ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Behavioural Economics

An Exploratory Study of Price Development in the Rapidly Growing Sneaker Resale Market

Student: Megan van der Ham

Student ID number: 483271

Supervisor: J.A. de Jong

Second assessor: S.C. van der Zee

July 23, 2022

Acknowledgement

I would like to start this paper off by deeply expressing my appreciation and gratitude to my supervisor dr. J.A. de Jong. His expertise has proven to be really helpful since the initial stage of conceptualization of this thesis until its final version, and his constant willingness to help allowed me to learn a lot and pursue a topic in which I have a personal interest.

Furthermore, I would like to thank Erasmus University Rotterdam for the challenging MSc in Economics and Business, for which I have completed both the Behavioural Economics and Data Science & Marketing Analytics specializations, allowing me to combine what I learned in both my fields of interests into a powerful skillset that I can put to good use in my future.

Abstract

Since the last two decades, a sneaker resale culture is developing and reselling sneakers has even become a profitable business model. While this market is growing rapidly, almost no research has been done about its occurring mechanisms, especially regarding the price development. Therefore, the goal of this thesis is to perform exploratory research concerning the price development in the sneaker resale market, using real-time data obtained from one of the biggest online resell platforms StockX. Since price development analysis is comparable to several events in the financial markets, standard financial analyses are taken as the basis for the analyses. Results have proven that the resale market introduction price of sneakers can be forecasted to some extent when employing the sneaker characteristics and fixed price indicators found in previous literature. Also, it is shown that certain sneaker characteristics such as model, colour, sex and retail price and the fixed price indicators regarding brand collaboration and nostalgia influence the variation of the resale market introduction price. Furthermore, this thesis has proven that there exists a declining trend in the market price development of sneakers over the first ten weeks after release, suggesting that selling early will, on average, result in the highest profit for resellers and simultaneously, purchasing later on will, on average, result in the lowest cost for consumers. Finally, this thesis has shown that in the latter case, unlike the first one, it is not necessary to introduce more complex methods, as the statistical method has proven to be sufficient.

Keywords: Sneaker resale market; Single price prediction; Price trend development; Multiple linear regression; Random forest regression; Linear trend model.

Table of content

1. Introduction

1.1 Sneaker culture

Sneakers (also known as *trainers, kicks* or *athletic shoes*) are shoes that were predominantly designed for physical exercise. However, they have evolved into shoes that are currently also used for everyday casual wear. In 2020, the sneaker market revenue (i.e., revenue obtained solely from the sales of the companies to consumers, excluding resale) was approximated to be $70 billion (Statista Research Department, 2021), with the biggest brands being Nike, Adidas and Puma, respectively (Statista, 2022a). Back to where it all started, the sneaker culture originates from the hip-hop and basketball community that arose in New York. According to Kawamura (2016) there were three key pillars within the sneaker culture development: (i) the emergence of hip-hop in the 1970s with Run DMC releasing a song called “My Adidas” in the 1980s; (ii) Michael Jordan releasing its first sneaker in collaboration with Nike in the 1980s; and (iii) the digital age facilitating growth in sneaker marketing and the resale culture.

It is only since the last two decades that the rise in the sneaker culture is accompanied by a resale culture (Weinhold, 2020). Where it started with predominantly sneakerheads (i.e., sneaker enthusiasts) camping outside of stores to get a pair of the newest sneakers to add to their personal collection, the product developed into a key product for a profitable business model. The sneaker market is currently at an all-time high when it comes to popularity, making it easy for sneaker resellers to profit from. A sneaker reseller is someone who aims to get their hands on deadstock sneakers, especially the most popular models, so they can sell it with the highest possible profit to consumers who are willing to pay more than the retail price to add a new pair of deadstock sneakers to their collection. It is based on the simple concept that occurs in many other businesses as well: buy low, sell high (Business Insider, 2022). The emphasis lies on the sneakers being deadstock, which means that they are still unused, in its original packaging and no longer available for sale on the primary market, leading to an increase in perceived value (Weinhold, 2020).

1.2 Research objective

One of the reasons for the success of sneakers is the athleisure trend that has been ongoing for the past few years. This trend comes with the demand for comfortable fashion, including sneakers (Braithwaite, 2021). Where sneakers started off as being footwear especially created for sportive activities, they currently evolved to be perceived as status symbols, a collectible, or even an alternative asset. With this, the sneaker resale market is a emerging rapidly. Since the popular sneakers are often sold out within seconds on the official websites and in-store (i.e., the primary market), leaving lots of consumers unfulfilled, they pursue them on other platforms (Watts, 2019), all together forming the sneaker resale market (i.e., the secondary market). Statista (2022b) estimated the global sneaker resale market to be valued around $6 billion in 2019, and Cowen Equity Research (n.d.) estimates this resale market to grow to a value of $30 billion by 2030, resulting in a growth rate of 500% over approximately one decade[1]. Therefrom can be concluded that this market has a high potential, emphasizing the importance to perform a more in-depth analysis.

The price of sneakers on this resale market, better known as the market price, is rather volatile since it changes often. One reason for this is because sneaker resellers add an arbitrary price premium, representing their personal desired profits. In general, the market price on the sneaker resale market is also more elastic than the retail price (Choi, 2017). Furthermore, almost no insights regarding (the development of) the market price in this specific market exist, making it interesting for further analysis. Additionally, not all sneakers are popular enough to make a profit by reselling them and therefore, sneaker resellers mainly aim to focus their sales on the popular, hyped sneakers. Chu and Liao (2010) found that resellers' awareness of future online resale potential can influence their decisions to invest in a sneaker. They seek successful sales transactions and do not want to risk holding excessive inventory (Chu and Liao, 2007). The ignorance and additional risk is what makes it interesting to perform an in-depth analysis on sneaker market prices, and to look at the possibility of sneaker price forecasting. Therefore, this thesis will consist of a rather exploratory research[2] with the focus on answering the following research question:

"To what extent can the market price of sneakers on the resale market be forecasted before release?"

Since price analysis is comparable to several events occurring in the financial markets, standard financial analyses are taken as basis for this research. To be more specific, the answer to the research question consists of two parts: (i) predicting the introduction price of sneakers on the resale market in the first week after the release date[3], and (ii) evaluating the average market price trend development of sneakers in the first ten weeks after release, both represented by their own sub questions.

[1] Which calculates to a Compound Annual Growth Rate (CAGR) of 15.76%.
[2] Focussing on the two biggest sneaker brands in the primary market.
[3] In the form of single price prediction, as compared to the second part which focuses on price forecasting over time.

1.2.1 Resale market introduction price prediction

To answer the main research question, the first part of the analysis regards to the prediction of the introduction price on the resale market after the release date. No research about this specific topic has been done yet. However, the concept is comparable to companies trying to raise capital by issuing stocks, also known as *Initial Public Offerings (IPOs)*. IPO refers to the situation in which a private company decides to become a publicly-traded and owned entity (Investopedia, 2021a). In this situation, the currently private shareholders' shares will become publicly traded and therefore, their price will be adjusted to the market value. This is comparable to a sneaker that is being introduced into the resale market and whose price will eventually adjust to the market value. In both cases, it is required to set an initial price for the product to enter the market with, aiming to accurately reflect its true market value. The development of stock prices shortly after it becomes tradeable in the public market will be referred to as IPO price prediction in this thesis. Lots of predictions are done regarding this price development of recently issued shares, and both statistical methods (e.g., regressions) and machine learning methods (e.g., decision trees or black box models) have proven to work sufficiently here. This part of the thesis will attempt to predict the resale market introduction price using both type of methods, to see which one performs better. Therefore, the first sub question that arises is:

1. Does a statistical method or a machine learning method perform better in predicting the introduction prices of sneakers in the resale market?

The results obtained from the analyses in this part will give insights in how well, in general, the introduction price of sneakers in the resale market can be predicted, as well as in what method performs best based on prediction accuracy. Lastly, since the relationships between the variables are identified as well using these methods, this thesis can provide insights into what factors are most important in predicting the introduction price of sneakers on the resale market.

1.2.2 Average market price trend development

The second part of the analysis focuses on the average development of the market price in the first ten weeks after the sneaker is released. This type of analysis is similar to time-series analyses for certain metrics in the financial markets, such as inflation, exchange rates, sales charts and interest rates. Usual analyses that are performed in these cases are trend analysis, often in combination with forecasting. Trend analyses may be well reflected in the goal of sneaker market price development, making this comparability a fair one. Here, the aim is to construct the average market price trend of sneakers and evaluate whether a trend exists in the first ten weeks after release. This helps understanding the average market price development and perhaps, when it is proven that there is a specific trend in market price, see whether phenomena such as the *no arbitrage condition* (i.e., stating that no profit can be made without risk and an initial endowment) hold, or whether it is possible to predict average market prices after these first ten weeks. To evaluate whether a trend exists, the goal is to compare the constructed regression of the average market price to a horizontal line (i.e., a line representing no trend). This leads to the following sub question:

2. Can the average market price trend be perceived to be different from a horizontal line?

The results obtained from this analyses will provide insights into the extent to which there is an average market price trend. If there indeed can be concluded that the trend differs from a horizontal line (i.e., the slope of the regression differs from zero), this could give insights into the best timing to sell and purchase sneakers on the resale market.

1.3 Research contributions

1.3.1 Comparison to other markets

As the sneaker market can be considered a special market, it is not a unique market. The same type of business mainly occurs in luxury markets, such as that of cars (e.g., brands such as Bugatti, Rolls-Royce and Ferrari) and watches (e.g., brands such as Rolex, Patek Philippe and Audemars Piguet) or other jewellery (e.g., brands such as Harry Winston and Bulgari). What all these luxury brands have in common is the fact that their business decision is to deliberately make (some of) their products scarce to create desire via exclusivity. This can be accomplished by developing limited edition products, only selling via certain stores or maintaining a waiting list for newly released items (WealthX, 2018). Besides this, some sneakers have even become comparable with more prestigious goods, such as art sold at auctions as well. Take for example the most expensive pair of sneakers ever, which is sold by the auction house Sotheby's for a price of $1.8 million.

One small difference between these luxury markets and the sneaker market is that these previously named brands already take this exclusivity into account in their business model. Their goal is to significantly improve profitability themselves by maintaining extremely high retail prices for their most exclusive products in the primary market, compared to similar products in this market (Growth and McDaniel, 1993). In contrast, companies selling sneakers (e.g., Nike and Adidas) always stick to a relatively low price cap, and a big part of profit on the most exclusive products is made via the resale market, being the secondary market. Even though these companies do not receive direct revenues from resale, they do act on the current situation, e.g., by increasing the number of limited edition drops and the additional hype around these launches (Cowen Equity Research, n.d.). By keeping their sneaker sales to a maximum quantity and acceptable price, the sneaker resale market is continued to be fuelled and no market share in the primary market gets lost (Highsnobiety, 2018). Additionally, it is a profitable marketing strategy since people continue to be hyped about the brands without them having to put any effort in their advertising for example. Regardless of the fact that the distribution of profit between the company and resellers differs between the different markets, they all deal with a disparity between the retail and market price of their products.

Examples of comparable industries in which price trend analysis is more common, are that of oil and electricity, but also in the stock market. The oil and energy prices have been more volatile since the 1973 oil crisis (Regnier, 2007; Verleger, 1994; Fleming and Ostdiek, 1999). Also, after no longer regulating electricity prices by the means of a fixed price as a function of supply costs, electricity prices modified to being dynamic and uncertain (Goto and Karolyi, 2004). Additionally, lots of researchers spent time on analysing the stock market volatility over time (Schwert, 1989). This occurring volatility has led to a bigger interest in understanding the price development of these goods, leading to many different types of analyses that are proven to be successful.

All in all, market prices of all categories are established by supply and demand of the product. Therefore, results of research on the price development occurring in the sneaker resale market can eventually be applied to these other product categories as well.

1.3.2 Comparison to other phenomena

Where sneakers seem comparable to other products, it is also possible to compare the action of price prediction and trend evaluation (as done in this thesis) to phenomena in financial markets. The first phenomenon that is focussed on in this thesis is IPO analysis, which is derived from stock markets. As previously mentioned, the sneaker resale market introduction price prediction can be compared to IPO short term price prediction. Previous literature regarding this topic uses both statistical models such as linear regressions and machine learning models such as decision tree models to predict the IPO price development. The previously mentioned techniques inspire the choice of methodology of the proposed analysis. However, since this is a newly analysed market, the performance of these different models in the case of sneaker resale prices cannot be assumed yet. For the second part of this thesis, the type of analysis is comparable to trend analysis, which is also common in financial markets. In general, this approach evaluates the trend development over time of certain metrics such as inflation, exchange rates, sales charts and interest rates, as discussed before. To conclude, the practice of applying existing knowledge of IPO or trend analysis on a more “mysterious” market is a relevant scientific insight that would not only fill a gap in single price prediction and market price trend development literature, but also in sneaker resale market literature.

2. Theoretical background

In this section, the focus lies on previous research regarding the sneaker resale market, including the influence of sneaker characteristics, fixed price indicators and behavioural characteristics of the resellers and consumers on the price development. Furthermore, it focuses on IPO price prediction and trend analysis, including a comparison of the financial market to the sneaker resale market. These findings in literature will give a better overview of the most important aspects to take into account in order to best perform both analyses, as well as help phrase hypotheses to test in order to answer the research question.

2.1 Sneaker resale market

In general, a resale market can be defined as the place where the transactions occur after the retail market (Choi, 2017), and is previously labelled as the secondary market. Examples of traditional resale markets are garage sales and consignment stores, however, with the rise of the internet this expended to online C2C markets (e.g., eBay or Craigslist) as well (Chu, 2013; Chu and Liao, 2007; Slaton and Pookulangara, 2020). Applying this to the sneaker market, the process of sneaker reselling has also developed over the last few decades. Where it started with forums for resellers and consumers to buy, sell or trade sneakers in the early 2000's, it developed into real-life global sneaker events in the late 2000's (i.e., events where sneakerheads come together to trade in sneakers and enjoy being part of a community), with one of the largest to date being Sneaker Con (Weinhold, 2020). While this event is more centred in the United States, another well-known sneaker event centred in Europe is Sneakerness. Next to these events, many online sneaker resale platforms such as StockX, GOAT and Stadium Goods make reselling even more accessible because of the worldwide (and real-time) connection between resellers and buyers they are providing. Additionally, sneakers are also sold through online public C2C markets such as eBay and local marketplaces such as the Dutch website Marktplaats. In this thesis, the focus lies solely on the online C2C resale market of sneakers.

2.2 Fixed price indicators

A fascinating element of the sneaker resale market is the market price, which can overall be considerably higher, equal or sometimes even lower than the retail price. This thesis focuses on the ability to predict the market price of sneakers and therefore, it is of importance to understand what influences (the development of) this market price. Choi and Kim (2019) have analysed the communication of sneakerheads and their attitudes towards sneakers. Interesting insights regarding their communication is that often the anticipated releases are discussed, including the purchase intensions and thoughts about materials, designs and release quantities. Also, sneakerheads' goals are to obtain the more rare sneakers (Cassidy, 2018). This shows that the sneaker characteristics seem to be of importance. When it comes to other fixed price indicators, several researchers such as Zhenxiang (2020) and Choi and Kim (2019) found that sneaker market prices are mainly influenced by (predetermined) scarcity in the market (i.e., due to demand surpassing supply), the star effect (i.e., the shoe is designed in collaboration with, or worn by, their idol or any other example figure), and nostalgia (i.e., emotional attachment to original models). Each of these aspects increases the hedonic value of the product, and are further elaborated in this section.

2.2.1 Scarcity of sneakers

It is proven that the scarcity, or rarity, of sneakers plays an important role in purchasing intensions, as it does in any collectible market. Explained from an economic perspective, providing a limited quantity of a wanted product at a price below the equilibrium price causes the market to be imbalanced and creates a gap between supply and demand (i.e., demand is surpassing the supply). This gap enhances the value of the product by introducing a subjective desirability (Lynn, 1991), which can be explained by the Commodity Theory introduced by Brock (1968). This theory explains that the value of a product is determined by its availability, meaning that when a product is scarce it is perceived to be of high value, making it more attractive to possess. Additionally, scarcity fulfils customers' desire to be unique and trendy (Choi and Kim, 2019). Lots of luxury brands incorporate the Rarity Principle that is based on this theory, which states that they *"must sustain high levels of awareness and tightly controlled brand diffusion to enhance exclusivity"* (Dubois and Paternault, 1995). This not only applies to luxury brands, but also to the sneaker market. One example of this is Nike releasing a certain quantity per sneaker such that only roughly 4% of it ends up in the resale market leading to a more exclusive availability, however, just enough so that resellers cannot charge big premiums where Nike only makes smaller revenues (Highsnobiety, 2018). With this strategy they sustain a cheap way of marketing, since the sneakerheads drive the hype and the brand remains prestigious due to unmatched demand (Luber, 2015).

As Cassidy (2018) rightfully concludes, the demand for sneakers is in part a function of the supply. Consumers are specifically looking for sneakers that have a limited supply. When a sneaker is still available in the primary market, chances are that this product has no added, or even a lower, value in the secondary market. To conclude, maintaining scarcity introduces a desire within the customers which creates and stimulates the success of the sneaker resale market as a whole.

2.2.2 Star effect

When it comes to sneakers, social identity is of importance as well. By wearing certain sneakers, people want to identify with an example figure they look up to that either designed or wears that sneaker. The agreement between sneaker brands and celebrities (or any other individual who is, and likes to be, well known) to work together is called celebrity endorsement, and the working is defined by Choi (2017) as the following: *"the image of a celebrity (e.g., talent, success and excellence) is transferred to the* *particular brand or goods. In this process, the product or brand acquires additional value"*. This additional value comes from the role of the celebrity in their own career which gives the celebrity a certain image (McCracken, 1989). Zhenxiang (2020) refers to this as the *star effect*, which will be the term used during the rest of the thesis. The star effect can be explained by the availability heuristic (Choi, 2017). This heuristic describes how people use availability, being the ease with which instances come to mind, to evaluate the frequency or probability of an event (Tversky and Kahneman, 1973). Linking this to celebrity endorsement, in the case of Michael Jordan, his accomplishments as a basketball player are linked to the sneaker he wears during the game instead of to his hard work and practice (Wilson and Sparks, 1996). This makes people want to purchase the sneaker, so they can identify with the famous basketball player, and especially, his skills.

It is proven that any type of endorsement (e.g., collaboration in designing process, advertisement, sponsorship, or usage of the product) by a known individual positively impacts the value perception and popularity of that good (Choi, 2017). When it comes to sneakers, the most occurring forms of celebrity endorsement are collaborations and usage of the product, with the latter one being harder to capture and being less of a fixed price indicator and more of an event based price indicator. Examples are the West and Kardashian families wearing Adidas Yeezy's before and after release, and Michael Jordan wearing a pair of Nike Air Jordan's during a basketball game. Collaborations, however, can be viewed as a fixed price indicator. They come in different forms, with one being multiple signature sneaker releases under a specific line such as the Nike Air Jordan's (collaborated with basketballer Michael Jordan) and the Adidas Yeezy's (collaborated with rapper Kanye West). However, there are also standalone releases in collaboration with, e.g., artists such as Travis Scott (e.g., Nike Air Jordan 1 Travis Scott) and Run-DMC (e.g., Adidas Superstar), food companies such as Ben&Jerry's (e.g., Nike SB Dunk Low Ben&Jerry's Chunky Dunky) and M&M's (e.g., Adidas Forum Low M&M's) and brands such as Dior (e.g., Jordan 1 Retro Dior), Off-White (e.g., Nike Air Jordan 4 Retro Off-White Sail) and Supreme (e.g., Nike Air Force 1 Low Supreme).

2.2.3 Nostalgia

Lastly, it is proven that nostalgia influences the demand (and therefore, the market price) for sneakers as well. As defined by Choi (2017): *"Nostalgia is a sentimental emotion related with the past. In the marketing context, nostalgia is related to the products and brands that have retro designs (e.g., retro branding)"*. With retro branding, companies use elements of designs from the past to give a nostalgic feeling, leading to an increase in the value perception (Brown, Kozinets and Sherry, 2003; Cattaneo and Guerini, 2012). This is also beneficial when it comes to time, energy and money, since the brand does not need a new brand design or marketing campaign (Cattaneo and Guerini, 2012).

This influence of nostalgia on the sneaker market prices is partly captured by the emotional attachment to original models (Choi and Kim, 2019). An explanation for this is the previous inability to fully engage in the sneaker culture in their younger years and being able to currently fulfil those desires (Matthews, Cryer-Coupet and Degirmencioglu, 2021). Nostalgia is also linked to perceived quality. Some people find that products nowadays lack quality while the quality of products from the past was perceived to be better (Thompson, Pollio and Locander, 1994). Lastly, the perceived authenticity of a nostalgic sneaker design proves to be of importance for consumers as well (Belk, 1995). Authenticity refers to something being real and genuine (Choi, 2017) and lots of people are attracted to this concept as a way to distinguish themselves from a larger group, or rather to feel like you belong to a certain smaller group.

Making use of nostalgia is often done in the athletic footwear world (Botterill, 2007; Brown, Kozinets and Sherry, 2003). Examples are the Adidas Superstar from the 1970s which again became a hype in the early 2010s, or Nike Air Jordan recreations with the key characteristics such as original logo, materials and other design details with premium materials (Nike News, 2015).

2.3 Behavioural characteristics

The market price of sneakers is not fixed, it changes over time. One reason for this fluctuating price is that each sneaker has a different demand function because of the different behavioural characteristics of the resellers and consumers that are entering and exiting the market on different moments in time. Where in Section 2.2 the focus mainly lies on the fixed factors of price indication that affect the market price in a constant rate (e.g., predetermined scarcity, collaborations and nostalgia), this section focuses on the behavioural characteristics of resellers and consumers that affect the price development in a variable rate. Even though these indicators cannot be included into the analyses because of the impossibility of collecting them online for each established sale, it is still interesting to discuss their influence on the sneaker resale market to get a better understanding of a how a sale is established and to inform hypotheses about price trends.

2.3.1 Resellers and consumers

Price development is generated by prices that are the result of all separate sales of a particular sneaker over a period of time. A sale is an agreement that occurs when the two parties involved (i.e., the reseller and the consumer) agree on the value of the product and thus, are both willing to sell/purchase the product under those particular circumstances. When it comes to resellers, this value mainly tends to be the monetary value, while for consumers this rather tends to be the hedonic value (Chu, 2013; Higgins, 2006). This is also known as the acquisition phase. To conclude, both parties influence the establishment of each single acquisition.

Since the hyped sneaker market is becoming more accessible to everyone, the type of consumers in the sneaker resale market evolved and can now be divided into sneakerheads (i.e., the core: the collectors and people who wear sneakers with passion) and the less enthusiastic, and rather short-term, sneaker consumers who are starting to join the community at an increasing pace. The latter category refers to people who are not willing to pay the highest price and who are not up to date when sneakers release or even where to purchase them. Those are the people who purchase more sporadically whenever they see a sneaker they like and the price seems good enough.

On the contrary, behaviour of resellers is easier to work out. In their paper, Chu and Liao (2007) divide online resellers into three categories: professional resellers (selling products with the goal of profit maximization), consumer resellers (selling products they previously owned) and mixed-role resellers (being a combination of both). When it comes to the current sneaker resale market, the dominating type is the professional reseller, who is accompanied by a smaller group of consumer resellers in the form of sneakerheads selling their personal collectables. Each type of reseller acts differently based on their motivations, eventually affecting the market in different ways (Chu and Liao, 2007). Professional resellers for example maintain a higher market price than consumer resellers because of the effort that goes into reselling (Chu, 2013; Chu and Liao, 2007). One reason for this required effort is because of the increasing competition when it comes to acquiring sneakers at all (Choi, 2017), making it harder for resellers to get their hands on the profitable sneakers. As discussed before, professional resellers' awareness of future online resale potential can influence their decisions to invest in a sneaker, since they seek successful sales transactions and do not want to risk holding excessive inventory (Chu and Liao, 2007; Chu and Liao, 2010).

2.3.2 Risk and time preferences

The amount of risk resellers prefer when investing in a sneaker depends on the level of risk aversion, which differs over individuals. Risk aversion refers to the tendency of an individual to prefer certainty over uncertainty (Corporate Finance Institute, n.d.), consisting of levels ranging from risk averse to risk seeking. Risk aversion is richly studied and can be easily explained when applied to stock markets. Lee, Rosenthal, Veld and Veld-Merkoulova (2015) found that a higher risk aversion is associated with lower stock market expectations and additionally, a lower probability of investing in stocks. Because of the comparability of the financial markets and the sneaker resale market, introducing this finding in the sneaker resale market leads to the assumption that sneaker resellers who are risk averse are probably more likely to have low expectations of the sales and thus, are less likely to invest in sneakers for reselling. Back to finance, when investors do take the step to participate in the stock market, their investments are according to their risk level (Investopedia, 2021b). Basic risk aversion theory explains that risk averse investors often go for the investments that are more certain and give a low profit, while risk seeking investors prefer risky investments with a possible high return. Since these are the default expectations of risk aversion, it is expected for sneaker resellers to exert this behaviour as well. For example, whereas risk averse resellers will mainly invest in sneakers of which the resale performance is known, risk seeking resellers will also invest in new sneakers with unknown performance, e.g., sneakers with extreme designs of which the demand is difficult to predict.

Once resellers invest in sneakers, this concept of risk goes hand in hand with the behavioural characteristic patience, representing their time preference. In one study by Watts (2019), the role of patience on the sneaker resale market is examined. Results proof that resellers who are patient are more likely to set a higher initial ask. Simultaneously, impatient resellers set lower asks in general since they want to establish a sale (and make profit) as soon as possible and thus, settle for an amount less than the maximum amount that is possible. This results in impatient resellers exiting the market earlier, because their products are sold the earliest[4] which leads to having no product left to sell and therefore, they cannot participate in the market anymore. This leaves patient resellers to rule the market after a certain period of time in which the supply has declined.

However, patience plays a role in consumer purchasing behaviour as well. Consumers who are patient are more likely to set a lower initial bid (Watts, 2019). Simultaneously, impatient consumers set higher bids in general since they want to purchase the sneaker as soon as possible and therefore, should be willing to pay more than the rest of the consumers. This leads to impatient consumers exiting the market earlier[2], leaving patient consumers to solely participate in the market after a certain period of time. In this case, with exiting the market is meant that the needs and wants of these consumers are fulfilled on the short term and therefore, they will not have to participate in the market anymore.

These behaviours of the reseller and consumer result in a decrease in number of sales over time since the asks and bids are less likely to be accepted because of the relatively bigger gap between them. When they will be accepted and at what price depends on who proves to be the least patient and thus, most willing to deviate from their initial offer. From this all it can be concluded that the goal of selling and purchasing, together with the different behavioural characteristics per individual, can influence the market price development over time. Therefore, it is of interest to see whether there is a trend in this price development or not.

[4] When assuming that consumers prefer lower prices, based on the Demand Theory.

2.4 Comparison to financial markets

As previously explained in Section 1.3.2, both parts of this thesis are comparable to analyses performed in financial markets, with the first one being a more specific phenomenon such as IPO short term price prediction and the second one being a more general occurring event such as trend analysis. This section discusses both type of analyses in a more specific way to get a good view of both type of analyses. Additionally, since the sneaker resale market and the financial markets are still two separate markets, the differences between them are discussed in the last section.

2.4.1 IPO price prediction

The prediction of the resale market introduction price is comparable to IPO price predictions, being a more concrete event in the financial markets. This refers to predicting the stock price at which shares trade the first day of being public, and their course shortly after. When IPO prices are set too low or too high this will lead to under- or overpricing, however, this is beyond the scope of this thesis. IPO price prediction comes with the difficulty of accurately predicting as many of the companies were not necessarily obliged to report their financial information, in addition to the fact that there is no previous trading history (Investopedia, 2021c). However, the establishment of IPO prices, their development shortly after and whether it is worth investing or not has already been richly analysed in existing literature. This is advantageous since the different types of methodology to predict the resale market introduction price (without historical data) can be inspired from this.

According to Quintana, Luque and Isasi (2005), most empirical analyses regarding IPO short term price prediction are based on linear regressions, often concerning the effect of the variables related to IPO. However, because this method is sensitive to outliers, which are often present in IPO data, it can be inefficient (Baba and Sevil, 2020). With the rise of Data Science, lots of these statistical IPO price predictions have been replaced by machine learning prediction methods which can deal with the issue of outliers and thus might perform better in this area. Additionally, researchers such as Quintana, Sáez and Isasi (2017) compared the prediction performance of several machine learning methods, with random forest being the best performing one. As confirmation of improvement, some researchers also compared the performance of statistical methods to machine learning methods when it comes to different kinds of IPO price predictions. For example, Esfahanipour, Goodarzi and Jahanbin (2015) compared a fuzzy regression to neural networks, with the latter one performing better, and Baba and Sevil (2020) tested the performance of a random forest against several robust regressions, with random forest outperforming the other methods. These types of studies show that overall, based on prediction accuracy, machine learning methods perform better than the (formerly default) statistical methods when it comes to IPO price predictions. Since the prediction of resale market introduction prices is comparable but not the same, it is of added value to compare a both methods in this thesis as well.

2.4.2 Trend analysis

The average market price trend development can be compared to a general type of analysis, being trend analysis. A trend is *"the overall direction of a market during a specified period of time"* (Investopedia, 2021d). One of the simplest ways of trend analysis is estimating a trend (using historical data) and using it to predict future values. However, trend analysis is also done for the sake of solely evaluating whether a pattern exists in the time-series data representing sales, exchange rates, etc.

When time series are growing by a constant amount or rate, they should be modelled by (transformations of) linear trend models, however, these models often do not capture the true course of the trend. It could for example be that the time series is covariance stationary or that the error terms are serially correlated. In those cases, it is possible to build better forecasting models than linear trend models, such as autoregressive models. However, since this thesis is rather exploratory and the goal is only to confirm whether there exists a trend in the average market price of sneakers in the first ten weeks after release, and not to also accurately forecast the prices after these ten weeks, the working of these type of models is not discussed any further.

2.4.3 Differences between the sneaker resale and financial markets

Since this thesis bases its analyses on the fact that the sneaker resale market and the financial markets are comparable, it is important to also discuss where both markets differ to see whether this could introduce problems in the methodology of choice. Starting off with focussing on the first part of the analyses, which is IPO inspired. One main difference that distinguishes the concept of IPO analysis on the sneaker resale market to IPO analysis on stock markets, is the fact that there is a natural moment at which the sneaker resale market starts. Whereas a sneaker's introduction in the resale market is naturally determined by the first person to put the sneaker up for sale, and does not necessarily align with a fixed moment in time (such as the release date in the primary market), IPOs are introduced at a fixed moment in time. Additionally, there are no opening hours for the sneaker resale market since it is not regulated and trading in sneakers can be done 24/7 from the moment the sneaker is firstly introduced to the market. This is also not the case for commodity and stock markets, where opening hours do exist.

Additionally, the product that is traded differs in nature. The essence of shares for example is having ownership in a company which is rather intangible and static, compared to sneakers which can be perceived as tangible, usable products or collectibles. Thereby, the IPO price of shares is driven by numerous factors such as market sentiment, interest rates, profitability and growth prospects of the company. Whereas some of these aspects can also be applicable to prices on the sneaker resale market, the essence of the product still differs significantly. However, since its effect on the market yields the same problem being the difficulty of predicting its market value before actually entering the market, the comparison of the analysis on both markets is justifiable.

Focussing on the second part of the thesis, being the average market price trend development, there are some differences between the two markets as well. In the case of this thesis, the average market price in the first ten weeks after release is analysed. However, in contrast to for example stocks, it could be that it matters how 'old' the sneaker model is, since new ones are released every day. It could be that there exists a decline in market price as the hype moves on to these other newly released sneakers, whereas for stocks this behaviour does not necessarily occur. This however does not make the comparison unjustified, since the concept of trend analysis stays the same. The only remark is that this difference may lead to a prior establishment of a hypothetical trend in the case of the sneaker resale market, where this is less straightforward in financial markets.

2.5 Hypotheses

2.5.1 Resale market introduction price prediction

By answering the first sub question, this thesis does not only give an answer to whether it is possible to predict the resale market introduction price of sneakers, but also whether a statistical method or a machine learning method performs better for these predictions. Both statistical methods and machine learning methods are often used for single price prediction purposes, including IPO price predictions. The difference between the two methods is the fact that machine learning models are trained to obtain the highest accuracy in predicting the data, while statistical models represent a fitted line that best describes the data (Stewart, 2019). In the first case, the emphasis lies on the results and the ability to repeatedly make accurate predictions. In the latter case, the predominant goal is to characterize the significant relationships in the data (also known as *statistical inferences*) and make it interpretable. Even though this is often catered for, the main goal here is not necessarily to make predictions.

With both types of methods having a different focus, they have proven to be of importance in achieving their own goal. As discussed in Section 2.4.1, both type of methods have proven to work for IPO price prediction and are still used. In the case of resale market introduction price prediction it is therefore interesting to also compare the two types of methods. This will not be done with the end goal of dishonouring one of the two, but rather to investigate what can be learned about the market introduction price as a whole, with machine learning methods focussing more on prediction performance and statistical methods focussing more on identifying relationships between variables. Researchers have found that, overall, machine learning methods are better in accurately predicting the IPO prices in stock markets than statistical methods. Based on this, the following hypothesis has been formed:

Hypothesis 1: The machine learning method outperforms the statistical method in predicting the resale market introduction price.

2.5.2 Average market price trend development

For the second sub question, this thesis focuses on the average market price trend development of sneakers on the resale market. People that are participating in the sneaker resale industry might have noticed a few general patterns in the demand and the accompanying market price trend of sneakers. For instance, there often is a big hype about the sneaker around the release date, with blog and forum posts just before the release discussing the purchase intentions and thoughts about materials and designs (the pre-acquisition phase), followed by the hysteria during the purchasing phase right after release, and often ending with some sort of validation seeking after purchasing the sneaker (the post-acquisition phase).

Once the sneakers are introduced in the resale market, there are different influences on the price development existing of sneaker characteristics, fixed price indicators (e.g., scarcity on primary market, star effect and nostalgia) and behavioural characteristics (e.g., patience and risk aversion). Eventually, some sneaker models have a higher demand accompanied by a higher value than other models, which is not uncommon in the retail industry, or any other industries. One of the explanations for this could be that the limited edition and more unique sneakers have a higher demand and a possibly lower supply, meaning that the price premium will be higher than for sneakers that are less unique and more accessible.

Knowing that all these price indicators affect the price trend in their own way, makes it interesting to analyse whether something can be concluded about the average market price trend development. After briefly evaluating the market price trends of several sneakers, three main price trends were established, being a high-to-low, a low-to-high and a flat trend. One reason for the existence of the high-to-low trend might be the age influencing the hype of the sneaker. One reason for the existence of the low-to-high trend might be the star effect in the form of a celebrity wearing the sneaker sometime after release. These first two trends dominated, however, within the same type of trend there were differences in timing of when the price changed to high or low. When averaging all these trends, the average market price trend development is expected to be a rather flat line since all differences might be averaged out. This leads to the following hypothesis:

Hypothesis 2: The average market price trend equals a horizontal line.

3. Data

This section covers all data related information, such as the data requirements, the data acquisition process and the descriptive statistics of the collected data, before starting this quantitative-based research.

3.1 Data collection

3.1.1 Data requirements

For the market introduction price prediction, the requirement is to have the average market introduction price of the first week after release, including possible influences on this price (i.e., sneaker characteristics and fixed price indicators). For the average market price trend development, the requirement is to have sales information from a significant period of time starting from the release date, in this case the average resale price per week over a period of 10 weeks. So, for each sneaker in the dataset it is required to collect the characteristics and the average resale price per week for the first ten weeks after the release date. Both requirements can be combined into one dataset, with the type of data being panel data.

3.1.2 Selection criteria

The selection of sneakers is based on the release dates obtained from the SneakerJagers website (https://www.sneakerjagers.com/), which is a Dutch platform existing of a sneaker search engine, release calendar and all other sneaker updates in Europe (SneakerJagers, n.d.). To create a representative set of sneakers and prevent selection bias as well as possible, the sneakers are selected based on their release date. The data collection started on May 27th 2022 and since the data has to include at least 10 weeks of resale data, the most recently released sneakers that are selected are from March 17th 2022. From here onwards, sneakers were selected up until back to the earliest release date of January 1st 2021. This leads to the dataset only consisting of sneakers that were released in this time frame of approximately fifteen months.

For selection criteria regarding the sneakers, only sneakers from the brands Nike and Adidas are included, since these two brands are the two biggest brands (Statista, 2022a) that also have had the most releases in the previously discussed time period. Since there exists a big number of models for both brands, only the models that occurred more than fifteen times in the given release period were included. The reason for this is that the data needs to have enough observations per category to be able to draw significant conclusions, and by including all existing brands and all their sneaker models this is harder to achieve. Lastly, only sneakers for men and women are included since they are most likely to purchase the sneakers for their own use – in comparison to kids and infant sneakers being purchased by adults – and thus, these purchases will best reflect the individuals behavioural characteristics.

3.1.3 Data source

All data regarding the market price trends of sneakers needed to answer both sub questions is manually obtained from the StockX website (https://stockx.com/), which is a widely known online resell platform that also posts information about models and other sneaker related news. Via this website, consumers and resellers can buy and sell authenticated sneakers, streetwear, electronics, collectibles, trading cards and accessories (StockX, n.d.). This platform distinguishes itself by only selling "deadstock" products, thus being new, unworn and in the original packaging (Watts, 2019). The previously made comparison of these sneakers to stocks is not random, even StockX calls itself the "first stock market for things" (Watts, 2019). One reason for this regards to the crucial aspect needed to start an exchange which is that goods on the market must be standardized, i.e., in the case of sneakers all being deadstock. This is something that is comparable to the stock market, and it distinguishes StockX from other trading sites such as eBay and Marktplaats. Additionally, all bids and asks are displayed on the website and once both parties agree, the transaction takes place automatically (without any agents), allowing for another possible comparison of this selling process to that of the stock market's.

3.2 Variables

This section discusses the required and used variables in this thesis, based on findings in previous literature. Figure 1 below visualizes what variables seem to have a relationship with the market price of sneakers on the resale market. The two solid arrows show the possible relationship of the included variables and the sneaker resale market price that this thesis will analyse, and the dotted arrow shows the possible relationship of the variables and the sneaker resale market price that this thesis cannot analyse but which have proven to be of influence in previous literature.

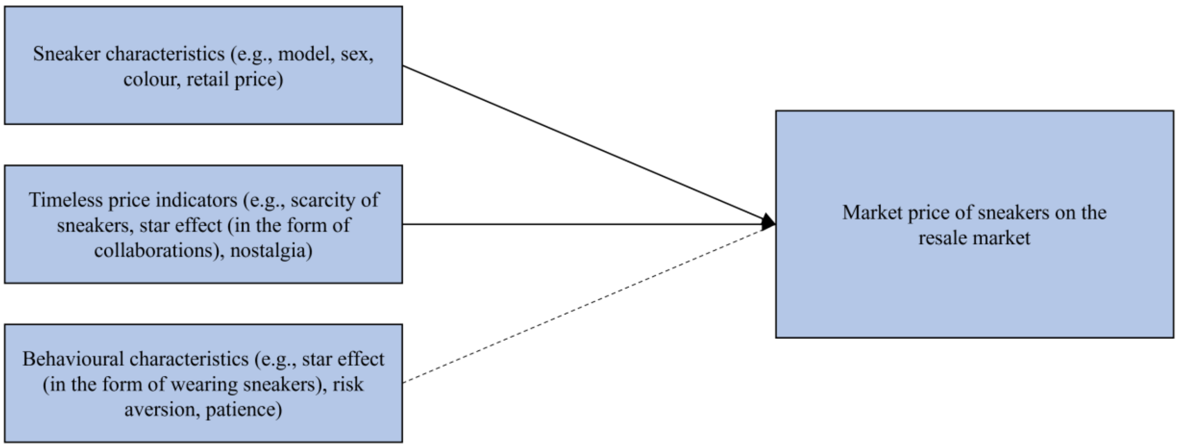


Figure 1. Conceptual model visualizing possible price indicators affecting the market price of sneakers on the resale market.

3.2.1 Sneaker characteristics and fixed price indicators

The first sub question regarding whether a statistical method or machine learning method performs better in predicting the resale market introduction price, is solely answered based on the sneaker characteristics and other external influences on purchasing behaviour discussed in the literature review, i.e., the fixed price indicators, since it is impossible to capture the behavioural characteristics for each established sale retrieved from the StockX website. The required variables that are included are the sneaker characteristics model, sex, colour and retail price, with the latter one representing the internal reference price to which consumers compare the observed price to (Kalyanaram and Winer, 1995) and which proves to affect the purchase intention (Chu and Liao, 2010). Since brand and model are dependent of each other, due to each brand releasing their own models, only one of the two is included in the analysis. In this case, model is chosen since it is more interesting to see whether there are differences between the models (also within one brand) than between the two included brands. Since the variable representing sex had to be created manually, an explanation follows here. When it comes to Nike sneakers, all sneakers that have “WMNS” or “(W)” in their name are perceived as sneakers for women. All sneakers that originally go from size M5.5 and up are perceived to be sneakers for men, and all sneaker that originally go from size M3.5 and up are perceived to be unisex sneakers. When it comes to Adidas sneakers, all Yeezy’s are perceived to be unisex sneakers.

Furthermore, the fixed price indicators proven by previous literature to have an effect, are whether the sneaker is scarce, whether it is linked to an example figure (i.e., the star effect) or whether it is a replica of an original model (introducing nostalgia). Because of the lack of information, the scarcity of each sneaker is too difficult to allocate and therefore, eventually cannot be included in the analysis. However, the other two fixed price indicators are manually captured in two additional variables. The variable representing the possibility of nostalgic feelings is created by analysing whether the sneaker model represents an original/retro model that could awake these feelings. Therefore, nostalgia is allocated in such a way that when the sneaker name includes “Retro”, “Vintage” or “Recon”, the sneaker is a re-release of an older sneaker and it is assumed that it introduces the feeling of nostalgia. Star effect is allocated to sneakers that were made in collaboration with a celebrity. This is the case for all Air Jordans, that are made in collaboration with Michael Jordan, and all Yeezy’s, that are made in collaboration with Kanye West. However, this also includes single collaboration with celebrities such as Serena Williams, Skepta and Travis Scott. Additionally, a variable representing whether the sneaker is a collaboration of two brands is also created. This represents collaborations with brands such as UNO, Kaws and Swarovski, but also apparel brands such as Supreme, Patta and Off-White.

As previously discussed, and alike the scarcity of the sneaker, the behavioural characteristics cannot be included in the dataset to test the effect on the resale market introduction price. Table 1 below displays all the between-subject variables included in the dataset, including the variable type and which values each variable could take on.

Table 1. Sneaker characteristic and other fixed price indicating variables with description.

Variable name	Type of variable	Values
Model	Categorical variable	Air Force, Air Jordan, Air Max, Blazer, Dunk, Waffle, Yeezy
Sex	Categorical variable	Men, Women, Unisex
Colour	Categorical variable	White, Black, Neutral*, Grey, Pink, Blue, Yellow, Green, Red, Orange, Purple, Special design**, GSM***
Retail price	Numeric variable (integer)	[80, 450]
Star effect	Dummy variable	0, 1
Brand collaboration	Dummy variable	0, 1
Nostalgia	Dummy variable	0, 1

Note. *neutral colour includes all colours in the brown colour scheme (e.g. sand, beige, copper, chestnut, chocolate, mocha); **special design includes sneakers that have a print, translucent, reflective or shiny parts, fur or charms; *** GSM stands for gold, silver or metallic. Integer means that the number does not include decimals.

3.2.2 Price information

When it comes to price variables used for the resale market introduction price prediction and average market price trend development, the data includes the average resale price per week for each sneaker for a period of the first ten weeks after the release date. Since all sneakers have different release dates but also a different velocity in which they are resold, leading to the StockX-data having different paces between the given prices on their website, this thesis cannot make use of calendar dates and instead the data needs to be divided into time periods after release. Based on this information, the time periods are constructed on a weekly basis, being the first week after release, the second week after release, etc., for the first ten weeks after release. To end up with a constant number of prices, the average price per week is taken per sneaker. The average price in the first week is used for both the resale market introduction price prediction and the average market price trend development analysis. The averages prices of week two to ten are solely used for the average market price trend development analysis. Table 2 below displays the within-subject variables included in the dataset, including the variable type and which values each variable could take on.

Table 2. Average market price variables with description.

Variable name	Type of variable	Values
P1	Numeric variable (integer)	[75; 3,022]
P2	Numeric variable (integer)	[24; 2,983]
P3	Numeric variable (integer)	[58; 2,991]
P4	Numeric variable (integer)	[84; 3,144]
P5	Numeric variable (integer)	[70; 3,220]
P6	Numeric variable (integer)	[71; 3,257]
P7	Numeric variable (integer)	[64; 3,486]
P8	Numeric variable (integer)	[53; 3,360]
P9	Numeric variable (integer)	[72; 3,391]
P10	Numeric variable (integer)	[63; 3,147]

Note. Integer means that the number does not include decimals. Furthermore, P denotes average market price and therefore, P1 refers to the average market price in period 1, P2 refers to the average market price in period 2, etc.

3.3 Descriptive statistics

The dataset has a sample size of 638 observations, each representing a single sneaker that is released within the period of which the data is collected. For now it is chosen to leave the outliers in since it might be of interest to see whether some characteristics of those sneakers lead to a higher resale market introduction price, however, this might change once it is proven that the outliers bias the models. Table 3 shows the descriptive statistics for the variables that represent the sneaker characteristics and fixed price indicators, and some of the remarks are discussed below.

As can be concluded from the variable *Model*, Nike has released substantially more sneakers over the period of fifteen months of which the data is collected (91.22% vs. 8.78% of sneakers released by Adidas represented by the model "Yeezy"). Also, only a small subset of the sneakers (5.96%) evokes nostalgic feelings. The rest of the sneakers seem to be new (or modern) releases. Larger subsets of the sneakers are in collaboration with a celebrity (41.54%) or a brand (8.93%). Furthermore, the largest part of the sneakers are perceived to be unisex (54.70%). However, in real life, sneakers that are perceived to be women (or men) are not limited to women (or men) only, eventually leading to even more sneakers being perceived to be unisex. Reviewing colours, white is the colour that occurs most (77.90%) on the sneakers in the dataset. One reason for this is the fact that the default colour of the sole of sneakers is white. The second and third most occurring colours are black and neutral. From this can be concluded that over the period of fifteen months of which the data is collected, a big part of the releases can be perceived to have more natural colourways than bright ones.

Table 3. Descriptive statistics of variables representing the sneaker characteristics and fixed price indicators.

Variables	**0 (No)**	**1 (Yes)**
Model		
Air Force	547 (85.74%)	91 (14.26%)
Air Jordan	433 (67.87%)	205 (32.13%)
Air Max	542 (84.95%)	96 (15.05%)
Blazer	606 (94.98%)	32 (5.02%)
Dunk	497 (77.90%)	141 (22.10%)
Waffle	621 (97.34%)	17 (2.66%)
Yeezy	582 (91.22%)	56 (8.78%)
Star effect	373 (58.46%)	265 (41.54%)
Brand collaboration	581 (91.07%)	57 (8.93%)
Nostalgia	600 (94.04%)	38 (5.96%)
Sex		
Men	543 (85.11%)	95 (14.89%)
Women	444 (69.60%)	194 (30.41%)
Unisex	289 (45.30%)	349 (54.70%)
Colour		
White	141 (22.10%)	497 (77.90%)
Black	323 (50.63%)	315 (49.37%)
Neutral*	400 (62.70%)	238 (37.30%)
Grey	485 (76.02%)	153 (23.98%)
Pink	557 (87.30%)	81 (12.70%)
Blue	465 (72.88%)	173 (27.12%)
Yellow	563 (88.24%)	75 (11.76%)
Green	508 (79.62%)	130 (20.38%)
Red	485 (76.02%)	153 (23.98%)
Orange	553 (86.68%)	85 (13.32%)
Purple	585 (91.69%)	53 (8.31%)
Special design**	524 (82.13%)	114 (17.87%)
GSM***	573 (89.81%)	65 (10.19%)

Note. N = 638. *neutral colour includes all colours in the brown colour scheme (e.g. sand, beige, copper, chestnut, chocolate, mocha); **special design includes sneakers that have a print, translucent, reflective or shiny parts, fur or charms; *** GSM stands for gold, silver or metallic. The total share of colours exceeds 100% since each sneaker can have more than one colour.

Table 4 shows the descriptive statistics for the variables that represent the release date, retail price and average market prices of the first ten weeks after release. As can be seen, only sneakers released in the period from January 8, 2021 until March 23, 2022 are included. When it comes to prices, the retail price ranges from $80 to $450. However, the market prices range from €24 to €3,486, showing that there can be substantially large differences between a sneaker's retail price and market price (over time). In the next section regarding the model free evidence, the prices are discussed further.

Table 4. Descriptive statistics of variables representing the release date, retail price and average market prices of the first ten weeks after release.

Variables	**Min**	**1st quantile**	**Median**	**Mean**	**3rd quantile**	**Max**
Release date	08-01-2021	06-05-2021	12-08-2021	16-08-2021	30-11-2021	23-03-2022
Retail price (in $)	80	110	125	142.2	170	450
P1 (in €)	86	140	174.5	203.2	223.8	3,022
P2 (in €)	24	137	167.5	198.3	214.8	2,983
P3 (in €)	58	134	165	198.2	213	2,991
P4 (in €)	84	132	163	197.2	212.8	3,144
P5 (in €)	70	131	163	197	214	3,220
P6 (in €)	71	130	163	197	214.8	3,257
P7 (in €)	64	129	161	196.3	214.8	3,486
P8 (in €)	53	126	160	194.8	213.8	3,360
P9 (in €)	72	127.2	160	195.7	218	3,391
P10 (in €)	63	126	158	193.8	215	3,147

Note. N = 638.

3.4 Model free evidence

Since the focus of this thesis lies on the different prices of sneakers in the resale market, this section discusses the price distributions before further analyses to get a better understanding of pricing decisions and development. As can be seen in Figure 2, the retail price of the sneakers included in the dataset is not normally distributed, but rather skewed to the right. This means that the retail prices are mainly between the $80 and $250. Besides this, there are only two sneakers included in the dataset with a retail price of $300. After the retail price of $300 there only exist four observations in the dataset that have a maximum retail price of $450, which are collaborations with the luxury brands Swarovski and Alyx.

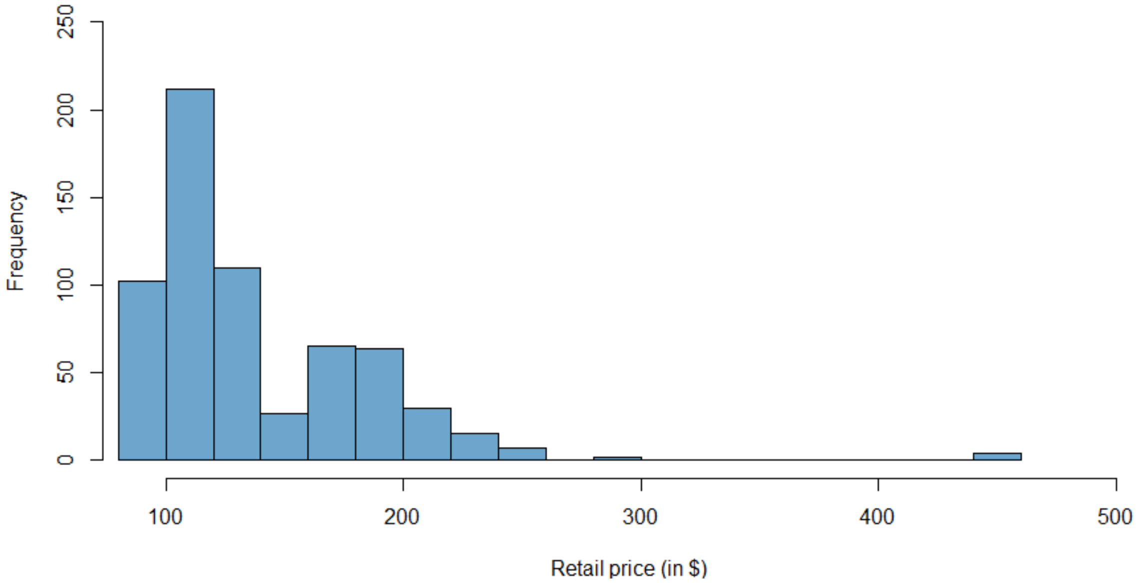
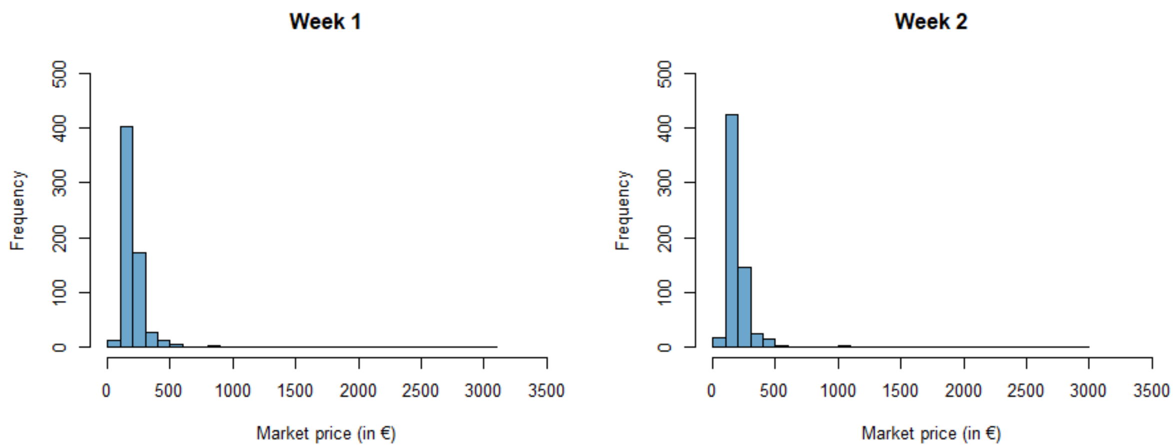


Figure 2. Distribution of retail price (in $).

When analysing the market prices in the first ten weeks after release, some sort of a pattern already occurs (see Figure 3 below). Again, all distributions are skewed to the right, meaning that the market prices are more balanced on the lower side of the price range. For all the weeks except week two and three, the resale market price ranges to a maximum of somewhere between €3,000 and €3,500. For weeks two and three, the resale market price only ranges to a maximum of somewhere between €2,500 and €3,000. These findings correspond to the numbers displayed in Table 4 in the previous section. When zooming in a bit more, it can be seen that for each week, the vast majority of the market price lies between €100 and €200 followed by market prices between €200 and €300, all with approximately the same frequency. The rest of the observations fall in the range of €300 to €500, with only a few ranging to the maximum of €3,500. Remarkably, when the weeks increase, the number of sneakers with a market price below €100 increases, suggesting that market prices might decline over time.



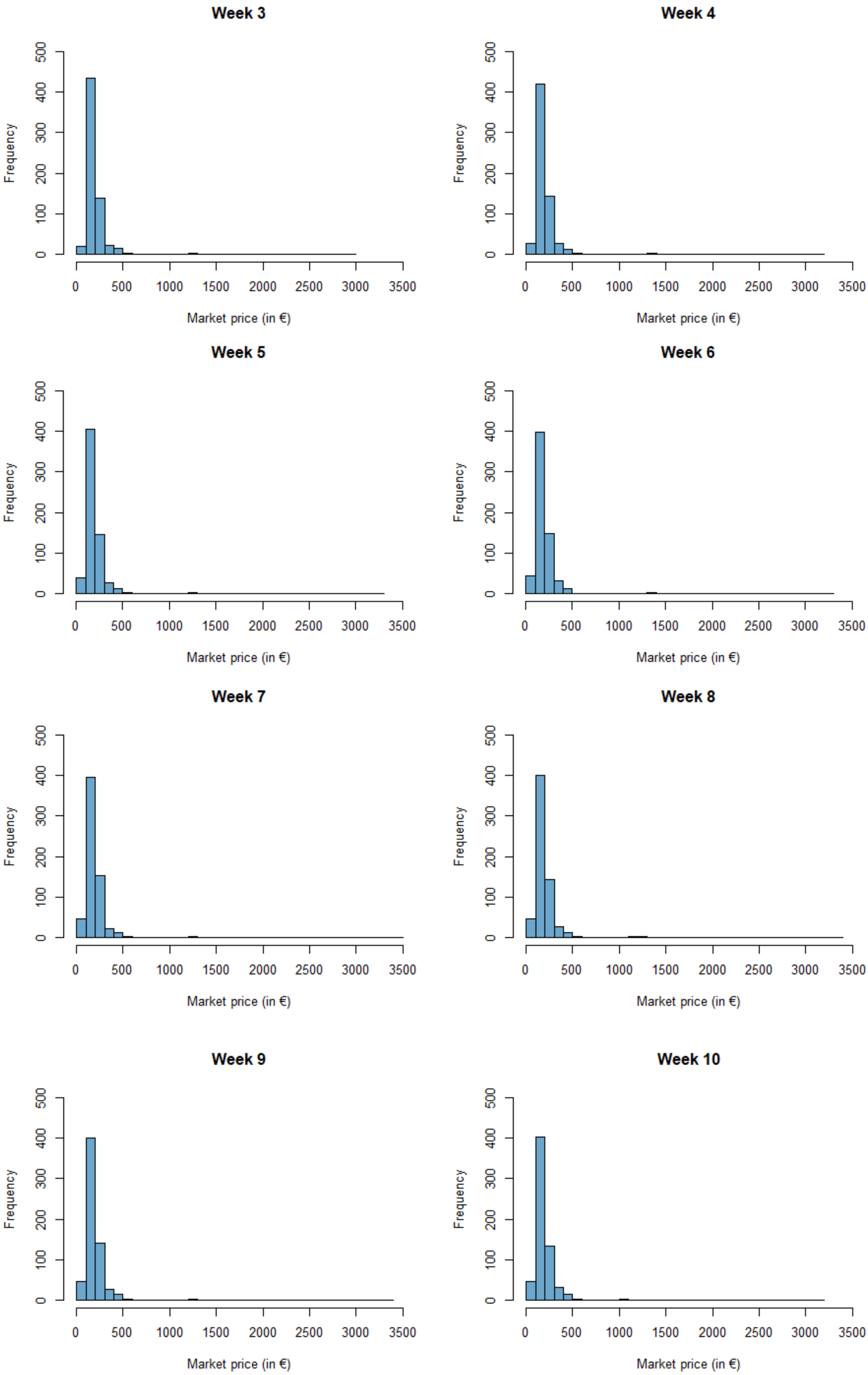


Figure 3. Distribution of average market price (in €) for each of the first ten weeks after release.

4. Methodology

This section is divided into two main parts, with the first one being the methodology of predicting the resale market introduction price. This part exists of an explanation of the two chosen models and how to best evaluate their performances so that it can be decided upon which model performs best in this analysis. The second part focusses on the methodology of evaluating the average market price trend development, and more specifically the comparison of this trend to a horizontal line.

4.1 Resale market introduction price prediction

Since no research about price prediction of sneaker market prices exists, this thesis will make use of the methods that have proven to best perform in comparable situations such as the introduction into the stock market. Previous literature has shown that both statistical methods and machine learning methods can be used for comparable IPO price predictions. Here, the most used statistical method for single price predictions (lacking historical data) is the multiple linear regression, and the best performing machine learning method has proven to be the random forest. Even though the sneaker resale market is less liquid than the stock market (Watts, 2019), there is no need to deviate from these models and design a new methodological approach, since this thesis will provide the first insights in this area.

Since the random forest model needs to be trained, it is of importance to split the dataset into a training and testing set. Normally, this is not needed for the multiple linear regression, however, for comparability purposes it is of importance that the models are created based on the same data. Therefore, the data is split into a training and testing set (with a ratio of 75:25) beforehand and both models are created using the training set. The performance of both models is evaluated using the testing set and is compared to find the highest accuracy with which single price predictions can be made in the case of resale market introduction price of sneakers.

4.1.1 Multiple linear regression

A linear regression estimates the value of the dependent variable by means of the value of the independent variable. A multiple linear regression simply refers to a linear regression that has multiple independent variables that affect the dependent variable. Stewart (2019) explains that a linear regression is a form of regression analysis that models a line that minimizes the mean squared error (MSE) for all data points, also known as the least squares approach. This means that a straight line is fitted in such a way that the discrepancies between the predicted and actual data points are minimized (IBM, n.d.). A standard multiple linear regression looks as follows:

$$(1) \qquad Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

With Y representing the dependent variable, α representing the intercept, X_1 until X_n representing the independent variables, β_1 until β_n representing the effect sizes of the corresponding independent variable, and ε representing the error term. Applying this multiple linear regression model to the data collected for this thesis, looks as follows:

$$(2) \qquad \textit{Resale market introduction price} = \alpha + \beta_1 \textit{ Model} + \beta_2 \textit{ Star effect} + \beta_3 \textit{ Brand Collaboration} + \beta_4 \textit{ Nostalgia} + \beta_5 \textit{ Sex} + \beta_6 \textit{ Colour} + \beta_7 \textit{ Retail price} + \varepsilon$$

One of the goals of multiple linear regression can be prediction, which also applies in this thesis. In that case the model is fitted on observed data to estimate the effect of each independent variable on the dependent variable ($\beta_1, \ldots, \beta_n$). The newly collected values of the independent variables ($X_1, \ldots, X_n$) can be filled in the model resulting in a newly estimated value of the dependent variable ($\hat{Y}$).

Many methods come with assumptions regarding the data and the relationships between the variables, that should be met. When this is not feasible, the model can reflect the data in an inaccurate way which leads to inaccurate predictions (Meinert, Patel and Li, 2019). These are the assumptions that should be met for a multiple linear regression model (Wooldridge, 2015):

1. Linearity (i.e., there needs to be a linear relationship between the independent and dependent variables).
2. Random sampling (i.e., every subject in the target population must have an equal chance of being included in the sample).
3. No perfect collinearity (i.e., independent variables should not be highly correlated to each other).
4. Zero conditional mean, or exogeneity, (i.e., the expected value of the error term is 0: $E(\varepsilon) = 0$, meaning that there is independency between the independent variables and the error term needed to find a causal effect).
5. Homoskedasticity (i.e., the residuals should be equal across the regression line, or the variance of the error term is the same for all observations regardless of their values).
6. Normality (i.e., error term should follow a normal distribution).

The first four assumptions ensure unbiasedness (and consistency) of the multiple linear regression, meaning that the estimated values equal the population parameters and thus, the model is able to rightly predict values. The last two assumptions ensure statistical inference, and are related to the characteristics of the distribution of the errors (Wooldridge, 2015). Before performing the analysis, it will be tested whether all assumptions are met.

4.1.2 Random forest

Machine learning models can identify patterns and make predictions about data that is unobserved (Murdoch, Singh, Kumbier, Abbasi-Asl and Yu, 2019). These models can learn from their experience (Kwartler, 2017), and this process is better known as *training* the model. There are different types of machine learning methods, however, for this thesis a random forest regression model will be used.

Random forest is an improved decision tree-like prediction model, representing one variable at each split node (being the place where one branch splits into two new ones). The problem of simple decision trees is that they easily overfit because they suffer from a high variance, and this is undesirable when wanting to build an accurate prediction model (James, Witten, Hastie and Tibshirani, 2013). As can be derived from the second part of the name, a random forest is a model that consists of multiple decision trees (i.e., a forest), ultimately decreasing variance and therefore, increasing prediction accuracy (Breiman, 2001). The first part of the name refers to how the sub-trees are made (i.e., randomly), eventually leading to a decrease in correlation between the predictions. There are two different versions of the random forest based on the type of the dependent variable (in this case also known as the predicter variable). There exists a random forest classification and random forest regression model. In this thesis, the latter one will be used. In contrast to the multiple linear regression, there are no assumptions that should hold for a random forest regression model since machine learning models have to be trained on the specific data at hand and therefore, will adapt to the characteristics of the data (to some extent, ideally preventing the model to overfit).

The training set is used to create multiple bootstrapped (i.e., simulated) samples of the same size as the original dataset, where repetition is allowed, with the goal to accurately represent the population. This leads to approximately one-third of the sample not being used in the bootstrapped samples, called the out-of-bag (OOB) samples, which are used to calculate the error rate of the model (James et. al, 2013). The other two-third of the training set (divided into bootstrapped samples) is used to create the decision trees. This is automatically different from the multiple linear regression where the entire training set is used to train the model, however, unfortunately this occurrence cannot be prevented and both models differ on this small aspect. Moving on, because each bootstrapped sample is used for a different tree, it results in as many trees as there are samples. Because there are multiple trees, the interpretability advantage of decision trees no longer holds. Nevertheless, the working is still important and goes as follows (James et. al, 2013): from each bootstrapped sample one decision tree is constructed by choosing one variable out of a random subset sample m of all p predictors/variables (where $m < p$) at each particular split node, representing the point in which the “branch” splits into two new ones. This is done by the means of the Gini Index, representing the variable importance and thus which variable should be placed highest in the tree. The formula of the Gini Index shows the probability of misclassifying an item when selected randomly (Tyagi, 2020), and can be written as follows:

$$(3) \qquad I_G(p) = 1 - \sum_{i=1}^{C} (p_i)^2$$

Where C represents the number of variables, p_i represents the probability of an item being classified to a particular class i.

Once all decision trees in the random forest model are designed by means of all bootstrapped samples, the OOB-samples are run through the model and give their own outcome of the response variable for each tree (which is comparable to the dependent variable in the multiple linear regression). Since this is a random forest regression problem, an average from all outcomes of all decision trees is taken as final prediction value (Breiman, 2001).

4.1.3 Performance evaluation

Since a multiple linear regression and a random forest regression are two different models, comparing the performance is not standard. The performance of a multiple linear regression is measured with the R-squared measure, which represents the proportion of the variance of the dependent variable that is explained by the independent variables (Investopedia, 2021e). Simultaneously, the performance of a random forest regression is also measured by means of variance explained, which seems comparable to the R-squared, since this also represents the variance in the response variable (i.e., dependent variable) explained by the change in predictor variables (i.e., independent variables). However, it is not. The big difference is that for the multiple linear regression, the variance explained (i.e., R-squared) is calculated over the entire training set which is used to build the model. This can be seen as the performance of the fit being measured, representing how well the model fits the data used to build it. Whereas for the random forest regression this variance explained is not calculated over the entire training set, but only over the OOB samples of the training set which are not used to build the model, and thus already representing how well the model can predict using new and unknown data.

Since it is not fair to compare these measures to each other, it is needed to decide upon one universal way to measure performance of both models. The performance measure that will be used in this thesis is the root mean squared error (RMSE), which is calculated using the following formula:

$$(4) \qquad RMSE = \sqrt{\sum(\hat{Y}_i - Y_i)^2 / n}$$

Where $\hat{Y}_i$ is the newly estimated value of the dependent variable for observation i, Y_i is the true value of the dependent variable for observation i, and n is the sample size. The RMSE will be calculated using the testing set and therefore, the predictions (i.e., all $\hat{Y}_i$) can be viewed as out-of-sample predictions.

As explained in Section 4.1.1, the MSE refers to the discrepancies between the predicted and actual data points. The RMSE is simply the square root of this value, representing the square root of the average sum of the squared differences. Compared to the MSE, it transforms the value to the same units as the dependent variable. Therefore, in this case it represents the average deviation between the predicted resale market introduction price and the actual resale market introduction price. As performance evaluation measure, the model that has the lowest RMSE performs best.

4.2 Average market price trend development

The second part of this thesis is regarding the evaluation of the average market price trend development. One way to assess this is by comparing the average market price trend to a horizontal line. The reason behind this approach is that when this linear regressed function does statistically differ from a horizontal line, it proves that there exists a trend in the average market price of the sneakers included in the dataset. If this is the case, it might be of interest for future research to evaluate whether the (average) market price of sneakers can be forecasted over time. This is valuable since it could prove abnormalities in the market, such as arbitrage opportunities.

As discussed in the descriptive statistics (Section 3.3), the data needed for this part exists of the average market price of the first ten weeks after release. To be able to perform this analysis, this data needs to be slightly transformed into a new data frame in the following way: the average market price per period is calculated as one variable and afterwards, the time period (i.e., $T = 1, 2, \ldots, 10$) belonging to that average is allocated as a second variable. This eventually leads to a dataset consisting of 10 observations and 2 variables. Continuing, the analysis requires an approach consisting of two steps. The first step is to construct the average market price trend that should be analysed. This is constructed by taking the average market price of all sneakers per time period and running a linear trend model on this data (quite like the linear regression model displayed in formula (1) in Section 4.1.1), which looks as follows:

$$(4) \quad \textit{Average market price} = \alpha + \beta_1 \textit{ Time period} + \varepsilon$$

With *Average market price* as dependent variable representing the average market price per week (and thus, changing at a constant rate with time), α representing the intercept, *Time period* as the (only) independent variable representing the time period (i.e., number of weeks after release), β_1 representing the slope of the average market price, and therefore the trend, and ε representing the error term.

The second step is evaluating the possible trend. Once the linear trend model is built, the slope (i.e., β_1) is evaluated to see whether the trend is statistically significant. This is done using the p-value of the t-statistic which comes with the following hypotheses: H_0: $\beta_i = 0$ *vs.* H_a: $\beta_i \neq 0$. If the p-value is below the significance level α, the null hypothesis can be rejected and it can be concluded that there is sufficient evidence for a significant relationship between the dependent variable and independent variable and thus, it is likely that an average market price trend exists in the case of the sneakers included in the dataset. If the null hypothesis cannot be rejected, it means that it less likely that there is a relationship between these two variables. In this case it can be concluded that the intercept-only model fits the data better[5] and thus, the average market price does not change over time. If this is the case, the best fitted linear trend model is a horizontal line and therefore, no trend seems to be present in the data.

For a linear trend model, the same assumptions should be met as for the multiple linear regression (explained in Section 4.1.1). Before performing the analysis, it will be tested whether all assumptions are met.

[5] Only when the intercept is proven to be significant.

5. Results

This section discusses the research findings of this thesis[6]. In the first part, two models (i.e., the multiple linear regression and the random forest regression) are compared to see which one performs better in predicting the resale market introduction price. Besides this, it is also discussed which variables have (the biggest of) an influence on this resale market introduction price. The second part focusses on the average market price trend development by analysing whether a significant trend exists in the first ten weeks after a sneaker is released. Because of the limited amount of data and the fact that this thesis is an exploratory research with the goal to find whether it is possible to draw any conclusions at all about the prices and its development (compared to other studies focussing on more specific details of a certain event), this thesis uses a minimum significance level α of 10%. This might lead to a broader understanding of prices and its development than only establishing results that are significant on a 5% significance level. However, for each statistically significant conclusion, the corresponding p-value is noted to give a better overview of which conclusions are more significant than others.

5.1 Resale market introduction price prediction

In this section, the resale market introduction price prediction is discussed. As explained before, this exists of a comparison of two models, one being a statistical method and the other being a machine learning method. Even though both methods have a slightly different end goal (i.e., making statistical inferences vs. repeatedly making accurate predictions), they are both used in the field of IPO price predictions, which is assumed comparable to the single price prediction that is performed in this thesis. Analysing both type of methods does not only show which one performs better, but also leads to better insights into the concept of sneaker resale market introduction price. First, the results and performance of the multiple linear regression (i.e., the statistical method) are discussed, using the training set. This is followed by the results and performance of the random forest regression (i.e., the machine learning method), using the training set as well. To conclude this section, the two models are compared to see which one performs best in the case of resale market introduction price prediction of sneakers, using the testing set.

5.1.1 Multiple linear regression

To perform the multiple linear regression in R, the lm()-function from the stats package (R Core Team, 2020) is used. The multiple linear regression was ran as displayed in formula (2) in Section 4.1.1. The total number of observations in the dataset exceeds the minimum required sample size of 233 observations for this analysis[7], of which the calculation itself is discussed in more detail in Appendix A.

[6] The analyses are performed in R, version 4.0.2.
[7] This number is calculated using the G*Power application, version 3.1.9.7.

All assumptions for the multiple linear regression, explained in Section 4.1.1, are tested and the corresponding results are extensively discussed in Appendix B. These tests show that only two of the six assumptions hold, being the linearity and no perfect collinearity assumptions. As regards to the random sampling and zero conditional mean assumptions, it is impossible to improve the model with the available variables and information at hand[8] and therefore, solutions to this are discussed in the Limitations. A solution for heteroskedasticity (as assumed here since the homoskedasticity assumption does not hold) is running the regression with robust standard errors, since the standard errors calculated for the initial model are incorrect and therefore, conclusions about significance are incorrect as well. Doing this has improved significance of some of the variables in the model. To improve normality, it is possible to remove outliers from the dataset[9] (e.g., by using the IQR[10]) and to transform the data such that numeric variables follow an approximate normal distribution (e.g., transforming the dependent variable to take the natural log of it, leading to a log-linear model)[11]. Following these steps has resulted in removing 37 observations in total and impacted the performance of the model, as shown in Table 5.

Table 5. Comparison of initial model and final model after improvements based on the assumption tests.

	Initial model	**Final model**
R-Squared	0.1663	0.4854
Number of significant variables	7	9
Number of training observations used to create the model	480	453

Note. Significant variables are statistically significant and different from zero with at least a significance level of 10%.

The results of the final model solely using the training set are given in Table 6 below. Since some variables are categorical variables (i.e., *Model*, *Sex* and *Colour*) and the data needs to adhere to at least one of the categories, they require a reference category that should not be included as a variable in the model. This reference category is chosen based on the highest number of observations, which for *Model* is "Air Jordan", for *Sex* is "Unisex" and for *Colour* is "White". *Star effect*, *Brand collaboration* and *Nostalgia* are dummy variables and do not require a reference category due to the fact that these are optional for sneakers and can take on values of either zero or one. To conclude, based on the reference

[8] When assuming the reason for rejecting the zero conditional mean assumption is that the model also estimates reverse causality, OVB, etc.
[9] Even though these observations do give interesting insights into what causes such a high resale market introduction price compared to other sneakers.
[10] The Interquartile Range (IQR) is often used to find outliers in a dataset. In this case, an observation is viewed as an outlier whenever its value is below Q1 − 1.5*IQR or above Q3 + 1.5*IQR.
[11] Which might also improve the zero conditional mean assumption when assuming the reason for the rejection is that the model misses nonlinearities.

categories chosen, the base sneaker in this model is a white, unisex Air Jordan that has no brand or celebrity collaboration and does not evoke nostalgic feelings and has a retail price of $0. Since the multiple linear regression is a log-linear model, the values of the coefficients should be transformed to be able to interpret the results. This is done by the following rule: a one unit increase in the independent variable (X_i) multiplies the expected value of the dependent variable by $e^{\beta i}$ (Benoit, 2011). In this case, the intercept which represents the magnitude of the resale market introduction price for the base sneaker, is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.01), and results in a resale market introduction price of $e^{4.7106}$ which is equal to €111.12 for the base sneaker.

Now continuing to the other significant insights. When it comes to *Model*, the only significant coefficients are the ones of the category "Air Force", "Blazer" and "Dunk". The models "Air Force" and "Blazer" have a coefficient of -0.1427 and -0.2218, respectively, and are statistically significant and different from zero on a 10% significance level (both with a p-value lower than 0.01). From this can be concluded that when the sneaker of interest belongs to one of the two models "Air Force" or "Blazer", using the same rule as in the previous section, they have a resale market introduction price that is 13.30% and 19.89% lower than an Air Jordan that has the exact same values for all the other variables in the model (i.e., ceteris paribus), respectively. When it comes to the model "Dunk", it has a coefficient of 0.1206 which is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.1). From this can be concluded that when the sneaker of interest is a Dunk, it has a resale market introduction price that is 12.82% higher than an Air Jordan, ceteris paribus. For the other sneaker models, it cannot be concluded with certainty that it influences the resale market introduction price compared to the sneaker being an Air Jordan.

Additionally, the fixed price indicator *Star effect* shows to have no significant effect on the resale market introduction price. However, the fixed price indicators *Brand collaboration* and *Nostalgia* do have a significant effects on the resale market introduction price. The coefficient of the variable representing collaboration with a brand has a magnitude of 0.1717 and is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.05). This means that whenever a sneaker is collaborated with another brand, the resale market introduction price increases with 18.73% compared to sneakers that are not collaborated with other brands, ceteris paribus. The coefficient of the variable representing evoking nostalgic feelings has a magnitude of 0.0811 and is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.1). This means that whenever a sneaker evokes nostalgic feelings, the resale market introduction price increases with 8.45% compared to sneakers that do not evoke nostalgic feelings, ceteris paribus.

When it comes to *Sex*, a sneaker that is perceived to be for women has a coefficient of -0.0727 which is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.05). This means that when a sneaker is perceived to be women's, it will have a 7.01% lower resale market introduction price compared to when a sneaker is perceived to be unisex, ceteris paribus. There seems to be no significant effect of a sneaker being perceived to be men compared to unisex. Furthermore, only one colour proves to have a significant effect on the resale market introduction price, which is the colour black. "Black" has a coefficient of -0.0435 which is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.1). This means that whenever a sneaker contains a black colour, the resale market introduction price will be 4.26% lower compared to a sneaker not containing black, ceteris paribus. All the other colours seem to not have a significant effect on the resale market introduction price. Lastly, the retail price seems to have a significant effect on the resale market introduction price as well. This variable has a coefficient of 0.0035 which is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.01). This means that for every dollar that the retail price increases, the resale market introduction price increases with 0.35%, ceteris paribus. This means that for each additional dollar in retail price there seems to be a possibility to make an average gross profit of 0.35%[12] when the sneaker is sold in the first week of being on the resale market, ceteris paribus.

Now that the insights are discussed, it is also of importance to analyse the performance of the model. As explained before, both models (i.e., the multiple linear regression and the random forest regression) have their own performance measure. In this part the default performance measure used for multiple linear regressions is used for model evaluation, which is the R-squared. The R-squared represents the proportion of the variance of the dependent variable that is explained by the independent variables (Investopedia, 2021e), and ranges from 0 to 1. In this case, the R-squared is 0.4854, meaning that when only including the sneaker characteristics and fixed price indicators in the model, 48.54% of the variance of the dependent variable is explained by these independent variables. This can be perceived as approximately half of the variance, meaning that the model still misses a lot of important information that could explain the resale market introduction price as well.

To conclude, the best performing model chosen to use as the multiple linear regression in this thesis has room for improvements. This could be done by adding important information beyond the sneaker characteristics and fixed price indicators that influence the resale market introduction price, such as the behavioural characteristics explained in Section 2.3. These or other missing information should be collected in different ways as currently done. Besides this, the model does give some insights into what existing variables are of importance when predicting the resale market introduction price[13], being the *Model, Brand collaboration, Nostalgia, Sex, Colour* and *Retail price*.

[12] Leaving additionally made costs such as shipment and import costs out.
[13] Based on the fact that there exists at least an association between these variables and the dependent variable.

Table 6. Results of the optimal multiple linear regression model using the training set and the natural logarithm of the resale market introduction price as dependent variable.

Variable	**Coefficient**
Intercept	4.7106***
Model	
Air Force	-0.1427***
Air Max	-0.0235
Blazer	-0.2218***
Dunk	0.1206*
Waffle	-0.0992
Yeezy	0.0526
Star effect	0.0225
Brand collaboration	0.1717***
Nostalgia	0.0811*
Sex	
Men	-0.0208
Women	-0.0727**
Colour	
Black	-0.0435*
Neutral	-0.0130
Grey	0.0033
Pink	0.0014
Blue	-0.0133
Yellow	-0.0441
Green	0.0006
Red	-0.0369
Orange	-0.0125
Purple	0.0061
Special design	0.0208
GSM	0.0085
Retail price (in $)	0.0035***

Note. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01. N = 453; Adjusted R-squared = 0.4854.

5.1.2 Random forest regression

To perform the random forest regression in R, the randomForest()-function from the randomForest package (Liaw and Wiener, 2002) is used. For this analysis, no minimum required sample size can or needs to be calculated. Even though machine learning models can deal better with outliers than statistical models, removing the outliers greatly improved the performance of this model as well (almost twice as good). Because of this reason and since both models should be built using the exact same dataset, outliers (found by means of the IQR range) are also not included here. For this method, no assumptions need to be tested. However, this rather advanced model requires the specification of several hyperparameters in its algorithm. The most important hyperparameters that need to be tuned in order to increase the predictive power are the sample size of the subset *m*, the number of trees and the three depth (Scornet, 2018). The variance explained calculated over the model, which is obtained from the OOB sample, is used as evaluation metric to find the hyperparameters that support the best performing model.

If the sample size of subset *m* is not defined, the default value is $\sqrt{p}$, however, it is likely that this is not the optimal number and therefore, tuning is required. Tuning refers to *"the process of maximizing a model's performance without overfitting or creating too high of a variance"* (Riskspan, 2017). The default random forest model on the current dataset uses a subset *m* of 6 variables tried at each node and 500 trees. The tree depth is captured in two hyperparameters, one representing the maximum number of nodes the trees can have and the other representing minimum number of observations that should be included in each node. For these hyperparameters, the default value is to grow the trees to the possible maximum and requiring a minimum of 5 observations in each node (Liaw and Wiener, 2002). Manually tuning (by means of trial and error) proves that the optimal values for this model are a subset *m* of 6 variables tried at each node, 400 trees, maximum number of 50 nodes and a minimum number of 5 observations that should be included in each node.

As discussed before, the random forest regression uses variance explained as performance measure to evaluate the model. Here, the variance explained is 47.57%, compared to the default model being able to explain 46.05% of the variance. This means that (based on the OOB samples) this final model can explain 47.57% of variance in the response variable based on change in the predictor variables. Thus, again approximately half of the variance can be explained, meaning that this model, solely including sneaker characteristics and fixed price indicators, misses important information that could explain the resale market introduction price as well, e.g., the behavioural characteristics mentioned in Section 2.3 or any other variables that are not included. Table 7 below shows the optimal model's hyperparameters and its performance in a clear manner.

Table 7. Hyperparameters and performance measurement of the final model using OOB samples of the training set and the resale market introduction price as the response variable.

	Subset *m*	**Number of trees**	**Maximum number of nodes**	**Minimum number of observations at each node**	**Variance explained**
Optimal model	7	400	75	5	47.57%

Since random forest models are black box models, it is infeasible to fully understand the decision processes, in this case the construction of all the decision trees. However, there are ways to approximately visualize the decision processes that have taken place while creating a random forest model and especially, which variables are most important based on their predictive power in the model. These are called variable importance plots. When it comes to the random forest regression model, as we built here, it is possible to evaluate this variable importance in two ways. The first one is based on the increase in MSE, which represents the mean decrease in accuracy when leaving out that particular variable (Datacamp, n.d.). Furthermore, the Gini Index (denoted as Formula (3) in Section 4.1.2) applies as the basis of constructing the optimal random forest model. However, the mean decrease in the Gini Index measure can also be visualized as the increase in node purity for variable importance purposes. This measure represents how each variable contributes to the homogeneity of the nodes and leaves (Martinez-Taboada and Ignacio Redondo, 2020).

Figure 4 shows the variable importance of the variables included in the optimal random forest regression model based on the increase in MSE (left) and the increase in node purity (right). These plots can be read as follows: the bigger the drop in increase, the more important the variable. However, it is not possible to extract the specific effect sizes from this plot so instead, these insights are relative to the other variables. As can be seen, both graphs show that the variables *Retail price* and *Model* have the highest increase in MSE and node purity compared to the other variables. First, this means that these two variables give the highest decrease in accuracy when leaving out and therefore, they seem to be most important for resale market price prediction purposes. Additionally, this means that these two variables are most important when it comes to creating homogeneity of the nodes and leaves. These findings correspond to the significant findings in the multiple linear regression.

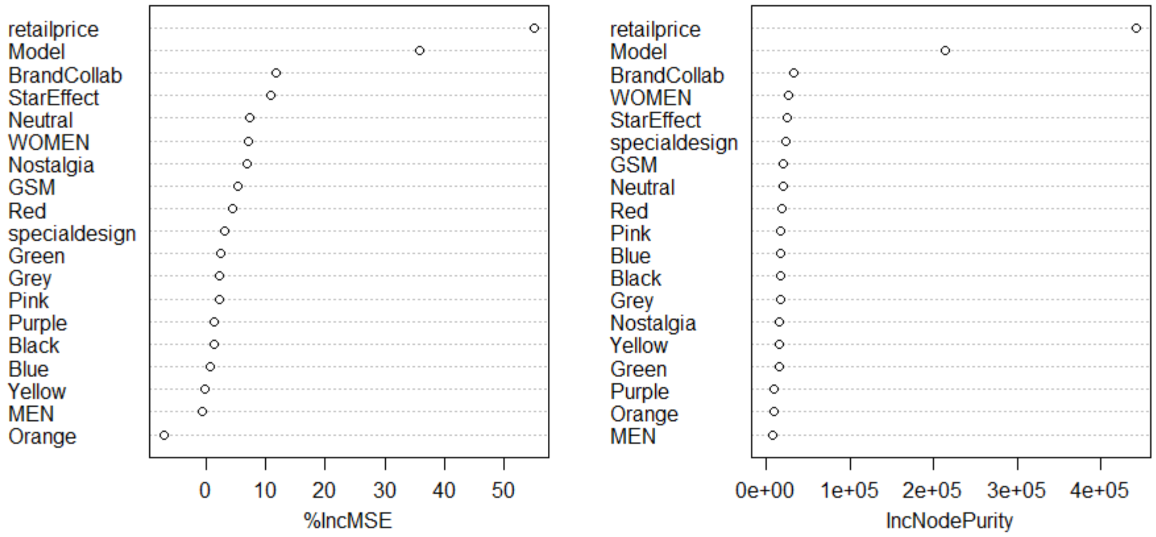


Figure 4. Variable importance using the increase in MSE (left) and the increase in node purity (right).

To conclude, just as the multiple linear regression, the best performing random forest regression model chosen in this thesis has room for improvements. The same solution holds here, namely making sure to add important information beyond the sneaker characteristics that might influence the resale market introduction price. However, this model too does give some insights into what existing variables are important for predicting the resale market introduction price, being *Retail price* and the *Model*.

5.1.3 Evaluation of methods

Now that both models are built and evaluated separately, the two methods' optimal models are compared to each other to see which method performs better in predicting the resale market introduction price, using the testing set. For this, the RMSE is used as comparable measure. As previously discussed, the RMSE refers to the square root of the discrepancies between the predicted and actual points. Table 8 below shows the RMSE for both models built. Interpreting these values results in the conclusion that the average deviation between the predicted resale market introduction price and the actual resale market introduction price is €40.37 in case of the multiple linear regression and €36.71 in case of the random forest regression. As can be seen, the random forest regression model has a lower RMSE than the multiple linear regression model, meaning that using the machine learning method leads to somewhat smaller discrepancies between the predicted and actual points. In other words, the random forest regression can predict the resale market introduction prices most accurately and therefore, outperforms the multiple linear regression.

Table 8. RMSE based on the results of the multiple linear regression and random forest regression using the testing set.

	Multiple linear regression	**Random forest regression**
RMSE	40.37	36.71

5.1.4 Robustness check: evaluation of hypothesis testing

Previously, the statistical inferences found using the multiple linear regression are discussed, however, statistical tests have proven that the assumptions needed to be able to perform hypothesis testing (i.e., being able to make pure statistical inferences) were not met. This has led to some changes in the model, such as running the model with robust standard errors (as improvement regarding the homoskedasticity assumption) and removing outliers and transforming numeric variables (as improvement regarding the normality assumption, and potentially the zero conditional mean assumption). The goal for this section is to check if the results from the multiple linear regression are robust to the possibility that a bigger sample size (i.e., including the full dataset, not just the training set) improves the performance of the model and whether it is able to assume that the assumptions needed for statistical inferences hold in this case, instead of having to make changes to the model or data. Additionally, it is checked whether the statistical inferences that are made before are robust to the increase in sample size, or whether different statistical inferences can be made in this case.

Running the initial multiple linear regression using the full dataset resulted in a R-squared of 0.1903, being quite similar to the initial model solely using the training set. However, it did not lead to a different conclusion in assumptions. As before, only the linearity and no perfect collinearity assumptions seem to hold. Additionally, the random sampling and zero conditional mean assumptions again do not hold, which is highly likely for the same reason, being additional information regarding influences on the resale market introduction price is needed to improve performance of the model. Again, even with the full dataset, it can be concluded that the heteroskedasticity and normality assumption do not hold. Therefore, moving on to finding whether the statistical inferences also differ when using the full dataset (and not just the training set), again the outliers are removed (based on the IQR range), the numeric variables are transformed and the regression is run with robust standard errors.

Table C1 in Appendix C shows the results of the optimal multiple linear regression model built using the full dataset. It shows that in total 11 variables seem to have a significant association with the resale market introduction price, being two more than in the model built solely using the training set. These two variables are the variables representing the model "Yeezy" and the colour "Yellow". Also, the R-squared is slightly higher for this final model using the full dataset than for the final model using the training set solely (49.02% vs. 48.54%). From this can be concluded that the model discussed here performs slightly better than the other model built using the training set solely.

This robustness check shows that increasing the sample size does not change the conclusion about the assumptions, however, it does slightly improve the model performance and the number of statistical inferences by introducing new statistically significant coefficients of two categories. Therefore, it can be concluded that the model is not robust to the sample size, meaning that a bigger sample size does improve the model, which seems a logical conclusion in the case of a linear trend model.

5.2 Average market price trend development

Analysing whether there exists a significant trend in the average market price is a stepping stone to knowing if it might be possible to actually forecast the average market price of sneakers in the resale market, and more specifically, to do this in time before the sneaker is released to evaluate its profitability in advance. Also, it might give insights into possible abnormalities in the market.

5.2.1 Average market price trend

Before starting the analysis, the average market price trend is established and displayed in Figure 5 below. This plot visualizes the average market price trend of the sneakers included in the dataset, and shows that it seems to decrease over time.

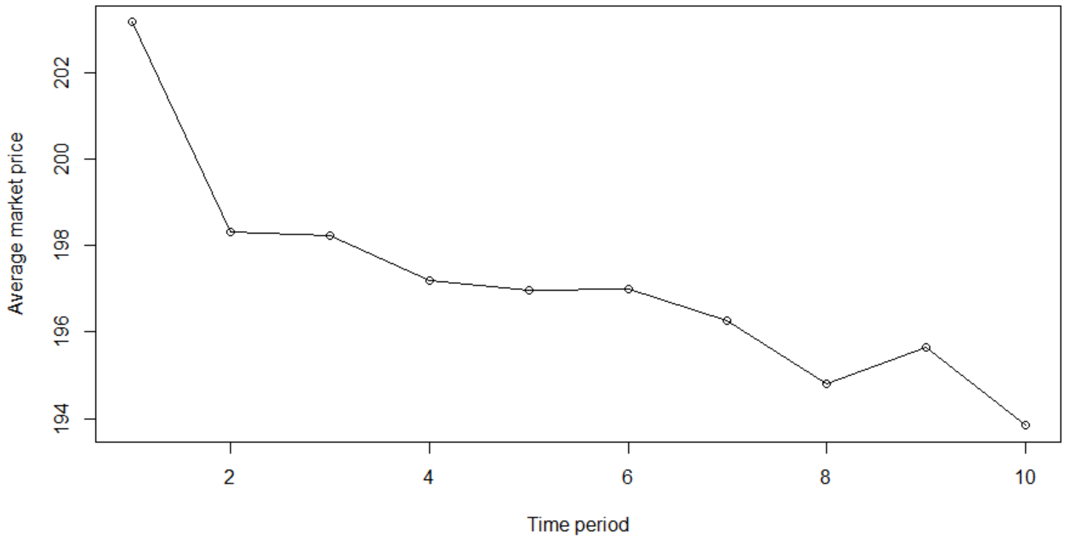


Figure 5. Average market price per time period after release.

To statistically test whether this declining trend of the market price of sneakers exists, the lm()-function from the stats package (R Core Team, 2020) is used. The linear trend model was ran as displayed in formula (4) in Section 4.2. The total number of observations in the dataset (being 10 instead of 638, as explained in Section 4.2) exceeds the minimum required sample size of 5 observations for this analysis[14], of which the calculation itself is discussed in more detail in Appendix D. All assumptions are tested and the corresponding results are extensively discussed in Appendix E. These assumption tests show that only two of the six assumptions hold, being the no perfect collinearity and normality assumption. Again, as regards to the random sampling and zero conditional mean assumptions, it is impossible to improve the model with the available variables and information at hand[15] and therefore, solutions to this are discussed in the Limitations. A solution for nonlinearity (as assumed here since the linearity assumption does not hold) is to use non-linear transformations of the independent variable, such as natural logarithm, square root or squared values[16]. A solution for heteroskedasticity (as assumed here since the homoskedasticity assumption does not hold) is running the regression with robust standard errors, since the standard errors calculated for the initial model are incorrect and therefore, conclusions about significance are incorrect as well. Transforming the independent variable into the natural logarithm of this variable (leading to the model becoming a linear-log model) and running the regression with robust standard errors, impacted the performance of the model from a R-squared of 0.7547 for the initial model to a R-squared of 0.8886 for this final model.

The results of the final linear trend model are given in Table 9 below. Since the linear trend model is transformed into a linear-log model, the values of the coefficients should be transformed to be able to interpret the results. This is done by the following rule: a 1% increase in the independent variable leads to an expected increase in the dependent variable of $\beta_i/100$ units (Benoit, 2011). In this case, the intercept which represents the magnitude of the average market price without linking a period to it, is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.01) and results in an average market price of €202.12. This table also shows that the coefficient of the natural logarithm of *Time period* is -3.2923 and is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.01). This means that for every 1% increase in time period that the sneaker included in this dataset is on the resale market, the average market price decreases with €0.03. Thus, the results suggest a declining trend in the average market price of the analysed sneakers in the first ten weeks after release. When evaluating the performance of the model it shows an R-squared of 0.8886, meaning that 88.86% of the variance in the average market price is explained by the time period. From this can be concluded that the model performs very well and explains a large part of the

[14] This number is calculated using the G*Power application, version 3.1.9.7.
[15] When assuming that the reason for rejecting the zero conditional mean assumption is that the model also estimates reverse causality, OVB, etc.
[16] Which might also improve the zero conditional mean assumption when assuming the reason for the rejection is that the model misses nonlinearities.

variance in the dependent variable. Since this number is not 1, it still means that there is some room for improvement.

Table 9. Results of the final linear trend model with average market price as the dependent variable.

Variable	**Coefficient**
Intercept	202.1177***
Log(Time period)	-3.2923***

Note. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01. N = 10; Adjusted R-squared = 0.8886.

As discussed, linear trend models often do not capture the true course of the trend. Even though the final model fits the data well, not all assumptions hold, leading to the possibility that this model might not be the best model for this analysis. Therefore, as additional step it is tested whether the time series might be covariance stationary or serially correlated, to see whether more advanced forecasting models such as autoregressive models are a better fit (CFA, 2022). Covariance stationary is tested using the Augmented Dickey-Fuller test, with the null hypothesis that the time series is non-stationary. This test results in a DF-statistic of -2.0755 which is not statistically significant and different from zero on a 10% significance level. This means that it is not possible to reject the null hypothesis and thus, the time series is not likely to be covariance stationary. This could indicate that a linear trend model is sufficient to model the time series. However, as confirmation, serial correlation is also tested for using the Durbin-Watson statistic, with the null hypothesis that the errors are not serially correlated. This test results in a DW-statistic of 0.5665 which is not statistically significant and different from two on a 10% significance level. This means that it is not possible to reject the null hypothesis and thus, the error terms are not likely to be serially correlated. From this can be concluded that the linear trend model regressed in this section is indeed sufficient to capture the dynamics of this time series (as confirmed by the high R-squared as well). This means that it is not needed to build more complex models since this model can be used to forecast future values.

5.2.2 Robustness check: random sampling assumption

Since the assumption of random sampling does not hold it might be of interest to see whether each model has their own average market price trend (since random sampling does hold when examining the data per sneaker model). Table F1 in Appendix F shows the linear trend model results for each of the sneaker models included in this dataset. As can be concluded from the intercepts which are all statistically significant and different from zero on a 10% significance level (all with a p-value lower than 0.01), all sneaker models have a different average market price without linking a period to it, ranging from a minimum of €133.11 for the model “Blazer” to a maximum of €247.85 for the model “Yeezy”. When it comes to average market price without linking a period to it, the models “Air Force”, “Air Max” and “Blazer” are comparable, the models “Air Jordan”, “Dunk” and “Waffle” are comparable and lastly, “Yeezy” seems to have the most expensive average market price without linking a period to it.

Moving on to the average market price trend, no significant trend in average market price is found for the models “Air Jordan”, “Dunk” and “Yeezy” which means that there seems to be no change in price (or profit) each week it takes longer to sell a pair of these sneakers. For the models “Air Force”, “Air Max”, “Blazer” and “Waffle”, a significant declining trend in the average market price is found in the first ten weeks after release. This means that for every 1% increase in time period that a pair of these sneakers is on the resale market, the average market price (or profit) decreases, however, in different magnitudes.

When looking at performances, especially the linear trend models regarding the sneaker models “Air Max” and “Blazer” seem to perform very well based on the R-squared (97.16% and 95.06% of the variance in the average market price is explained by the change in time period after release, respectively). The linear trend models for the sneaker models “Air Force” and “Waffle” seem to perform less good, however, can still explain a big part of the variance in average market price by the change in time period after release. And the linear trend model for the sneaker models “Air Jordan”, “Dunk” and “Yeezy” perform worst, since only a small part of the variance in the average market price is explained by the change in time period after release, which is logical since this variable is not significant. Therefore, reasons for the latter linear trend models performing worse could be missing information besides the time period that influences the average market price of these sneakers more than that of the sneakers that have a better performing linear trend model and are mainly influenced by the time period.

This robustness check shows that when focussing on the linear trend per sneaker model (where the random sampling assumption holds), there seems to exist a statistically significant trend in the average market price for approximately 43% of the sneaker models included in this dataset. These trends are somewhat comparable to the results of the linear trend model including all models (where the random sampling assumption does not hold), which means that the initial results of the main analysis are somewhat robust to the fact that the random sampling assumption does not hold.

5.2.3 Robustness check: normality assumption

As shown in Section 3.4, the retail price and average market prices over the first ten weeks after release were skewed to the right (and thus not normally distributed). However, using statistical tests we proved that the normality assumption holds. To check if the results are robust to the possibility that this assumption seems to not be true, the same analysis has been re-run on a subset of sneakers belonging to different retail price categories. Table F2 in Appendix F shows the results of the optimal linear trend models for sneakers having a retail price below $100, from $100 to $200, and $200 and above.

As can be concluded from the intercepts, which are all statistically significant and different from zero on a 10% significance level (all with a p-value lower than 0.01), all sneakers falling in the three retail price categories have a different average market price without linking a period to it, ranging from a minimum of €146.20 for the sneakers with a retail price below $100 to a maximum of €307.79 for the sneakers with a retail price above $200, with the middle category falling in between those two extremes. Moving on to the average market price trend, for the subset of sneakers that have a retail price below $100 or $200 and above there seems to be no significant trend in average market price development. This means that it cannot be concluded whether an average market price trend exists for sneakers that are sold for a retail price of $100 and less or $200 and more on the primary market. Additionally, the R-squared is quite low for both categories (0.0593 and 0.1069, respectively), meaning that only a small part of the variance in the dependent variable can be explained by change in the independent variable, which is logical since this variable is not significant. One reason for this could be that there exists other information that explains the development in average market price for these sneakers that are not included in this model. However, when it comes to the subset of sneakers that have a retail price from $100 to $200, it can be concluded that the coefficient of -3.54 is statistically significant and different from zero on a 10% significance level (with a p-value lower than 0.01). This means that for every 1% increase in time period that a pair of these sneakers is on the resale market, the average market price (or profit) decreases with €0.04. When evaluating the performance of this model, it shows an R-squared of 0.9297, meaning that 92.97% of the variance in the average market price is explained by the change in time period. From this can be concluded that this model performs well, and even better than the model built using the original data. However, since this number is not 1, it still means that there is some room for improvement.

This robustness check shows that when focussing on the sneakers that have a retail price below $100 or $200 and above, there seems to exist no such thing as a statistically significant trend in the average market price. For the other category representing sneakers that have a retail price between $100 and $200, the linear trend model shows evidence of a trend in the average market price of sneakers on the resale market that is comparable to the trend found using the original dataset. This latter category is the biggest category and also the middle one, thus leaning most towards a version of the full dataset where normality holds. Combining this with the findings means that the results are mainly robust to the possibility that the normality assumption might not be true.

6. Discussion

In this last section, a conclusion is drawn about resale market introduction price prediction and average market price trend development. Additionally, the limitations of this thesis are discussed including the possibilities for future research based on this thesis' findings.

6.1 Conclusion

The combination of the lack of existing knowledge on the sneaker resale market and the interesting comparable aspects of this market to other markets inspired the goal for this thesis, which is expressed in the main research question:

"To what extent can the market price of sneakers on the resale market be forecasted before release?"

The findings of this thesis contribute both to the general knowledge of market prices, which can also be applied to other product categories, and to the still limited amount of existing knowledge on the rapidly emerging sneaker resale market. In addition, because of the ability to compare events in the more "mysterious" sneaker resale market to events happening in other well-known financial markets, findings of this thesis also fill a gap in single price prediction and market price trend development literature.

The main research question was divided into two parts being (i) predicting the introduction price on the resale market in the first week after the release date, and (ii) evaluating the average market price trend development in the first ten weeks after release. To conclude, it is proven that without historical price data, thus when only employing the sneaker characteristics and fixed price indicators found in previous literature, the market price of sneakers on the resale market (in the form of market introduction price) can be forecasted to some extent, however, improvement is possible. When historical price data does exist and thus, a trend analysis can be done, the market price of sneakers on the resale market seems to follow a negative trend and can be forecasted quite accurately, even without adding any other information that could be of an influence to that trend. The comprehensive implications of both parts, each representing their own sub question, are discussed below.

6.1.1 Implications for resale market introduction price prediction

This section discusses the results regarding the resale market introduction price prediction, and especially answers the sub question asking whether a statistical method or a machine learning method performs better in predicting the introduction price of sneakers in the resale market. The way this single price prediction occurs is comparable to IPO short term price predictions, of which previous literature has shown that machine learning methods outperform statistical methods, leading to the following hypothesis being formulated:

Hypothesis 1: The machine learning method outperforms the statistical method in predicting the resale market introduction price.

The findings in this thesis prove that the machine learning method used in this thesis (i.e., the random forest regression) indeed outperforms the statistical method used in this thesis (i.e., the multiple linear regression), based on the performance evaluation metric RMSE. One reason for this could be because of the fact that the main goal of machine learning methods is to repeatedly make accurate predictions, whereas the goal of statistical methods is to make statistical inferences. Also, machine learning methods try to learn from the data in such a way that it can work well on data that it has never seen before, whereas statistical methods only make sure to fit the given data in the best way possible and thus, might not at all fit new data that it has never seen before. Even though the behavioural characteristics could not be included, both models have proven to be able to predict the resale market introduction price of sneakers to some extent. All in all, it can be concluded that this hypothesis can be accepted. The machine learning method outperforms the statistical method in predicting the resale market introduction price.

Apart from which method is best, and what might even be more insightful, is the fact that these analyses have shown what specifically influences this resale market introduction price. When it comes to sneaker characteristics and fixed price indicators (both types of variables that were included in the models), it is proven that the type of sneaker model, which sex the sneaker is labelled as, the colours that are included in the sneaker, the retail price, whether the sneaker is created in collaboration with a brand and whether the sneaker evokes nostalgic feelings are of importance when predicting the resale market introduction price of a sneaker. These findings are based on the multiple linear regression, however, the random forest regression confirms the importance of both the sneaker model and retail price. Robustness checks show that when increasing the sample size, the multiple linear regression model performs better than before and even more statistical inferences can be made, making these results not robust to sample size.

6.1.2 Implications for average market price trend development

In this section the results regarding average market price trend development are discussed, with the focus on answering the second sub question asking whether the average market price trend can be perceived to be different from a horizontal line (i.e., whether there exists a trend). Knowing that different price indicators (e.g., sneaker characteristics, fixed price indicators and behavioural characteristics) exist that affect the price trend in their own way, leads to no straightforward conclusion about the price development. Additionally, briefly evaluating the market price trends of several sneakers have led to the insights that multiple price trends exist. Averaging the potential trends possibly results in a rather flat line, leading to the following hypothesis being formulated:

Hypothesis 2: The average market price trend equals a horizontal line.

However, the findings in this thesis have proven that there exists a slightly declining trend in the average market price of sneakers included in the dataset in the first ten weeks after release and therefore, it can be concluded that this hypothesis can be rejected. The average market price trend development does not equal a horizontal line, however, declines over time. As discussed before, the main reason for this could be that it matters how old the sneaker model is. New sneakers are released every day, which means that the hype for a certain sneaker fades rather quickly and moves on to the newly released ones. This might make the demand peak in the beginning leading to higher willingness to pay and/or ask, which probably declines once time passes. Additionally, it is proven that using a linear trend model is sufficient to forecast future values and it is not needed to build more complex models such as autoregressive models. Since each sneaker is different, robustness checks are performed to test whether the trend (existence) changes when applying different variations of the included observations. These results show that when testing per model, there only exist a trend in the average market price of sneakers for about half of the sneaker models in the first ten weeks after release, however, these slopes are in line with the key findings, making the results somewhat robust to non-random sampling. They also show that the existence of a trend varies for sneakers subdivided by retail price, making the results robust to non-normality as well.

6.2 Limitations and implications for future research

As established from the beginning of this thesis, the single price prediction models might suffer from *Omitted Variable Bias (OVB)* since it was impossible to include the behavioural characteristics, which seem to affect the resale market introduction price and average market price trend development as well. This could possibly be confirmed by the fact that the zero conditional mean assumption does not hold and therefore, it might be possible that no consistent causal effect can be estimated. OVB refers to a bias that occurs when leaving out relevant variables, leading to this effect partly being captured in the error term and, unjustified, partly being attributed to the effect size of the included variables. As can be concluded from the performance metrics, the variance in the dependent/predictor variables was not fully explained by change in the included independent/response variables. Thus, the single price prediction models have possibilities for improvements. Therefore, for future research, it is suggested to perform the same comparison of analysis including the behavioural characteristics (e.g., by turning it into an field experiment) and other possibly affecting information to see whether the prediction performances increase and thus, the single price prediction models become more accurate in their predictions. However, even though the current results may suffer from omitted variable bias and the quality of the data could be improved, all models were still able to both explain a substantial part of the variance in the resale market introduction price and to prove the existence of a declining trend in the market price development.

A second limitation regards to the data collected leaning towards selection bias. The dataset only includes the sneakers of the two biggest brands Nike and Adidas over a period of fifteen months due to time constraints, since collecting this data already resulted in dozens of hours of manual data acquisition. Another reason for choosing to only include these sneakers is to get enough observations per category of the variable *Model*. Since each brand has their own models which they continue to release over a longer time period than of which this data is collected, including all existing brands and all their models in the chosen time frame would lead to not enough observations per category to be able to draw a conclusion from it. Therefore, this thesis can rather be viewed as an exploratory research of the resale market price of sneakers from Nike and Adidas, from which cannot be concluded whether external validity holds. For future research, it is suggested to expand the dataset by taking a longer time period into account and including all sneakers released in that time period. This might lead to a higher quality dataset of a bigger size which eliminates the possible selection bias, and when used to re-run the analysis might in turn improve the quality of the results and possibly improve external validity.

Another implication for future research concerns the trend evaluation. Since this thesis can be considered an exploratory study, only the basic analysis is performed to find whether an average market price trend exists. It is not studied what might have an influence on this trend. Future research could include possible variables affecting the average market price trend (e.g., sneaker characteristics, fixed price indicators and behavioural characteristics), and thus expand this linear trend model to see whether it is possible to find what influences the slope of this trend.

6.3 Final words

To conclude, this thesis serves as an exploratory study and has proven that analyses of single price prediction (in the form of resale market introduction price prediction) and price trend development in the sneaker resale market lead to the possibility to subtract interesting (initial) insights from it. Also, in the latter case, no complex models are needed to improve the results obtained with the simpler statistical models, leading to a lower threshold for researchers to start analysing this market. All in all, this thesis can be seen as a pre-review and future research is encouraged to contribute to the existing knowledge of this rapidly emerging, but under-analysed market.

References

Baba, B., & Sevil, G. (2020). Predicting IPO initial returns using random forest. *Borsa Istanbul Review, 20*(1), 13-23.

Belk, R. W. (1995). Collecting as luxury consumption: Effects on individuals and households. *Journal of Economic Psychology*, *16*(3), 477–490.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London, 22*(1), 23-36.

Bhandari, A. (March 20, 2020). *What is Multicollinearity? Here's Everything You Need to Know.* Retrieved from: https://www.analyticsvidhya.com/blog/2020/03/what-is-multicollinearity/

Botterill, J. (2007). Cowboys, Outlaws and Artists The rhetoric of authenticity and contemporary jeans and sneaker advertisements. *Journal of Consumer Culture*, *7*(1), 105–125.

Braithwaite, N. (May 17, 2021). *The history of sneakers: from commodity to cultural icon.* Retrieved from: https://theconversation.com/the-history-of-sneakers-from-commodity-to-cultural-icon-127268

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brock, T. C. (1968). Implications of commodity theory for value change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological Foundations of Attitudes*, 243-275. New York: Academic Press.

Brown, S., Kozinets, R. V., & Sherry, J. F. (2003). Sell me the old, old story: Retromarketing management and the art of brand revival. *Journal of Customer Behaviour, 2*(2), 133-147.

Bruin, J. (2006). *newtest: command to compute new test.* UCLA: Statistical Consulting Group. Retrieved from: https://stats.oarc.ucla.edu/stata/ado/analysis/

Business Insider, by Ciment, S. (January 4, 2022). *SNEAKER RESELLING SIDE HUSTLE: Your guide to making thousands flipping hyped pairs of Dunks, Jordans, and Yeezys.* Retrieved from: https://www.businessinsider.com/ultimate-guide-to-make-6-figures-reselling-sneakers-nike-yeezy?international=true&r=US&IR=T

Cassidy, N. G. (2018). *The effect of scarcity types on consumer preference in the high-end sneaker market,* Doctoral dissertation, Appalachian State University.

Cattaneo, E., & Guerini, C. (2012). Assessing the revival potential of brands from the past: How relevant is nostalgia in retro branding strategies? *Journal of Brand Management*, *19*(8), 680–687.

CFA (2022). *Refresher Reading: Time-Series Analysis*. CFA Program Level II. Retrieved from: https://www.cfainstitute.org/en/membership/professional-development/refresher-readings/time-series-analysis

Choi, J.W. (June 12, 2017). *Sneakerheads' assessment of sneaker value and behaviors throughout the sneaker ownership cycle*. Doctoral dissertation, Oregon State University.

Choi, J. W. J., & Kim, M. (2019). Sneakerhead brand community netnography: An exploratory research. *Fashion, Style & Popular Culture, 6*(2), 141–158.

Chu, H. (2013). A conceptual model of motivations for consumer resale on C2C websites. *The Service Industries Journal*, *33*(15-16), 1527-1543.

Chu, H., & Liao, S. (2007). Exploring consumer resale behavior in C2C online auctions: taxonomy and influences on consumer decisions. *Academy of Marketing Science Review, 11*(3).

Chu, H., & Liao, S. (2010). Buying while expecting to sell: The economic psychology of online resale. *Journal of Business Research, 63*(9–10), 1073–1078.

Corporate Finance Institute (n.d.). *What is Risk Aversion?* Retrieved from: https://corporatefinanceinstitute.com/resources/knowledge/finance/risk-aversion/

Cowen Equity Research, by Kernan, J., Chen, O., Zuber, K. and Orr, J. (n.d.). *Sneakers as an alternative asset class, part II*. Retrieved from: https://www.cowen.com/insights/sneakers-as-an-alternative-asset-class-part-ii/

Datacamp (n.d.). *Introduction to Machine Learning in R*. Retrieved from: https://campus.datacamp.com/courses/introduction-to-machine-learning-in-r/how-much-will-i-earn?ex=6

Dubois, B., & Paternault, C. (1995). Understanding the world of international luxury brands: the "dream formula". (Special Issue: Research Input into the Creative Process). *Journal of Advertising Research, 35*(4), 69-77.

Esfahanipour, A., Goodarzi, M., & Jahanbin, R. (2016). Analysis and forecasting of IPO underpricing. *Neural Computing and Applications*, *27*(3), 651-658.

Fleming, J., & Ostdiek, B. (1999). The impact of energy derivatives on the crude oil market. *Energy Economics*, *21*(2), 135-167.

Goto, M., & Karolyi, G. A. (2004). Understanding electricity price volatility within and across markets. *Dice Center Working Paper No. 2004*(12).

Higgins, E. T. (2006). Value from hedonic experience and engagement. *Psychological Review*, *113*(3), 439-460.

Highsnobiety, by Leach, A. (August 7, 2018). *How Nike Controls the Sneaker Resell Market*. Retrieved from: https://www.highsnobiety.com/p/sneaker-resell-market/

IBM (n.d.). *Linear regression*. Retrieved from: https://www.ibm.com/nl-en/analytics/learn/linear-regression#:~:text=What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable

Investopedia, by Hall, M. (November 30, 2021a). *What "Going Public" Means*. Retrieved from: https://www.investopedia.com/ask/answers/what-does-going-public-mean/

Investopedia, by Chen, J. (June 29, 2021b). *Risk Averse*. Retrieved from: https://www.investopedia.com/terms/r/riskaverse.asp#:~:text=The%20term%20risk%2Daverse%20describes,slowly%20and%20steadily%20over%20time

Investopedia, by Gad, S. (July 28, 2021c). *How an Initial Public Offering (IPO) is priced*. Retrieved from: https://www.investopedia.com/articles/financial-theory/11/how-an-ipo-is-valued.asp

Investopedia, by Hayes, A. (September 5, 2021d). *Trend Analysis*. Retrieved from: https://www.investopedia.com/terms/t/trendanalysis.asp#:~:text=What%20Is%20Trend%20Analysis%3F,term%20direction%20of%20market%20sentiment

Investopedia, by Fernando, J. (September 12, 2021e). *R-Squared*. Retrieved from: https://www.investopedia.com/terms/r/r-squared.asp

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer: New York.

Kalyanaram, G., & Winer, R. S. (1995). Empirical generalizations from reference price research. *Marketing Science, 14*(3), 161-169.

Kawamura, Y. (Jan 28, 2016). *Sneakers. Fashion, Gender and Subculture*. Bloomsbury Publishing.

Kwartler, T. (2017). *Text mining in practice with R (First Edition).* John Wiley & Sons Ltd.

Landsburg, S. E. (2012). *The armchair economist: economics & everyday life (Rev. ed.).* New York: Free Press.

Lee, B., Rosenthal, L., Veld, C., & Veld-Merkoulova, Y. (2015). Stock market expectations and risk aversion of individual investors. *International Review of Financial Analysis, 40*, 122-131.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22.

Luber, J. (October, 2015). *Why sneakers are a great investment*. TedTalk, Retrieved from: https://www.ted.com/talks/josh_luber_why_sneakers_are_a_great_investment

Lynn, M. (1991). Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology & Marketing, 8*(1), 43-57.

Martinez-Taboada, F., & Ignacio Redondo, J. (2020). *Variable importance plot (mean decrease accuracy and mean decrease Gini).* Retrieved from: https://plos.figshare.com/articles/figure/Variable_importance_plot_mean_decrease_accuracy_and_mean_decrease_Gini_/12060105/1#:~:text=The%20mean%20decrease%20in%20Gini,the%20variable%20in%20the%20model

Matthews, D., Cryer-Coupet, Q., & Degirmencioglu, N. (2021). I wear, therefore I am: investigating sneakerhead culture, social identity, and brand preference among men. *Fashion and Textiles, 8*(1), 1-13.

McCracken, G. (1989). Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research, 16*(3), 310-321.

Meinert, R., Patel, A. & Li, S. (May 24, 2019). *The Importance of Analyzing Model Assumptions in Machine Learning.* Retrieved from: https://towardsdatascience.com/the-importance-of-analyzing-model-assumptions-in-machine-learning-a1ab09fb5e76

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences, 116*(44), 22071-22080.

Nike News (December 2, 2015). *Jordan Brand Evolves The Retro Remaster.* Retrieved from: https://news.nike.com/news/jordan-brand-evolves-the-retro-remaster

Quintana, D., Luque, C., & Isasi, P. (June, 2005). Evolutionary rule-based system for IPO underpricing prediction. *In Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, 983-989.

Quintana, D., Sáez, Y., & Isasi, P. (June 20, 2017). Random forest prediction of IPO underpricing. *Applied Sciences, 7*(6), 636-652.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/

Regnier, E. (2007). Oil and energy price volatility. *Energy Economics, 29*(3), 405-427.

Riskspan (November 7, 2017). *Tuning Machine Learning Models*. Retrieved from: https://riskspan.com/tuning-machine-learning-models/#:~:text=Tuning%20is%20the%20process%20of,of%20a%20machine%20learning%20model

Schwert, G. W. (1989). Why does stock market volatility change over time?. *The Journal of Finance, 44*(5), 1115-1153.

Scornet, E. (2018). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys, 60,* 144-162.

Slaton, K., & Pookulangara, S. (December, 2020). Stompin' in my air force ones: An investigation into the sneaker resale market. In *International Textile and Apparel Association Annual Conference Proceedings, 77*(1). Iowa State University Digital Press.

SneakerJagers (n.d.). *Over ons*. Retrieved from: https://www.sneakerjagers.com/about

Statista, by Richter, F. (May 4, 2022a). *Nike Reigns Over the Sneaker World*. Retrieved from: https://www.statista.com/chart/13470/athletic-footwear-sales/

Statista, by Smith, P. (Jan 13, 2022b). *Value of the U.S. and global sneakers resale markets in 2019.* Retrieved from: https://www.statista.com/statistics/1202148/sneaker-resale-market-value-us-and-global/

Statista Research Department (May 31, 2021). *Global sneakers market revenue from 2012 to 2025.* Retrieved from: https://www.statista.com/statistics/1017918/sneakers-market-value-forecast-worldwide/

Stewart, M. (March 25, 2019). *The Actual Difference Between Statistics and Machine Learning*. Retrieved from: https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3

StockX (n.d.). *The Current Culture Marketplace.* Retrieved from: https://stockx.com/about/how-it-works/

Thompson, C. J., Pollio, H. R., & Locander, W. B. (1994). The spoken and the unspoken: a hermeneutic approach to understanding the cultural viewpoints that underlie consumers' expressed meanings. *Journal of Consumer Research, 21*(3), 432-452.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232.

Tyagi, N. (March 24, 2020). *Understanding the Gini Index and Information Gain in Decision Trees.* Retrieved from: https://medium.com/analytics-steps/understanding-the-gini-index-andinformation-gain-in-decision-trees-ab4720518ba8

Verleger, P. K. (1994). *Adjusting to volatile energy prices. (Vol. 39)*. Peterson Institute.

Watts, A. (2019). Auctions with different rates of patience: Evidence from the resale shoe market. *Managerial and Decision Economics, 40*(8), 882-890.

WealthX (January 10, 2018). *Luxury brand strategy tactics to effectively engage customers*. Retrieved from: https://www.wealthx.com/luxury/2018/luxury-brand-strategy-tactics-to-effectively-engage-customers/

Weinhold, M. (December 2, 2020). *The history of sneaker reselling and the current state of the game.* Retrieved from: https://releaserecovery.com/the-history-of-sneaker-reselling-and-the-current-state-of-the-game/

Wilson, B., & Sparks, R. (1996). "It's gotta be the shoes": Youth, race, and sneaker commercials. *Sociology of Sport Journal, 13*(4), 398-427.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Zhenxiang, G. (2020). Irrational Consumer Behaviors in the Sneaker Market. *Management Science and Engineering, 14*(1), 49-52.

Appendix A – G-Power calculation: multiple linear regression

In this Appendix, the G-Power calculation for the multiple linear regression is discussed in more detail. This calculation is made using the G*Power application (version 3.1.9.7). Power refers to the probability of detecting a “true” effect when such an effect exists (Bruin, 2006). The values displayed in Figure A1 are decided manually and here, an explanation follows. For the Effect size, a number of 0.10 is chosen. An medium effect size of 0.15 is commonly used, however, since this research is rather exploratory, it is of interest to also get findings that have a smaller effect size. As α error probability, a number of 0.10 is chosen, since the minimum significance level in this thesis is 10%, as discussed in the Results. As Power (1 – β error probability), representing the treatment effect when it occurs, a relatively high number is chosen, being 0.90, with the goal to find the truest effect. The number of predictors is 7, since there are 7 predictor/independent variables that are included in the regression (see Section 4.1.1). This leads to a minimum required sample size of 233 observations.

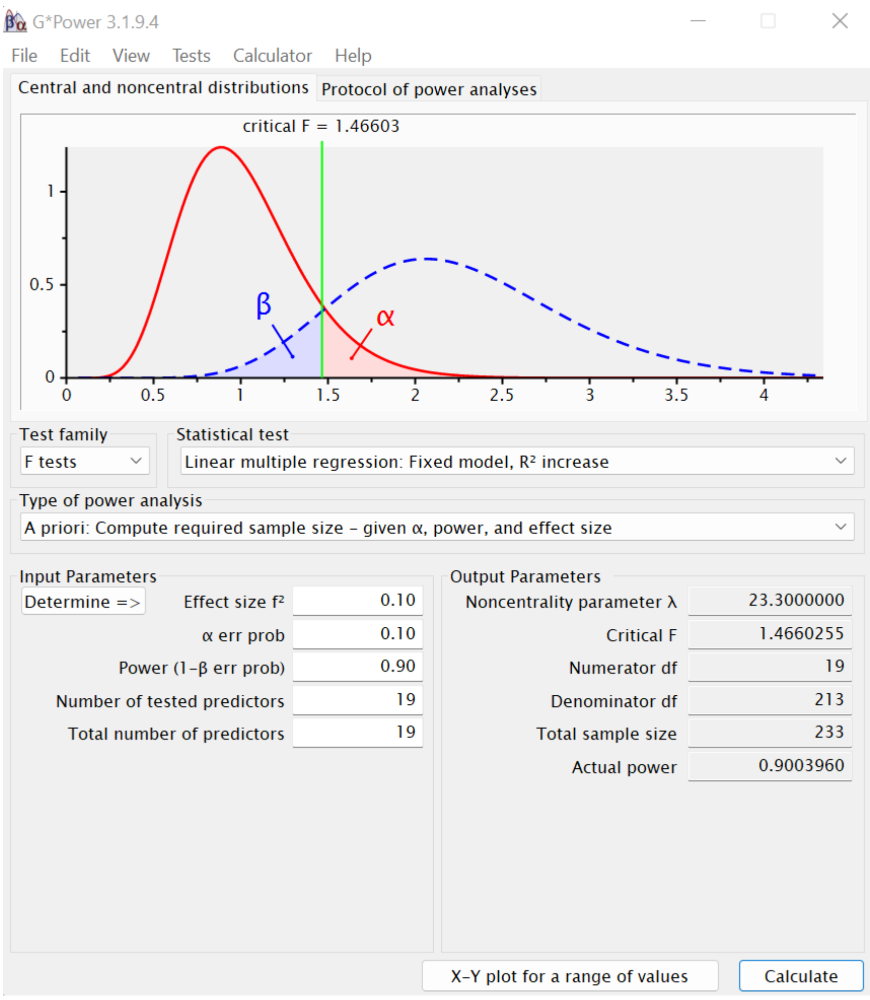


Figure A1. G-Power calculation for the multiple linear regression.

Appendix B – Multiple linear regression assumptions

In this Appendix, the six assumptions of the multiple linear regression are discussed in more detail. This contains the tests used to see whether the assumptions hold, including their corresponding results, in case of the initial (i.e., default) multiple linear regression model.

Assumption 1: Linearity

The linearity assumption refers to the linear relationship between the dependent variable and the independent variables and error term (Wooldridge, 2015). This is tested by the means of a scatterplot visualizing the residuals and fitted values, displayed in Figure B1 below. Since this plot shows no pattern and the red line is approximately horizontal at zero, it is possible to assume a linear relationship between the dependent variable and the independent variables and error term.

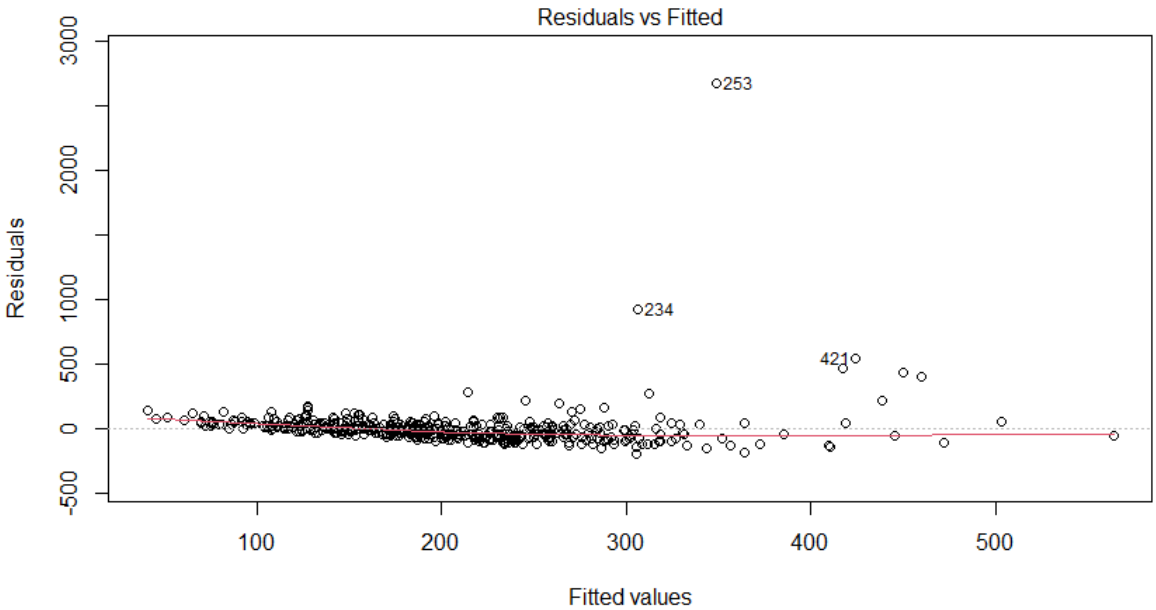


Figure B1. Scatterplot of residuals and fitted values.

Assumption 2: Random sampling

The random sampling assumption ensures that the data is collected randomly, following the population values (Wooldridge, 2015). As discussed in Section 3.1.2, the sneakers were selected based on their release date in order to avoid selection bias as much as possible, however, only the most popular sneaker models from two brands are included to make sure enough data per category is included. Therefore, it cannot be concluded that the random sampling assumption holds.

Assumption 3: No perfect collinearity

The collinearity assumption refers to two things, being: (i) none of the independent variables is constant, and (ii) there are no perfect linear relationships among the independent variables (e.g., when a categorical variable is included, one category should be used as reference category) (Wooldridge, 2015). This is tested by the means of the Variance Inflation Factor (VIF). The VIF score of each independent variable shows how well that variable is explained by the other independent variables, and is predicted by regressing this independent variable against every other variable (Bhandari, 2020). Table B1 below displays the (Generalized) VIF score per variable. A VIF score of 1 means that there is no correlation between this independent variable and the others. A VIF score exceeding 5 indicates high multicollinearity between this independent variable and the others. As can be seen, none of the variables' VIF score exceeds 5 and therefore, it is possible to assume no perfect collinearity between each independent variable and the other independent variables.

Table B1. Generalized VIF score for each variable included in the multiple linear regression.

Variable	**Generalized VIF score**
Model	1.35
Star effect	3.98
Brand collaboration	1.17
Nostalgia	1.12
Men	1.10
Women	1.22
Black	1.07
Neutral	1.08
Grey	1.07
Pink	1.06
Blue	1.02
Yellow	1.06
Green	1.06
Red	1.04
Orange	1.03
Purple	1.05
Special design	1.05
GSM	1.06
Retail price (in $)	1.34

Assumption 4: Zero conditional mean

The zero conditional mean assumption refers to the expected value of the error term taking on a value of zero (i.e., $E(\varepsilon) = 0$). If this is the case, there is independency between the independent variables and the error term and thus, it is possible to find a true causal effect (Wooldridge, 2015). It is the key assumption that ensures unbiasedness of the multiple linear regression and once it holds, it can also be concluded that exogeneity holds. This thesis uses the RESET test to see whether this assumption holds, which exists of the following steps (Wooldridge, 2015): (i) estimate the model of interest; (ii) obtain the predicted values; (iii) re-estimate the model by adding the previously predicted values as additional independent variables; and (iv) perform F-test of joint significance on the additional independent variables. Performing these steps show that the final F-test is statistically significant and different from zero at a 10% significance level (with a p-value lower than 0.01), indicating misspecification and thus, it is not possible to assume that the zero conditional mean assumption holds. There can be two reasons for this, with the first one being that the model could miss important nonlinearities. Secondly, it could mean that the model does not estimate the causal effect consistently, instead this effect becomes a mixture of all things such as reverse causality, OVB, etc. In this latter case, the model does not estimate the causal effect, but rather an association of the variables.

Since two out of four assumptions ensuring unbiasedness seem to not hold, it cannot be concluded that the multiple linear regression model is unbiased.

Assumption 5: Homoskedasticity

The homoskedasticity assumption ensures that the importance of the error term is the same for all individuals independent of their values (Wooldridge, 2015). In this case, the variance of the error term should be the same for all observations. First, this is tested by the means of a scale-location plot visualizing the root of the standardized residuals and fitted values, displayed in Figure B2 below. This figure shows doubtful results, since the red line is not horizontal, and the residuals are not all equally spread. To be able to draw a real conclusion, a Non-Constant Variance score (NCV) test is performed to test for a constant error variance. This test is statistically significant at a 10% significance level (with a p-value lower than 0.01), leading to a rejection of the null hypothesis stating that the variance is constant. This means that it is not possible to assume homoskedasticity and therefore, it is assumed that this model represents heteroskedasticity. In this case, the estimated coefficients of the model are still unbiased (in case all assumptions ensuring this hold), but not efficient.

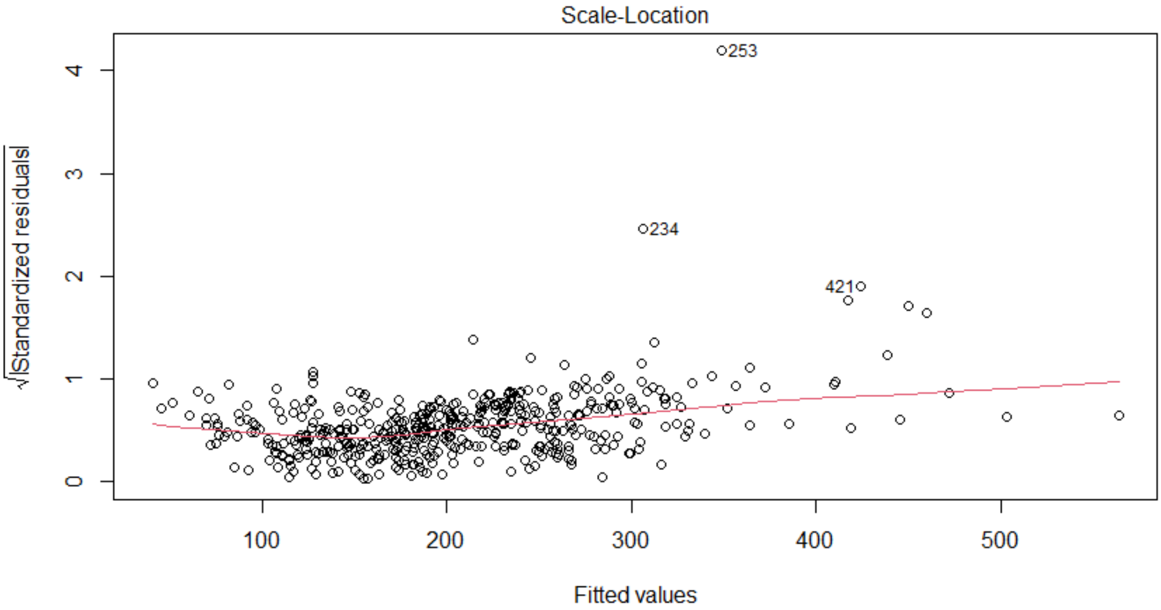


Figure B2. Scatterplot of root of standardized residuals and fitted values.

Assumption 6: Normality

The normality assumption ensures that the error term follows a normal distribution, and that it is independent of the independent variables (with the latter one also being assumed by the zero conditional mean assumption) (Wooldridge, 2015). This assumption is mainly the indicator for exact statistical inference (i.e., characterizing significant relationships in the data). This is tested by the means of a Q-Q plot, displayed in Figure B3 below. Since all data points fall approximately along the reference line (except for at the far right), it is possible to assume normality. However, since this is not entirely the case at the right side and the model free evidence in Section 3.4 has shown that both the resale market introduction price and the retail price are not normally distributed, this conclusion will be confirmed using the Shapiro-Wilk normality test. This test is statistically significant at a 10% significance level (with a p-value lower than 0.01), leading to a rejection of the null hypothesis stating that the population is normally distributed, meaning that we cannot assume normality.

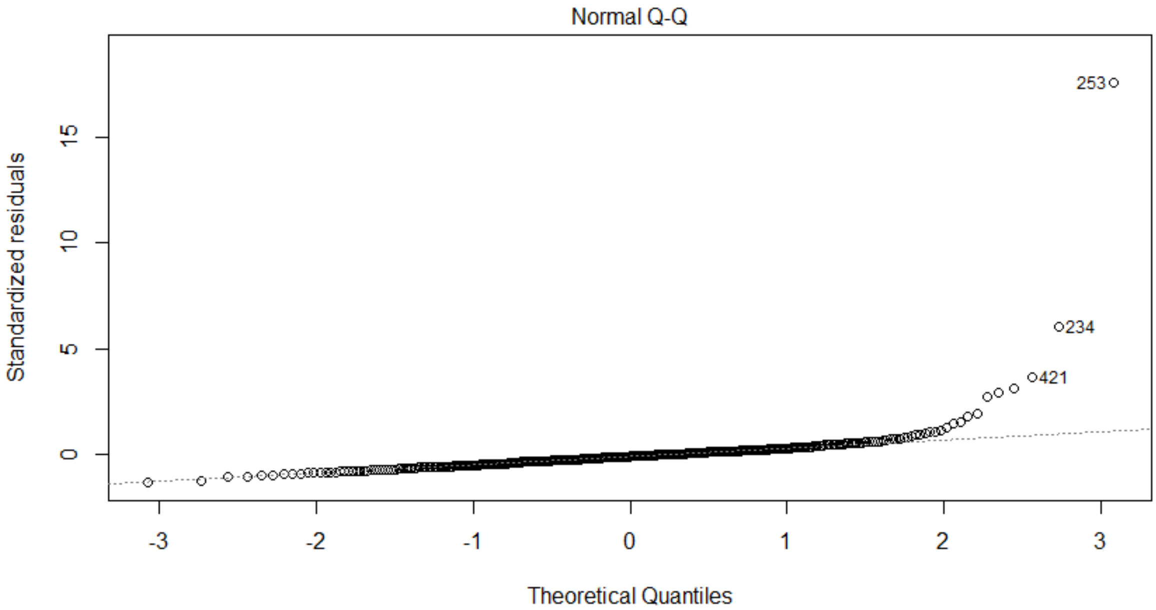


Figure B3. Q-Q plot to test for normality.

Since both assumptions ensuring statistical inference seem to not hold, it cannot be concluded that the multiple linear regression model can be used for hypothesis testing, i.e., characterizing significant relationships in the data.

Appendix C – Robustness check: multiple linear regression

In this Appendix, the results of the multiple linear regression's robustness check regarding the evaluation of hypothesis testing are displayed.

Table C1. Results of the optimal multiple linear regression model using the full dataset and the natural logarithm of the resale market introduction price as dependent variable.

Variable	**Coefficient**
Intercept	4.7394***
Model	
Air Force	-0.1423***
Air Max	-0.0255
Blazer	-0.2259***
Dunk	0.1195**
Waffle	-0.1136
Yeezy	0.0901*
Star effect	0.0248
Brand collaboration	0.1815***
Nostalgia	0.0977**
Sex	
Men	-0.0289
Women	-0.0796***
Colour	
Black	-0.0544**
Neutral	-0.0313
Grey	0.0011
Pink	0.0068
Blue	0.0082
Yellow	-0.0843**
Green	-0.0017
Red	-0.0211
Orange	-0.0065
Purple	0.0069
Special design	-0.0067
GSM	0.0262
Retail price (in $)	0.0034***

Note. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01. N = 601; Adjusted R-squared = 0.4902.

Appendix D – G-Power calculation: linear trend model

In this Appendix, the G-Power calculation for the linear trend model is discussed in more detail. This calculation is made using the G*Power application (version 3.1.9.7). Power refers to the probability of detecting a “true” effect when such an effect exists (Bruin, 2006). The values displayed in Figure D1 are decided manually and here, an explanation follows. Just as before, as α error probability a number of 0.10 is chosen, since the minimum significance level in this thesis is 10%, as discussed in the Results. Again, as Power (1 – β error probability), representing the treatment effect when it occurs, a relatively high number is chosen, being 0.90, with the goal to find the truest effect. The slope of H_0 is chosen to be 0, since this represents the horizontal line. The slope of H_1 is determined by filling in the values in the separate calculation window given in Figure D2. Here, the correlation ρ, standard deviation of the residual σ and the standard deviation of the independent variable σ_x are calculated in R using the data. From this the slope of H_1 and the standard deviation of the dependent variable σ_y were calculated and filled in in the G*Power calculator. This leads to a minimum required sample size of 5 observations.

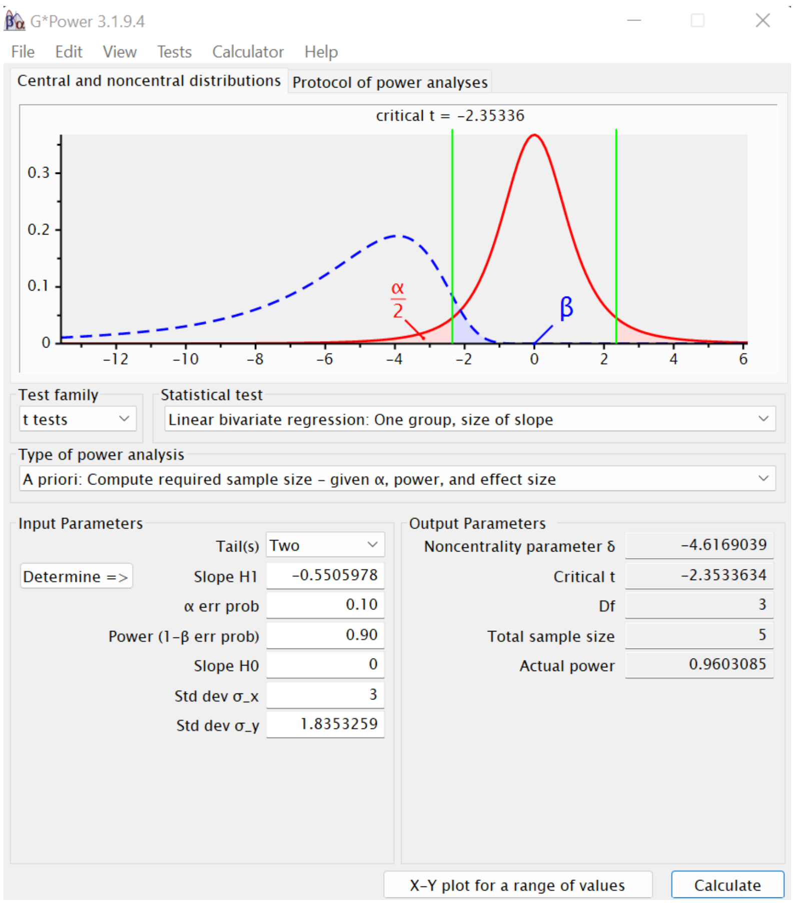


Figure D1. G-Power calculation for the linear trend model.

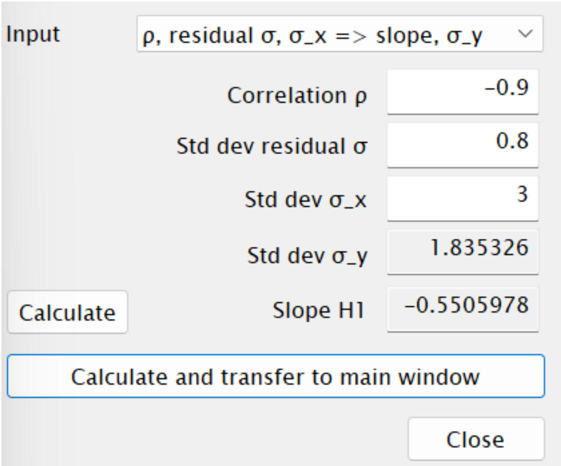


Figure D2. G-Power additional calculation window.

Appendix E – Linear trend model assumptions

In this Appendix, the six assumptions of the linear trend model are discussed in more detail. This contains the tests used to see whether the assumptions hold, including their corresponding results, in case of the initial (i.e., default) linear trend model.

Assumption 1: Linearity

The linearity assumption refers to the linear relationship between the dependent variable and the independent variables and error term (Wooldridge, 2015). This is tested by the means of a scatterplot visualizing the residuals and fitted values, displayed in Figure E1 below. Since this plot shows a pattern and the red line is not horizontal at zero, it is not possible to assume a linear relationship between the dependent variable and the independent variables and error term and thus, this assumption does not hold.

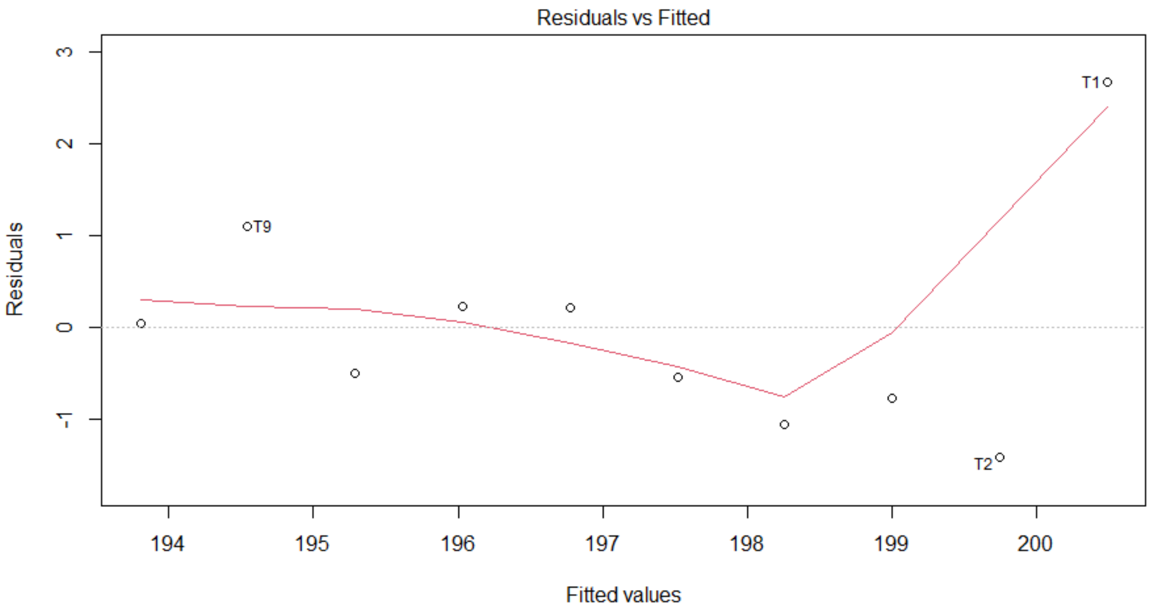


Figure E1. Scatterplot of residuals and fitted values.

Assumption 2: Random sampling

The random sampling assumption ensures that the data is collected randomly, following the population values (Wooldridge, 2015). As discussed in Section 3.1.2, the sneakers were selected based on their release date in order to avoid selection bias as much as possible, however, only the most popular sneaker models from two brands are included to make sure enough data per category is included. Therefore, it cannot be concluded that the random sampling assumption holds.

Assumption 3: No perfect collinearity

The collinearity assumption refers to two things, being: (i) none of the independent variables is constant, and (ii) there are no perfect linear relationships among the independent variables (e.g., when a categorical variable is included, one category should be used as reference category) (Wooldridge, 2015). As explained in Appendix B, this can be tested by the means of the Variance Inflation Factor (VIF). However, in the case of the linear trend model there is only one independent variable and therefore, it can already be assumed that no perfect collinearity exists and thus, this assumption holds.

Assumption 4: Zero conditional mean

The zero conditional mean assumption refers to the expected value of the error term taking on a value of zero (i.e., $E(\varepsilon) = 0$). If this is the case, there is independency between the independent variables and the error term and thus, it is possible to find a causal effect (Wooldridge, 2015). It is the key assumption that ensures unbiasedness of the multiple linear regression and once it holds, it can also be concluded that exogeneity holds. This thesis uses the RESET test to see whether this assumption holds, which exists of the following steps (Wooldridge, 2015): (i) estimate the model of interest; (ii) obtain the predicted values; (iii) re-estimate the model by adding the previously predicted values as additional independent variables; and (iv) perform F-test of joint significance on the additional independent variables. Performing these steps show that the final F-test is statistically significant and different from zero at a 10% significance level (with a p-value lower than 0.01), indicating misspecification and thus, it is not possible to assume that the zero conditional mean assumption holds. There can be two reasons for this, with the first one being that the model could miss important nonlinearities. Secondly, it could mean that the model does not estimate the causal effect consistently, instead this effect becomes a mixture of all things such as reverse causality, OVB, etc. In this latter case, the model does not estimate the causal effect, but rather an association of the variables.

Since three out of four assumptions ensuring unbiasedness seem to not hold, it cannot be concluded that the multiple linear regression model is unbiased.

Assumption 5: Homoskedasticity

The homoskedasticity assumption ensures that the importance of the error term is the same for all individuals independent of their values (Wooldridge, 2015). In this case, the variance of the error term should be the same for all observations. First, this is tested by the means of a scale-location plot visualizing the root of the standardized residuals and fitted values, displayed in Figure E2 below. This figure shows that the red line is not horizontal, and the residuals are not all equally spread. To be able to confirm this conclusion, a Non-Constant Variance Score (NCV) test is performed to test for a constant error variance. This test is statistically significant at a 10% significance level (with a p-value lower than

0.5), leading to a rejection of the null hypothesis stating that the variance is constant. This means that it is not possible to assume homoskedasticity and therefore, it is assumed that this model represents heteroskedasticity. In this case, the estimated coefficients of the model are still unbiased (in case all assumptions ensuring this hold), but not efficient.

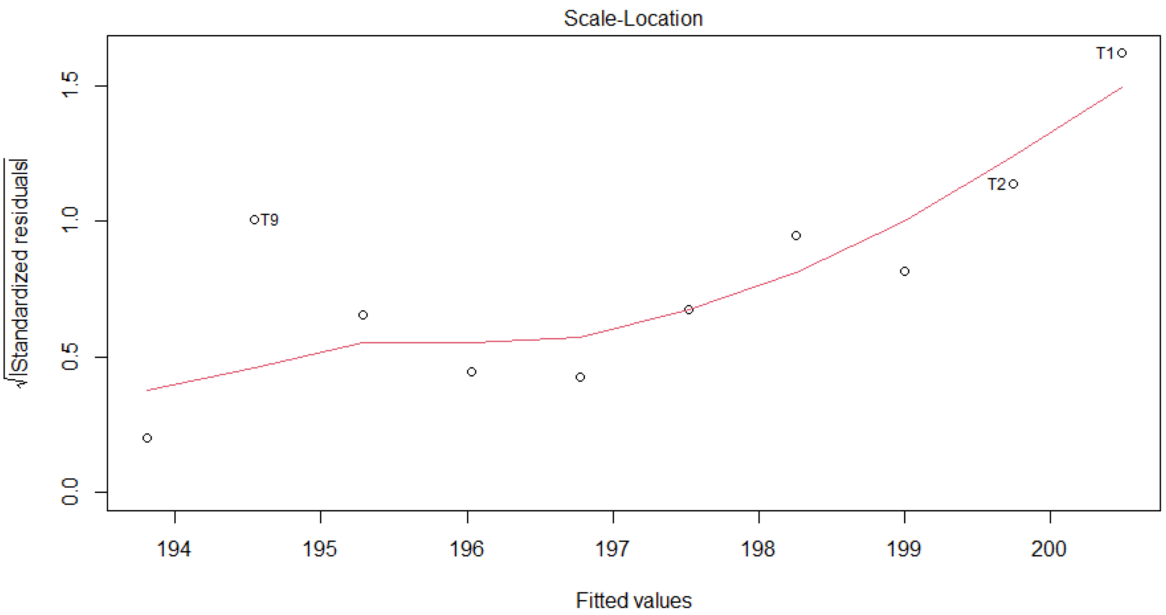


Figure E2. Scatterplot of root of standardized residuals and fitted values.

Assumption 6: Normality

The normality assumption ensures that the error term follows a normal distribution, and that it is independent of the independent variables (with the latter one also being assumed by the zero conditional mean assumption) (Wooldridge, 2015). This assumption is mainly the indicator for exact statistical inference (i.e., characterizing significant relationships in the data). This is tested by the means of a Q-Q plot, displayed in Figure E3 below. Since all data points fall approximately along the reference line (except for at the far right), it is possible to assume normality. However, since this is not entirely the case at the right and the model free evidence in Section 3.4 has shown that none of the average market prices per week are normally distributed, this conclusion will be confirmed using the Shapiro-Wilk normality test. This test shows no significant results, leading to the impossibility of rejecting the null hypothesis stating that the population is normally distributed, meaning that we can assume normality.

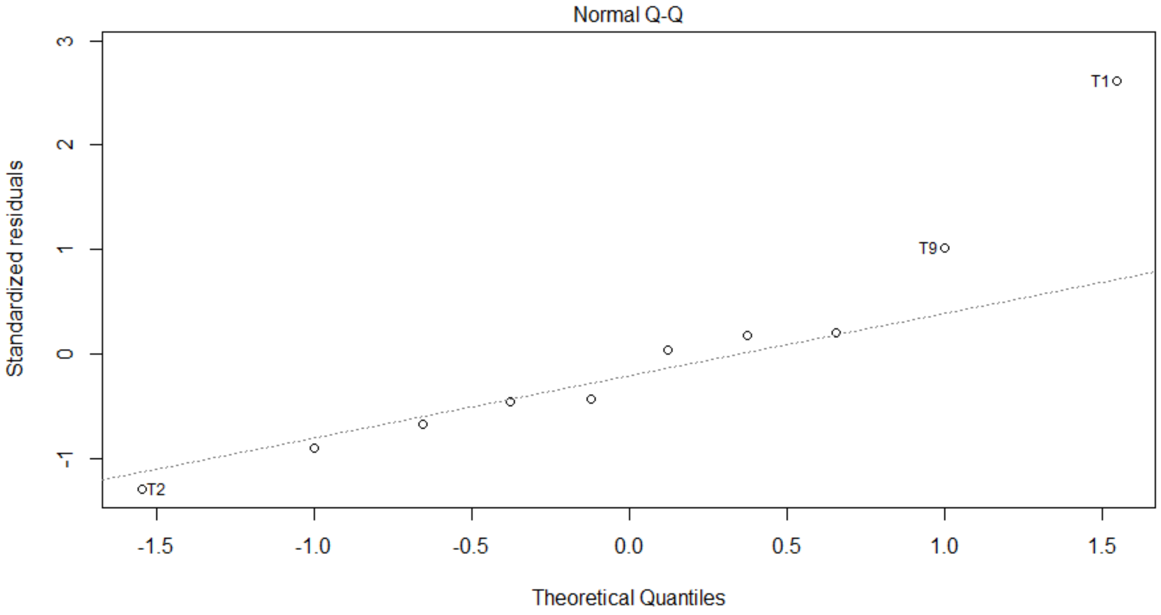


Figure E3. Q-Q plot to test for normality.

Since only one of the two assumptions ensuring statistical inference seem to hold, it cannot be concluded that the multiple linear regression model can be used for hypothesis testing, i.e., characterizing significant relationships in the data.

Appendix F – Robustness checks: linear trend model

In this Appendix, the results of the linear trend model's robustness checks regarding random sampling and normality are displayed.

Table F1. Results of the optimal linear trend models per sneaker model with average market price as the dependent variable.

Model category	**Coefficient Intercept**	**Coefficient log(Time period)**	**R-squared**
Air Force	156.00***	-7.63*	0.7515
Air Jordan	215.21***	-1.80	0.2955
Air Max	183.47***	-11.68***	0.9716
Blazer	133.11***	-12.21***	0.9506
Dunk	221.31***	2.970	0.1306
Waffle	216.44***	-6.20**	0.6861
Yeezy	247.85***	2.87	0.2860

Note. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Table F2. Results of the optimal linear trend models per retail price category with average market price as the dependent variable.

Retail price category	**Coefficient Intercept**	**Coefficient log(Time period)**	**R-squared**
Below $100	146.20***	0.50	0.0593
From $100 to $200	186.84***	-3.54***	0.9297
$200 and above	307.79***	-2.39	0.1069

Note. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.