2014	457	1955	8
2013	428	1962	9
2020	415	1965	12

Source: Author's calculations

Age To provide a better readability and interpretation of the results, the *Year* variable was subsequently transformed to the *Age* variable. This variable represents the age of a specific LEGO set. More precisely, it specifies the number of complete years that have passed since the set was first released. Thus, it was simply computed using the following formula:

$$Age = 2024 - Year \tag{3.2}$$

where:

- Age is the age of the specific LEGO set,
- 2024 represents the current year
- $Year$ means the year when the set was released

We assume the age to have a positive effect on the annual growth in value of a set, since the older the LEGO set is, the higher demand for such a set is expected to exist.

Retail price The actual retail price the LEGO set was initially released for in the European Union (i.e. in €) - it reflects the manufacturer's suggested retail price. We might expect the retail price to have almost zero or probably a slightly negative effect on the growth in value as the more expensive the LEGO set is, the slower the growth in the value could occur compared to the cheaper sets.

Theme (& Subtheme) The theme and the subtheme of a set are determined by aligning with some of the industry standard sites, mostly Brickset and Bricklink. All the LEGO sets are sorted by their theme and subsequently divided into subthemes (i.e. each LEGO theme can be a combination of subthemes). Further, the Theme variable was used for the analysis, mainly due to easier interpretation and for better clarity as there are 145 Themes and 611 Subthemes (only the summary table of a regression while using the Subthemes variable would cover 36 pages). Last but not least, incorporating all the groups of Themes variable (i.e. including more than 140 dummy variables in the model) is out of scope of this paper. Therefore, a group of 15 most frequent themes (listed in Table 3.2) was kept and the rest is called as "other".

An implementation of the Theme directly in a form of a factor variable in the regressions instead of manually creating dummy variables was considered as this approach offers the following advantages:

1. Efficiency and Simplicity: R automatically handles the transformation of factor variables into dummy variables. This not only streamlines the modeling process but also ensures that the code remains concise and readable. The automatic handling by R reduces the likelihood of errors that might arise from manual dummy variable creation and addition of each dummy variable into the specific model.

2. Consistency and Accuracy: By using factor variables, we leverage R's built-in capabilities to ensure consistent treatment of categorical data. R selects and manages the reference level appropriately, enhancing the accuracy and reliability of our statistical conclusions.

3. Readability and Maintainability: The direct inclusion of factor variables in the model formula leads to cleaner code. This readability is crucial for maintaining the codebase and for facilitating collaboration with other researchers who might review or extend this work.

By directly incorporating categorical variables into the regression models, we would ensure a methodologically sound, efficient, and reproducible analysis, aligning with best practices in statistical modeling and data science. But in the end, we created dummy variables for the 16 mentioned groups (15 most frequent LEGO themes + group of the rest themes) as it enabled us to set a base group of the LEGO themes easily. The 15 dummies for the listed themes were incorporated into the models, briefly described in Chapter 4, and the group of other themes was set as a base group.

In analyzing the evolution of LEGO set values, it is crucial to consider the diverse range of themes that LEGO offers. Each theme, from classic cityscapes to fantasy adventures and licensed special sets, possesses unique characteristics and appeal, which can significantly influence the market value of the sets over time. These variations in themes result in differing collector interests, production quantities, and cultural relevance, all of which contribute to the distinct impact on the value development of LEGO sets.

Table 3.2: Frequency by 15 Top Themes

Theme	frequency
Duplo	829
Minifigure Series	656
City	618
Star Wars	611
Town	588
Technic	394
Friends	350
Creator	347
Bionicle	341
Service Packs	312
Space	301
Ninjago	277
Castle	256
Racers	227
Classic	219

Source: Author's calculations

Pieces This variable provides information about the number of individual pieces included in a specific set. This count includes all distinct pieces that need to be assembled, such as bricks, minifigures, and specialized components.

Minifigures number This characteristic offers knowledge about the count of the minifigures in the set if there are any.

As the LEGO Minifigure is a registered trademark (LEGOGroup (2018)) item, it has to dispose of specific parameters and features. Such part represents

Table 3.3: The 10 Most Common Amount of Minifigures

Number of minifigures	frequency
no minifigure	5924
1	2071
2	959
3	702
4	553
5	309
6	192
7	118
8	86
9	45

Source: Author's calculations

a specific part of the LEGO sets and is usually valued higher by the collectors or buyers. The amount of minifigures in a set has been observed only by sets possessing this "type" of the minifigures. The LEGO brand produces not only the minifigures themselves but also other types of figures, such as Duplo (theme of the LEGO sets designed for children from 1,5 to 5 years old) figures or Friends (sets designed primarily for girls) figures or others usually called dolls. For the consistency of the data and holdback of the number of observations, we added zeroes to all the LEGO sets not having the information about Minifigures number in the original dataset downloaded from Brickeconomy.

Value used The Value used is the approximate current value of a used set in the 'like-new' (i.e. complete with all parts) condition. This element represent the mean value, as pricing of used set can vary significantly due to numerous factors.

3.4.3 Variables not included in the model

In this subsection, a list describing the variables that were not incorporated in the model described in Chapter 4 is provided. Further, reasons for not using these variables are provided.

Subtheme As discussed in Section 3.4.2 already, over 600 subthemes of LEGO sets exist, therefore, the Theme variable was finally chosen for the analysis. Last but not least, some subthemes are called the same across the theme groups or stated as unknown by numerous sets. That might lead to misleading results and false grouping of the observed sets.

Set number The set number, as described in Section 3.2, serves as a unique identifier for the sets and was used for downloading the data and, at the same time, ensured the prevention of duplicate observations.

Growth The growth shows the percent change in the value of the set based on the retail price and its current value. The *Growth* variable is not used in the analysis as the *Annual growth* is present there. Incorporating both of the mentioned variables could lead to collinearity bias.

Price per Piece As the *Price per Piece* variable means the average cost per LEGO piece in the set, that is simply the retail price of the set divided by the sum of all the pieces including bricks and minifigure parts. The *Price per Piece* variable is not incorporated in the analysis because both, the *Retail price* and the number of *Pieces* are present, and using the *Price per Piece* variable would lead to perfect multicollinearity between these three variables and would cause that the results would not be reliable anymore.

Value of Minifigures The Value of the minifigures represents the current estimated value of the minifigures in the specific set.

Several reasons for omitting this piece of information from the analysis exist. Firstly, the variables *Value sealed* and *Value used* were utilized. Secondly, by some of the sets in the dataset, the estimated value of minifigures is even higher than the estimated current value of sealed sets. Lastly, the information about the value of the minifigures is missing by many sets in the dataset and that would lead to lowering the number of observations.

Chapter 4

Methodology

This chapter presents the methodology used for the analysis. In Section 4.1, the model is introduced. The following Section 4.2 briefly explains the assumptions required to obtain the best results.

4.1 Model

Because of the structure of the obtained data, Multiple Linear Regression Multiple Linear Regression (MLR) technique with Ordinary Least Squares Ordinary Least Squares (OLS) was chosen for the analysis. This method provides the following benefits: clear interpretation, prevalence, and conciseness. Conversely, the results are representative and unbiased if and only if all the Gauss-Markov assumptions for cross-sectional regression are satisfied (Wooldridge (2012)).

The model looks followingly:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \epsilon_i, i = 1, 2, ..., n \tag{4.1}$$

where, for $i = n$ observations:

- y_i is the response (also dependent) variable
- $x_{1i}, x_{2i}, ..., x_{ki}$ are k explanatory (also independent) variables
- β_0 is the y-intercept
- $\beta_1, \beta_2, ..., \beta_k$ are coefficients corresponding to $x_1, x_2, ..., x_k$
- ϵ_i is the error term

- n stands for the number of observations

In this study, two main forms of regression are used. In the first model, the annual growth of the value of the sealed sets represents the response variable. In the second model, the average annual growth of the value of used sets stands for the dependent variable.

4.2 Gauss-Markov Assumptions

For the Multiple Linear Regression to provide representative and reliable results of the OLS estimators, several assumptions, also called Gauss-Markov Assumptions Gauss-Markov Assumptions (GMAs), are required. For the MLR with cross-sectional analysis, the assumptions are following:

1.Linearity in parameters A linear relationship between the dependent variable and the independent variables is present in the model. Thus, the relationship needs to be correctly specified by the model.

2.Random Sampling The second assumption claims that the sample has to be random.

3.No Perfect Collinearity In the sample, and thus in the population, none of the independent variables is constant. That means, there are no straight linear relationships among the independent variables.

4.Zero Conditional Mean This assumption sometimes called an exogeneity assumption, requires a zero conditional mean of the error term which means that the independent variables are not correlated with the error term, ensuring that the estimates are unbiased.

5.Homoskedasticity When homoskedasticity is present, the error term has a constant variance. Accordingly, the variability of the error term is consistent across all levels of the independent variables, leading to efficient estimates.

When the first four listed assumptions are satisfied, the OLS estimators are unbiased and consistent. To obtain the best linear unbiased estimator, fulfillment of all the five assumptions is needed Wooldridge (2012).

Chapter 5

Results Discussion

In this chapter, the results from the conducted analysis are discussed. Two dependent variables, the annual growth in value of the sealed sets and the average annual growth (of the value) of the used sets, were modelled employing a decent number of independent variables using R Studio. Therefore, robust standard errors were computed for the interpretation of results. Section 5.1 comments on the results of the annual growth in value of sealed sets modelling. Consequently, Section 5.2 discusses the results by the used sets. Further, Section 5.3 performs the robustness check. Lastly, the Section 5.4 provides a discussion about the fulfillment of the Gauss-Markov assumptions in the indicated models.

5.1 Sealed Sets Annual Growth Modelling

Firstly, the findings from the multiple linear regression analysis conducted to investigate the relationship between annual growth in value of the sealed sets and specific characteristics of the LEGO sets are presented. The analysis aims to determine the extent to which each independent variable predicts the annual growth. The regression results of the first model are depicted in the Table 5.1.

An interaction term among the number of pieces in a set and its retail price was added to the model as we assume that there might be a relationship between these two variables. We believe that the more pieces a specific LEGO set contains, the higher will be its retail price as it costs more to produce such a set. Therefore, the interaction term is included in the model to check whether our assumption holds.

Initially, let us state that the performance of the models is measured by adjusted R-squared value as it better deals with the number of explanatory variables incorporated in the model in contrast to R-squared (Wooldridge (2012)). The model has adjusted R-squared around 0.158 which is quite low value and according to that we might conclude that the model explains ca 16% of the reality only. This state could be caused by the fact that we ultimately included only the 15 most frequent themes of LEGO sets as analysis of the whole existing scale is out of scope of this work. However, the results of the conducted regression might still be significant. Firstly, the age of the set was found to have a negative effect on the annual growth in value of the sealed sets. Specifically, each added year of age is supposed to decrease the annual growth by 0.115 of the basis point and is significant at a 99% significance level. That contradicts our assumption from Chapter 3.

Further, almost all the variables were found to be statistically significant at a 99% significance level. Two exceptions occur, the LEGO sets of the Town theme and the number of pieces in a LEGO set are not statistically significant. To continue with the Themes specification, 7 out of the 15 listed themes showed to have a negative effect. The highest decrease was observed by the Friends theme which is supposed to be almost 3 basis points compared to the group of other sets which was set as the basis group. The other relatively big negative effects are of themes Service Packs, City, Technic, Creator and Racers representing a decrease of approximately 1.76, 1.75, 1.6, 1.44 and 1.42 basis points respectively. The results are in line with our expectations as these themes are considered to be the basic ones. On the other hand, both Bionicle and Ninjago themes have a positive effect of around 2 basis points compared to the 'other sets' group.

As expected, the value of sealed sets showed a positive impact of 0.003 basis points. Another expected outcome is a negative effect of the retail price of 0.025.

Surprisingly, the number of minifigures turned out to affect the annual growth in the value of the sealed sets negatively by 0.052 basis points.

Last but not least, the interaction term was identified to be statistically at a 99% significance level, presenting a little positive effect of the number of pieces on the retail price.

5.2 Used Sets Average Annual Growth Modelling

The second model aims to examine the impact of the same set of independent variables on the average annual growth of the value of the used LEGO sets.

Table 5.1: Regression output - Annual growth of sealed sets

	Dependent variable:	
	Annual growth of sealed sets	
Age	−0.115***	(0.005)
Pieces	0.0004	(0.0003)
Retail_price	−0.025***	(0.003)
Minifigs_num	−0.052***	(0.017)
Value_sealed	0.003***	(0.0003)
Bionicle	2.016***	(0.186)
Castle	0.905***	(0.187)
City	−1.751***	(0.179)
Classic	1.679***	(0.260)
Creator	−1.443***	(0.268)
Duplo	−0.641***	(0.147)
Friends	−2.975***	(0.235)
Minifigures	0.793***	(0.286)
Ninjago	2.006***	(0.299)
Racers	−1.415***	(0.215)
ServicePacks	−1.759***	(0.148)
Space	0.699***	(0.179)
StarWars	1.366***	(0.207)
Technic	−1.609***	(0.174)
Town	0.052	(0.118)
Pieces:Retail_price	0.00001***	(0.00000)
Constant	10.324***	(0.149)
Observations	11,114	
R^2	0.159	
Adjusted R^2	0.158	
Residual Std. Error	4.286 (df = 11092)	
F Statistic	100.049*** (df = 21; 11092)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
Note:	Robust Standard Errors in parentheses	

Identically as in the case of sealed sets, the interaction term between the number of pieces and the retail price of a LEGO set was included in the model. The regression output is presented in Table 5.2.

The adjusted R-squared in this case is 0.436, i.e. significantly higher than in the case discussed in the previous section. As expected, the age of the sets, the number of pieces, the retail price, and the number of minifigures were found to be statistically significant again, all excepting the number of minifigures (that was by 95% significance level) by 99% significance level. By contrast to the sealed sets, the age was revealed to have a positive impact of 0.447 basis points by the used sets which aligns with the expectations. On the contrary, the number of pieces turned out to have a negative effect of 0.015 basis points.

Similarly to the previous case, the retail price and the number of the minifigures revealed to influence the dependent variable negatively, by 0.3 and 0.188 subsequently.

Further, the value of the used sets exhibited having a positive impact of 0.228 basis points on the average annual growth in the value of used sets at 99% significance level.

Moving to the Theme dummy variables, numerous themes turned out to have a low statistical significance including Castle, Creator, Friends, Racers, Star Wars, and Technic theme. Further, the Ninjago and Space themes revealed to be statistically significant only at 90% significance level. Out of the statistically significant themes, the Classic theme showed a positive impact of 30.501 basis points compared to the other themes group at 99% significance level. Further, the Town theme was revealed to have a negative impact of 11.378 basis points. At 95% significance level, City and Bionicle themes showed to have a positive impact of 2.958 and 3.413 basis points respectively.

Nevertheless, the interaction term was fund to be statistically significant (as by sealed sets) at 99% significance level, showing some positive impact of the number of pieces on the retail price.

5.3 Robustness Check

To verify the reliability of the results, robustness checks were conducted. Thus, both response variables were modelled once again utilizing only those variables having a p-value lower than 0.05 meaning being statistically significant at least at 95% significance level. New regression and their outputs can be find in Appendix B (Table B.1 and Table B.2). By comparing the adjusted R-squared

Table 5.2: Regression output - Annual growth of used sets

	Dependent variable: Average annual growth of used sets	
Age	0.447***	(0.045)
Pieces	−0.015***	(0.003)
Retail_price	−0.300***	(0.029)
Minifigs_num	−0.188**	(0.083)
Value_used	0.228***	(0.013)
Bionicle	3.413**	(1.451)
Castle	−0.125	(1.738)
City	2.958**	(1.406)
Classic	30.501***	(5.238)
Creator	1.608	(1.453)
Duplo	−5.688***	(0.722)
Friends	−1.432	(1.131)
Minifigures	6.391***	(0.852)
Ninjago	3.847*	(1.984)
Racers	−0.578	(1.900)
ServicePacks	−5.879***	(1.329)
Space	−4.201*	(2.329)
StarWars	−0.771	(1.000)
Technic	−1.328	(1.040)
Town	−11.378***	(0.981)
Pieces:Retail_price	0.0001***	(0.00001)
Constant	0.807	(0.838)
Observations	11,114	
R^2	0.437	
Adjusted R^2	0.436	
Residual Std. Error	28.812 (df = 11092)	
F Statistic	410.356*** (df = 21; 11092)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
Note:	Robust Standard Errors in parentheses	

values with the previous results, we can claim that the performance of the models remained unchanged.

5.4 Testing GM assumptions

The satisfaction of GMA.1 (Linearity in parameters) is supposed to be a consequence of the specification of the model. In our case, it is originated in the Chapter 4 (equation 4.1).

GMA.2 (Random sampling) is intricate not to violate at all. As the chosen source of the data for our analysis is one of the platforms storing information about all the existing LEGO sets and a description of all the listed LEGO sets has been downloaded, we are convinced that it should be a convenient sample. Furthermore, we applied a filter to sets designed to the examination - only the retired sets were taken into account.

To inspect whether the GMA.3 (No perfect collinearity) holds, several measures were implemented. We computed the Variance Inflation Factor Variance Inflation Factor (VIF) of the linear regression using the vif() function in R. It is a regularly used method for identifying multicollinearity presence in regression models. When categorical variables (the dummy variables depicting the Themes in our case) are present in the dataset or the model, the so-called Generalized Variance Inflation Factor Generalized Variance Inflation Factor (GVIF) needs to be computed. GVIF is an expansion of VIF that is commonly used for detecting the multicollinearity in regression models including categorical variables with more than two levels. We used the type 'predictor' that is suitable for models including interaction variables. As we obtained values lower than the maximum threshold which is approximately 3.16 (i.e. a square root of 10), we may conclude that the perfect collinearity is not a problem in the models identified in this work. Further, standard VIF was plotted and is depicted in Figure A.1 and Figure A.2. The rule of thumb for VIF is that the value should not be higher than 5 which is fulfilled in both cases for all the variable. Thus, our previous conclusion that perfect collinearity is not present in the model is verified by visualization of VIF.

For checking GMA.4 (Zero conditional mean) As the intercept is included in both models, we can believe that this assumption is fulfilled. Furthermore,

As we analyse the data in cross-sectional form, the assumption GMA.5 (Homoskedasticity) needs to be examined as well. To test whether homoskedasticity is present in the models, the Breusch-Pagan test was applied. In both cases, we rejected the null hypothesis that there is homoskedasticity at a 99% significance level. Therefore, robust standard errors were used. Nevertheless, it is recommended to always utilize heteroskedasticity-robust standard errors while having a large sample. These adapted errors are valid in the presence of heteroskedasticity of unknown form.

Chapter 6

Conclusion

LEGO sets have evolved beyond mere toys; they are now being recognized as valuable assets in investors' portfolios. Therefore, gaining insight into the determinants that influence the annual growth in value of these sets is highly beneficial. This thesis aims to study the specifications of retired LEGO sets to determine the key specifications driving the annual growth in the value of the sets.

Data utilized for the conducted analysis were primarily derived from Brickeconomy.com, one of the biggest websites accumulating all the existing data about existing LEGO sets. To the author's best knowledge, this is the first study using this resource of data for examining LEGO sets. Thus, the main goal of this thesis was to determine the characteristics affecting the annual change in the value of LEGO sets using current data from a source that has not been used yet. The data retrieval and the data preprocessing and cleaning represented a demanding part of this examination as the dataset was collected by using web scraping techniques from Brickeconomy.com. The collection of the data proceeded in April 2024. The downloaded dataset contained 16,219 listings. After applying some filters and cleaning the data set, a dataset of more than 11,000 sets was used for the analysis. To examine the sets' characteristics impact on the annual growth in value of the dataset, MLR and OLR were utilized to model two dependent variables, the annual growth in value of sealed sets and the average annual growth in value of used sets. Fulfillment of GM assumptions was considered.

The variables affecting the annual growth were identified. Surprisingly, different effects on the sealed LEGO sets and on the used sets recognized. Especially by the Themes dummy variables.

Though, the work has many limitations. Firstly, the average annual growth of value of the used sets was computed using the available information about retail price, age and current value. Therefore, the values might be imprecise. Therefore, further research concerning the used LEGO sets is recommended. Furthemore, incorporating more specifics of the LEGO sets (e.g. even subthemes) would possibly bring better performing model (by both the groups of LEGO sets). Unfortunately, such analysis was out of scope of this work. Another possible expansion of the analysis would be using panel data and examining the behaviour of the LEGO sets' value in time.

Bibliography

Ali, S. (2021): "Surprising new study finds investing in legos better than gold, stocks, bonds and art." https://thehill.com/changing-america/enrichment/arts-culture/585311-surprising-new-study-finds-investing-in-legos-better/. [Online; accessed 2024-04-26].

Auer, B. R. & F. Schuhmacher (2013): "Diamonds â€" a precious new asset?" *International Review of Financial Analysis* **28**: pp. 182–189.

Barclays (2012): "Profit or Pleasure?" *Technical Report Volume 15*, Barclays Wealth.

Baumol, W. J. (1986): "Unnatural value: or art investment as floating crap game." *The American Economic Review* **76(2)**: pp. 10–14.

Brickeconomy.com (2024a): "Welcome to brickeconomy." `https://www.brickeconomy.com/`. Accessed: 2024-04-25.

Brickeconomy.com (ca 2023a): "About brickeconomy." `https://www.brickeconomy.com/about`. Accessed: 2023-04-25.

Brickeconomy.com (ca 2023b): "Lego minifigure series 10 mr. gold." `https://www.brickeconomy.com/set/71001-19/lego-minifigure-series-10-mr-gold`. Accessed: 2023-04-20.

Cameron, S. & H. Sonnabend (2020): "Pricing the groove: hedonic equation estimates for rare vinyl records." *Applied Economics* **52(50)**: pp. 5516–5530.

Dimson, E., P. L. Rousseau, & C. Spaenjers (2015): "The price of wine." *Journal of Financial Economics* **118(2)**: pp. 431–449.

Dobrynskaya, V. & J. Kishilova (2022): "Lego: the Toy of Smart Investors." *Research in International Business and Finance* **59(101539)**.

GEIER, B. (2023): "11 different types of investments and how they work." https://smartasset.com/investing/types-of-investment. [Online; accessed 2023-06-16].

GOETZMANN, W. N. (1993): "Accounting for taste: Art and the financial markets over three centuries." *The American Economic Review* **83(5)**: pp. 1370–1376.

HANDLEY, L. (2018): "How marketing built Lego into the world's favorite toy brand." https://www.cnbc.com/2018/04/27/lego-marketing-strategy-made-it-world-favorite-toy-brand.html. [Online; accessed 2024-04-26].

JORDĂ , Ă., K. KNOLL, D. KUVSHINOV, M. SCHULARICK, & A. M. TAYLOR (2019): "The Rate of Return on Everything, 1870â€“2015*." *The Quarterly Journal of Economics* **134(3)**: pp. 1225–1298.

KOFORD, K. & A. E. TSCHOEGL (1998): "The market value of rarity." *Journal of Economic Behavior Organization* **34(3)**: pp. 445–457.

KOURTIS, A., R. N. MARKELLOS, & D. PSYCHOYIOS (2012): "Wine price risk management: International diversification and derivative instruments." *International Review of Financial Analysis* **22**: pp. 30–37.

LEGOGROUP (2018): "FairPlay - a guide compiled by the LEGO Group." `https://www.lego.com/cdn/cs/legal/assets/blt1a4c9a959ce8e1cb/LEGO_Fairplay_Nov2018`. [Online; Accessed 2024-07-09].

LOW, R. K. Y., Y. YAO, & R. FAFF (2016): "Diamonds vs. precious metals: What shines brightest in your investment portfolio?" *International Review of Financial Analysis* **43**: pp. 1–14.

LUCEY, B. M. & L. DEVINE (2015): "Was wine a premier cru investment?" *Research in International Business and Finance* **34(C)**: pp. 33–51.

MASSET, P. & J.-P. WEISSKOPF (2010): "Raise your glass: Wine investment and the financial crisis." *SSRN Electron. J.* Accessed: 2024-7-15.

MEI, J. & M. MOSES (2002): "Art as an investment and the underperformance of masterpieces." *The American Economic Review* **92(5)**: pp. 1656–1668.

PESANDO, J. E. (1993): "Art as an investment: The market for modern prints." *The American Economic Review* pp. 1075–1089.

Renneboog, L. & C. Spaenjers (2012): "Hard assets: The returns on rare diamonds and gems." *Finance Research Letters* **9(4)**: pp. 220–230.

Renneboog, L. & C. Spaenjers (2013): "Buying Beauty: On Prices and Returns in the Art Market." *Management Science* **59(1)**: pp. 36–53.

Shanaev, S., N. Shimkus, B. Ghimire, & S. Sharma (2020): "Children's toy or grown-ups' gamble? LEGO sets as an alternative investment." *The Journal of Risk Finance* **21(5)**: pp. 577–620.

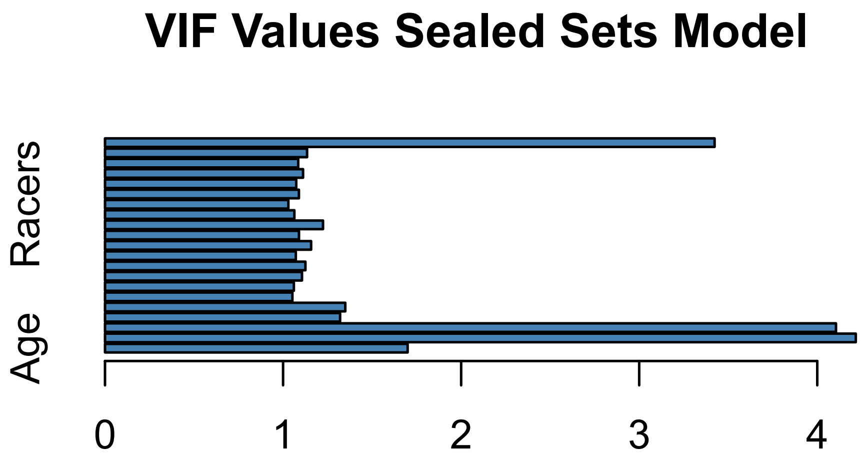
Statista (2024): "Lego minifigure series 10 mr. gold." `https://www.statista.com/statistics/282870/lego-group-revenue/`. Accessed: 2023-04-20.

Wooldridge, J. M. (2012): *Introductory Econometrics: A Modern Approach*. ISE - International Student Edition. South-Western, 5th edition.

Appendix A

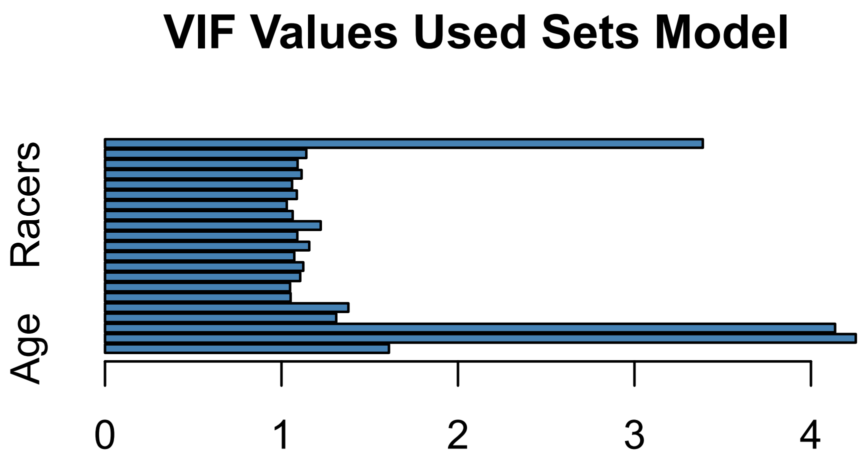
VIF Values

Figure A.1: VIF of Sealed Sets Model



Source: Author's calculations.

Figure A.2: VIF of Used Sets Model



Source: Author's calculations.

Appendix B

Robustness Check

Table B.1: Robustness Check - Sealed Sets

	Dependent variable:	
	Annual_growth	
Age	−0.115***	(0.005)
Retail_price	−0.023***	(0.003)
Minifigs_num	−0.051***	(0.016)
Value_sealed	0.003***	(0.0003)
Bionicle	1.991***	(0.184)
Castle	0.884***	(0.185)
City	−1.748***	(0.178)
Classic	1.689***	(0.253)
Creator	−1.390***	(0.262)
Duplo	−0.709***	(0.141)
Friends	−2.974***	(0.235)
Minifigures	0.739***	(0.283)
Ninjago	2.011***	(0.299)
Racers	−1.423***	(0.214)
ServicePacks	−1.792***	(0.146)
Space	0.679***	(0.176)
StarWars	1.358***	(0.207)
Technic	−1.574***	(0.171)
Retail_price:Pieces	0.00001***	(0.00000)
Constant	10.370***	(0.145)
Observations	11,114	
R^2	0.159	
Adjusted R^2	0.158	
Residual Std. Error	4.287 (df = 11094)	
F Statistic	110.357*** (df = 19; 11094)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
Note:	Robust Standard Errors in parentheses	

Table B.2: Robustness Check - Used Sets

	Dependent variable:	
	Av_annual_growth_used	
Age	0.433***	(0.043)
Pieces	−0.015***	(0.003)
Retail_price	−0.303***	(0.028)
Minifigs_num	−0.173**	(0.082)
Value_used	0.228***	(0.013)
Bionicle	3.570**	(1.445)
City	3.016**	(1.383)
Classic	31.024***	(5.216)
Duplo	−5.481***	(0.695)
Minifigures	6.378***	(0.862)
ServicePacks	−5.523***	(1.267)
Town	−11.044***	(0.893)
Pieces:Retail_price	0.0001***	(0.00001)
Constant	0.915	(0.851)
Observations	11,114	
R^2	0.437	
Adjusted R^2	0.436	
Residual Std. Error	28.820 (df = 11100)	
F Statistic	661.454*** (df = 13; 11100)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
Note:	Robust Standard Errors in parentheses	