CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

The Price Determinants of Investment Rums

Master's thesis

Author: Martin Polanka

Study program: Economics and Finance

Supervisor: doc. PhDr. Jozef Baruník Ph.D.

Year of defense: 2024

Declaration of Authorship

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

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

Prague, April 26, 2024

Martin Polanka

Abstract

This study investigates the pricing determinants of rum and explores its viability as an alternative investment asset. Utilizing a diverse methodological approach, the first inquiry employs Hedonic analysis through Weighted Least Squares, Stepwise regression, Lasso regression, and Bayesian Model Averaging. The second analysis adopts Repeat-sales regression to probe rum's potential as an investment asset.

Key findings reveal that aging and reputation emerge as pivotal variables influencing rum pricing dynamics. Further, the study suggests that rum does not align with traditional collectible assets, as its price fails to exhibit consistent appreciation over time.

In conclusion, this research contributes to a deeper understanding of the factors shaping rum pricing and offers insights into its suitability as an investment asset.

JEL Classification	D44, D47
Keywords	Rum, investment rums, price determinants, pricing
Title	The Price Determinants of Investment Rums

Abstrakt

Tato studie zkoumá faktory ovlivňující cenu rumu a zkoumá jeho využití jako alternativní investiční aktivum.

K analýze těchto faktorů používá autor hedonickou analýzu pomocí Weighted Least Squares, Stepwise regrese, Lasso regrese a Bayesian Model Averaging. K prozkoumání rumu jako alternativního investičního aktiva použil autor Repeat-sales regresi.

Hlavní výsledky odhalují, že staření a reputace se ukazují jako klíčové proměnné ovlivňující dynamiku cen rumu. Dále studie naznačuje, že rum nezapadá mezi tradiční sběratelská aktiva, jelikož jeho cena neprojevuje konzistentní růst v průběhu času.

Závěrem tato práce přispívá k hlubšímu porozumění faktorů, které formující cenu rumu, a poskytuje náhled na jeho vhodnost jako investičního aktivum.

Klasifikace JEL	D44, D47
Klíčová slova	Rum, investiční rumy, cenové determinanty, cenotvorba
Název práce	Cenové Determinanty Investičních rumů

Acknowledgments

The author is grateful especially to doc. PhDr. Jozef Baruník Ph.D. for his valuable advices and patience. Further, the author thanks to his family and friends for support and comments.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Polanka, Martin: *The Price Determinants of Investment Rums* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 64. Advisor: doc. PhDr. Jozef Baruník Ph.D.

Contents

List of Tables

List of Figures

Acronyms

ABV Alcohol By Volume

BMA Bayesian Model Averaging

OLS Ordinary Least Squares

WLS Weighted Least Squares

Chapter 1

Introduction

Due to an extensive diversity in rum production processes, aging techniques, and flavor profiles, the pricing of rums exhibits a substantial range, encompassing both accessible options available in supermarkets for immediate consumption, and exclusive rums that can hold a collectible status. Notably, the exclusive category of rums presents an intriguing and untapped market segment, which has garnered substantial attention and generated enticing investment prospects. Like other esteemed commodities, such as fine wines, luxury cars, and high-end watches, passion-driven investments can be effectively monetized. Despite extensive exploration of investment prospects in other alcoholic beverage domains like wine and whiskey, research on the dynamics and determinants of investment rum prices is notably lacking due to insufficient available data. Hence, there exists an unexplored opportunity to examine the behavior and influencing factors governing the pricing patterns of investment-grade rums.

This thesis has two objectives. The first objective of this thesis is to gather rum data, create a pilot dataset, and delve into the price determinants of exclusive rums. The researcher aims to identify the major determinants that significantly impact rum pricing and explore the specific effects of these determinants on the price. A comprehensive analysis will be conducted, encompassing various potential determinants, including distillery parameters, aging techniques, ratings, and Alcohol By Volume (ABV), among others.

The second objective of this thesis is to evaluate the suitability of collectible rum as an alternative investment asset through a Repeat-sales analysis. By examining historical transactional data from an auction house and employing a Repeat-sales methodology, this study aims to discern the investment potential inherent in collectible rum bottles. This analysis involves tracking the price movements of specific rum bottles over time and potential value appreciation. Through rigorous statistical modeling and econometric techniques, this investigation seeks to unveil patterns in the performance of collectible rum as an investment asset, providing insights into its potential as a lucrative avenue for investors seeking diversification within the alternative assets sphere.

Ultimately, by combining the two analyses this research aims to contribute empirical evidence and informed perspectives to the discourse surrounding the investment prospects of collectible rum within the broader financial landscape.

The thesis is organized into several chapters, each serving a distinct purpose in elucidating the nuances of rum pricing and collectible rum as an investment asset. Chapter two offers a comprehensive overview of rums, delineating their varied characteristics to better comprehend the factors influencing their pricing disparities. In chapter three, an extensive literature review explores prior research concerning the pricing mechanisms of spirits and collectibles, providing a foundational understanding for the analytical framework adopted in this thesis. Chapter four meticulously outlines the datasets utilized in the analyses. Methodological intricacies employed in each analysis are expounded upon in chapter five. Chapter six presents the outcomes of the analytical approaches, synthesizes the findings, and emphasizes the thesis's limitations. Finally, chapter seven concludes with insights into the implications of the research conducted.

Chapter 2

What is rum?

Rum, often regarded as one of the most diverse and storied spirits globally, holds a significant place in the world of alcoholic beverages. Derived from sugarcane or its by-products, rum embodies a rich history intertwined with the colonial past, trade routes, and cultural exchanges. Its origins can be traced back centuries to the Caribbean islands, where sugarcane plantations led to the development of this distilled elixir.

Fundamentally, rum production involves a meticulous process that begins with the harvesting of sugarcane or its by-products, such as molasses or sugarcane juice. This raw material serves as the foundation for rum, and its quality and characteristics profoundly influence the final product. The fermentation and distillation methods employed, the aging process in barrels made from various woods, and the blending techniques applied contribute significantly to the complexity and flavor profile of the rum.

Understanding the intricate process of rum creation is vital in comprehending the multitude of factors that influence its price. The quality and type of sugarcane used, the distillation process (pot vs. column stills), the aging period, the type of barrels used for aging (e.g., oak barrels), and even environmental factors like climate and geography, all play pivotal roles in determining the flavor, aroma, and overall quality of rum. Consequently, each step in the production process adds value and uniqueness to the final product, directly impacting its market value and pricing.

Thus, comprehending the intricate process of rum production becomes imperative in analyzing the price determinants. Factors ranging from raw materials to production methods to aging and blending intricacies collectively contribute to the diverse spectrum of rum offerings and their corresponding market values.

Categorization of rums plays a substantial role in determining their price due to the significant impact of various factors. Age, rarity, and production methods heavily influence pricing. Aged rums, especially those with extensive maturation periods, command higher prices due to the investment in time, storage, and evaporation losses involved in their production. Rums from specific regions with unique geographical indications, such as Agricole rums from Martinique or premium offerings from renowned distilleries, often carry higher price tags due to their limited availability and distinctive qualities. Additionally, specialized production techniques, such as pot still distillation or small-batch releases, contribute to higher costs, reflecting in elevated prices due to the craftsmanship and intricacies involved in creating these exceptional rums.

2.1 Geographical Categorization

Geographical categorization for rums is crucial for preserving authenticity, ensuring quality, supporting local economies, and aiding consumers in recognizing distinct characteristics and traditions of rum production from specific regions. These categorizations safeguard traditional methods and regional nuances while providing assurance of quality, cultural significance, and differentiation in the market, ultimately benefiting both producers and consumers alike.

On a global scale, the Caribbean region stands out as the primary hub for rum production, with numerous countries contributing significantly to the industry. Countries such as Jamaica, Barbados, Cuba, the Dominican Republic, Trinidad and Tobago, and Puerto Rico are among the leading rum-producing nations in the world. Each of these Caribbean nations boasts a rich heritage in rum-making and hosts several distilleries renowned for their distinct rum styles, flavors, and production methods. Additionally, countries in Central and South America, such as Guyana, Belize, and Brazil, also contribute significantly to global rum production, further diversifying the origin and variety of rums available worldwide.

Geographical location also serves as a fundamental determinant of rum pricing, shaping various cost factors that influence the final retail price consumers encounter. From production and distribution expenses to market dynamics and currency fluctuations, the interplay of these elements underscores the complex relationship between geography and rum pricing.

Firstly, production costs vary significantly depending on where rum is produced. Factors such as labor costs, raw material availability, energy expenses, and tax regulations differ across regions. For instance, rum manufactured in regions with higher labor costs or stricter environmental regulations may incur greater production expenses, consequently affecting its retail price.

Moreover, import and export taxes, along with tariffs, exert a substantial influence on the final retail price of rum. These levies are contingent upon trade agreements, governmental policies, and international relations. Therefore, rum imported into countries with high import taxes or tariffs will inevitably be priced higher to accommodate these additional expenses.

Transportation costs represent another significant component of rum pricing affected by geographical location. The expenses associated with shipping rum from production facilities to consumer markets are contingent upon factors such as fuel prices, transportation infrastructure, and the distance between the two points. Consequently, rum distributed to regions with higher transportation costs may be priced higher to offset these logistical expenses.

Furthermore, local demand and market dynamics play a pivotal role in determining rum prices within a given geographical region. In areas where rum enjoys widespread popularity and high demand, prices may be driven upward due to heightened competition among producers and distributors competing for market share.

The reputation and perceived quality of rum originating from specific regions also impact pricing strategies. Certain locales have gained renown for producing high-quality rum with distinct flavor profiles and production techniques, thus commanding premium prices in the market. Consumers often associate these regions with superior craftsmanship and are willing to pay a premium for the perceived authenticity and excellence of their products.

Lastly, currency exchange rates can significantly affect the cost of imported and exported rum. Fluctuations in exchange rates between countries with different currencies can lead to adjustments in retail prices to compensate for currency value disparities. Consequently, consumers may observe fluctuations in rum prices as a result of changes in currency exchange rates over time.

Examples of the most well-known rum-producing regions and their traditional features:

Jamaica

– Jamaican rum is renowned for its robust and flavorful profile achieved through traditional pot stills. Distilleries such as Hampden Estate and Worthy Park are celebrated for their high-ester rums, offering funky, fruity, and aromatic spirits. Longer fermentation periods, sometimes utilizing wild yeast strains, contribute to complex flavors. Extended aging in oak barrels further enriches Jamaican rums, enhancing depth and uniqueness.

Barbados

– Barbadian rum production combines pot and column stills, resulting in a balanced and versatile spirit. Mount Gay and Foursquare distilleries prioritize aging in used bourbon barrels, delivering smooth tastes highlighted by vanilla, caramel, and tropical fruit notes. Foursquare's innovative aging techniques uphold tradition, crafting rums with exceptional depth and complexity.

Martinique

– Rhum Agricole, distinct for using fresh sugarcane juice, creates a unique grassy, vegetal flavor profile with floral and herbal nuances. Martinique's A.O.C. (Appellation d'Origine Contrôlée) designation ensures strict adherence to production regulations, highlighting terroir-driven practices and the island's sugarcane varieties and geographical influences.

Cuba

– Cuban rum, crafted predominantly through column distillation, yields a lighter, cleaner spirit. Brands like Havana Club emphasize smoothness and approachability, aging their rums in various barrels for subtle vanilla, oak, and spice flavors.

Dominican Republic

– Dominican rums offer a diverse range, spanning from light and sweet to robust and complex expressions. Various aging techniques and barrel choices contribute to this spectrum. Some distilleries utilize the solera aging method, imparting depth and complexity to the final product.

2.2 Raw material

There are two main types of rums when considering raw materials used: molasses-based and cane juice-based rums. Molasses-based rums are the most prevalent, deriving from the fermentation and distillation of sugarcane molasses, a byproduct of sugar production. On the other hand, cane juice-based rums, known as Agricole rums, are crafted from fresh sugarcane juice rather than molasses, often associated with French-speaking Caribbean islands.

These distinct categories yield different flavor profiles, with molasses-based rums showcasing a broader spectrum of flavors, while cane juice-based rums typically possess grassier and more vegetal notes, reflecting the raw material's freshness.

The pricing differs in the production process and the cost of raw materials. Molasses-based rums, being more prevalent and utilizing a byproduct of sugar production, may generally be less expensive to produce. In contrast, cane juice-based rums, particularly Agricole rums, often undergo more artisanal production methods and may have limited availability of fresh sugarcane, leading to higher production costs and potentially higher prices for consumers.

Additionally, the flavor profiles associated with each type of rum may also influence pricing. Molasses-based rums, with their broader spectrum of flavors, may appeal to a wider audience and thus be priced competitively. Cane juice-based rums, with their distinct grassy and vegetal notes, may cater to a niche market willing to pay a premium for unique flavor experiences.

2.3 Distillation Methods

Distillation methods are primarily classified into two main types: pot still and column still rums. Pot still rums are typically crafted in traditional copper pot stills, a method known for producing rich, full-bodied spirits with pronounced flavors and aromas. These rums often retain more congeners and flavorful compounds due to the batch distillation process.

In contrast, column still rums, also called continuous still or column/columnnary still rums, are produced in continuous column distillation apparatus. This method allows for higher proof and a lighter, more neutral spirit by continuously separating alcohol from the fermented liquid, resulting in a smoother and more refined taste profile. The choice of distillation method greatly influences the character, complexity, and mouthfeel of the resulting rum.

The price difference lies in the costs associated with each distillation method and the perceived value of the resulting rum.

Pot still distillation, with its batch process often requires more labor-intensive methods and time compared to column still distillation, which is more automated and continuous. Pot still rums may command higher prices in the market due to the perceived quality and craftsmanship associated with traditional methods. The artisanal nature of pot still distillation, along with its ability to produce complex and unique flavor profiles, can justify premium pricing.

On the other hand, column still rums may benefit from the efficiency of automation in production, potentially leading to lower production costs and more competitive pricing.

Generally, premium brands may invest in pot still distillation to create distinctive and high-quality rums, while mass-market brands may opt for column still distillation to produce more affordable options for a broader consumer base. Overall, while both pot still and column still distillation methods contribute to the diversity of rum offerings, the choice of distillation method can impact production costs and, consequently, influence pricing strategies in the rum market.

2.4 Aging

The pricing of rum is closely intertwined with its aging process, which encompasses a spectrum from unaged to extensively aged varieties. Aging rum is a time-consuming endeavor, requiring significant resources to store and maintain it in barrels over time. The longer the aging period, the greater the investment in time and resources, directly influencing the cost of production and subsequently, the pricing of the final product. Additionally, as rum ages, it undergoes chemical reactions within wooden barrels that enhance its flavor profile, complexity, and smoothness.

This quality enhancement contributes to the perceived value of aged rums, allowing producers to justify higher prices for these premium products. Furthermore, extended aging results in a reduction of available stock due to evaporation, also known as "the Angel's share", further contributing to the scarcity of aged rums. This limited supply, coupled with growing consumer demand for well-aged spirits, drives up the prices of aged rums in the market. Collectors and connoisseurs alike seek out aged rums for their unique flavor experiences, further bolstering their market value. In essence, the interplay between aging and pricing establishes a hierarchy within the rum market, where extensively aged rums command higher prices due to their perceived quality, scarcity, and alignment with market demand for premium spirits.

Generally, there are three major categories of rum aging:

- **White/Silver**: Unaged or lightly aged rums that retain a clear or slightly tinted appearance due to minimal contact with wooden barrels. These rums typically exhibit the raw, fresh flavors of the base ingredients.
- **Gold/Amber**: Rums that undergo moderate aging, acquiring a richer color and subtle flavors from the wooden barrels. They strike a balance between the freshness of young rums and the complexity of aged ones.
- **Dark/Aged**: Rums with prolonged aging periods in oak barrels, resulting in deep colors, complex flavors, and nuanced aromas. These rums often exhibit notes of caramel, vanilla, spice, and wood from the extended maturation process.

2.5 Flavor Profiles

Rums are categorized based on a diverse range of flavor profiles that cater to varied preferences. Within each flavor profile, variations in aging, production methods, and regional influences further contribute to a vast array of distinctive taste experiences in the world of rum. Basic distinction of profiles divides rums into these categories:

- **Light and Floral:** Rums with light and floral profiles often undergo shorter aging periods and may use column still distillation methods, resulting in a smoother and milder taste. These rums are generally more accessible and appeal to a broader consumer base. Consequently, they tend to be more competitively priced compared to rums with richer or more complex flavor profiles.
- **Rich and Spicy:** Rums characterized by rich and spicy flavors typically undergo longer aging processes, sometimes in charred barrels, allowing for greater flavor development and complexity. The use of high-quality ingredients and traditional production methods also contributes to their premium status. As a result, rums in this category often command higher

prices, appealing to connoisseurs and enthusiasts willing to invest in a more indulgent and sophisticated drinking experience.

- **Smoky or Funky:** Rums exhibiting smoky or funky characteristics are often associated with unique production techniques or regional influences, such as Jamaican pot still distillation or aging in heavily charred barrels. These rums offer distinctive and sometimes polarizing flavor profiles that appeal to adventurous drinkers seeking something out of the ordinary. Due to their niche appeal and often limited availability, smoky or funky rums can be priced at a premium, reflecting their rarity and the craftsmanship required to produce such distinctive flavors.

Chapter 3

Literature review

The world of collectible spirits, particularly collectible rums, remains a domain largely uncharted within academic discourse. The comprehensive exploration of collectible rum as an investment asset is a burgeoning field that might interest not only scholars but investors as well. Within the purview of investment assets, two predominant analytical approaches emerge: the hedonic approach, centered on unraveling the characteristic features influencing the value of an asset, and the risk-return approach, pivotal in evaluating the investment potential and performance of an asset over time.

The literature landscape pertaining to these methodologies has been extensively applied across various asset classes, yet the integration of these approaches within the context of collectible rum remains notably scarce. The scarcity of empirical research dedicated to the analysis of collectible rums underscores an intriguing research gap, compelling the need for an in-depth exploration into the determinants of collectible rum values and its viability as an investment asset.

It is pertinent to explore the extant academic literature concerning the dynamics of price determination in the realm of collecting for investment purposes, as these dynamics may exhibit nuances distinct from those observed in conventional asset markets.

This literature review aims to delve into existing studies employing the hedonic approach and risk-return analysis, setting the stage for the present thesis's empirical investigation into this problematic within the alternative assets landscape.

3.1 Hedonic approach

The hedonic approach represents a robust analytical methodology extensively utilized within the realms of economics and statistics to estimate the economic worth associated with specific attributes or characteristics inherent in a product or service. This methodological framework finds considerable application, particularly in the evaluation and assessment of goods or services endowed with multifaceted features or qualities. By adopting a hedonic pricing model, the approach dissects the total price of a product, dissecting it into its individual constituent attributes or characteristics. Each distinct attribute's contributory value to the overall worth of the product is meticulously quantified, elucidating its respective impact on the product's market value.

This method thus provides a comprehensive and structured mechanism for delineating the nuanced economic significance linked to various attributes or qualities inherent in a product or service, contributing significantly to informed pricing and valuation analyses within diverse economic landscapes.

For instance, Taylor (2003) applied a hedonic pricing model to ascertain the intrinsic value of vehicles by meticulously considering an array of attributes intrinsic to these automobiles. These attributes spanned a spectrum encompassing factors like engine type, transmission mode (manual or automatic), presence of air conditioning, and others of similar nature. Through an examination of a substantial dataset comprising numerous car transactions, economists and analysts endeavor to unravel and estimate the distinct value associated with each individual attribute.

This rigorous analysis aids in deciphering the intricate interplay between diverse characteristics and their respective impacts on the ultimate price determination within the automobile market. This methodological approach, rooted in the hedonic pricing model, not only facilitates a complex comprehension of the contributing worth of disparate attributes, but also offers valuable insights into the mechanisms governing consumer preferences and market dynamics within the automotive industry.

In the domain of collectibles, Chanel *et al.* (1992) undertook a comprehensive exploration into the viability of employing the hedonic methodology to evaluate paintings. Asserting the necessity of constructing price indices for paintings, the paper advocates for regression analyses that encompass the entirety of sales data, rather than solely relying on resale figures. While focusing specifically on paintings, the study underscores the idiosyncratic nature of the art market, characterized by its sporadic transactions and personalized negotiations, setting it apart from conventional financial markets. Despite prevailing claims regarding market efficiency within the art world, empirical validation encounters substantial hurdles, chiefly due to the irregularity inherent in trading activities. Consequently, the paper posits the utilization of the hedonic methodology across expansive datasets as a means to more thoroughly investigate the predictability of returns and assess market efficiency. This methodological approach promises to shed light on the intricate dynamics of similar markets with irregular trading activities, and contribute to a deeper understanding of its underlying mechanisms.

3.1.1 Wine

Wine, with its rich diversity of flavors and characteristics, is an excellent candidate for hedonic price estimation, given its inherent complexity. However, objectively assessing wine quality presents challenges due to subjective perceptions and the multifaceted factors influencing it, such as grape variety, terroir, and winemaking techniques. Despite these hurdles, the intricate nature of wine allows for detailed analyses that can capture consumer preferences and pricing trends. Thus, while the objective evaluation of wine quality remains challenging, the hedonic approach offers a valuable means of exploring its pricing dynamics.

The empirical evidence in the literature suggests that prices of differentiated products, such as wine, depend on various factors, including quality, reputation, and objective characteristics. However, Oczkowski (2001) argues that standard Ordinary Least Squares (OLS) procedures may distort the statistical significance of these attributes and the predictions of average prices. To address this issue, Oczkowski employed factor analysis and Two-Stage Least Squares (2SLS) estimation methods.

To overcome the objective quality limitation, Oczkowski treated quality and reputation as latent constructs and applied Confirmatory Factor Analysis (CFA) to supplement the estimation of hedonic price functions.

In the context of hedonic wine functions, Oczkowski categorized the wine attributes into sensory, chemical, objective, and climatic factors. Sensory attributes involve subjective assessments of a wine's characteristics, chemical attributes represent technical measures (such as acidity, alcohol volume, etc.), objective attributes include easily recognizable traits like vintage year and grape variety, and climatic attributes measure the impact of weather on grape production. Considering this division of attributes, Oczkowski summarized the previous literature as follows:

1. *When both individual sensory and objective characteristics are examined, the former tends to be insignificant and the latter significant.* (Pierre Combris & Visser 2000)

2. *When wine-guide overall sensory current quality scores are employed with objective characteristics, quality scores and objective traits are significant.* (Golan & Shalit 1993; Oczkowski 1994; Landon & Smith 1997; Schamel 1998; Wade 1999; Angulo *et al.* 2000; Pierre Combris & Visser 2000)

3. *When reputation (lagged quality) sensory measures appear with objective characteristics, reputation scores and objective traits are significant.* (Landon & Smith 1997; Wade 1999)

4. *When both reputation and current quality appear with objective characteristics, reputation, quality and objective traits are significant but reputation is far more economically important than quality.* (Landon & Smith 1998).

Therefore, evidence exist that significance of the attributes differ for different literature. The paper's main conclusion highlights the statistical insignificance of the quality variable, instead reputation is the main attribute which drives the price of wine.

Similar problematic was investigated by Luigi Benfratello & Sacchetto (2009). The authors delved into the reputation versus quality problem. Based on the conclusions drawn from the study, it becomes evident that the debate surrounding reputation versus quality in the context of wine pricing is nuanced. Through empirical analysis focusing on premium Italian red wines, specifically Barolo and Barbaresco, the research aimed to fill the gap in understanding this issue within the Italian wine market. By constructing a comprehensive database encompassing objective, sensorial, and reputation variables, the study shed light on the intricate dynamics influencing consumers' preferences and willingness to pay for wine.

The findings revealed that while both sensorial characteristics and reputation play crucial roles in shaping consumers' choices, reputation emerged as the dominant factor in driving market prices. The preference for wines and producers with established reputations underscores the significance of promotional activities aimed at bolstering reputation, such as participation in wine exhibitions and garnering citations in renowned guides. These insights have important implications for firms' strategic decisions, suggesting a shift towards cultivating a strong reputation alongside efforts to enhance taste quality.

Paroissien & Visser (2020) delved into another aspect affecting wine pricing: the influence of wine medals on producers' pricing strategies. Their objective was to gauge how winning these accolades in competitions impacts the prices set by producers. Employing a unique identification approach, they analyzed the causal effects of medals on prices, exploiting the timing of medal awards in relation to transaction dates. Their findings revealed that winning a medal could lead to a 13% price increase for producers. Furthermore, they observed that the impact of winning a gold medal was notably more pronounced than that of silver or bronze, although the correlation with quality across all medal types was not significantly different. Additionally, the study assessed the potential profits for producers participating in competitions, uncovering significant incentives for involvement. Lastly, they scrutinized the efficacy of competitions by evaluating the reliability of awarded medals as indicators of quality. Surprisingly, only a minority of competitions awarded medals that were significantly correlated with quality, particularly those with a longer establishment history where judges evaluated fewer wines per day.

Lastly, the study by Breeden & Liang (2017) delves into the intricate dynamics of wine prices in auctions, employing an age-period-cohort (APC) algorithm to dissect the factors influencing these dynamics. By scrutinizing a vast database comprising 1.5 million auction results, the researchers aimed to disentangle the effects of wine age, market conditions, and vintage specificity on price appreciation.

The findings not only illuminate the relationship between wine prices and their ages but also offer insights into the long-term potential for price appreciation. Additionally, the analysis elucidates the nonlinear relationship between wine ratings and prices. Furthermore, a comparative analysis across nine auction houses reveals significant price disparities for similar wines, underscoring the influence of auction dynamics on final prices.

3.1.2 Whiskey

Cordiez (2020) delved into the realm of collectible whiskey, examining the factors that determine their prices and their unique position as an alternative investments. This investigation is rooted in the understanding that collectible whisky satisfies the criteria of being a collectible, characterized by limited quantity, a finite period of production, and an appreciating market value.

The study emphasized that the nature of collecting is deeply rooted in the passion for acquiring and possessing items removed from ordinary use, often exceeding their intrinsic value. Collectors often preserve these items, sometimes with the intention of profiting from future sales, while others see them as family heirlooms. This unique perspective is aptly termed "emotional assets."

The world of collectible whisky has witnessed a surge in attention, primarily attributed to the increased accessibility through online auctions and its resilience to shocks in global financial markets. This surge has sparked curiosity about the risk-return profile of collectible whisky as well as its potential for diversification in comparison to traditional financial assets.

The study employed two distinct approaches to shed light on this subject: the hedonic regression, which explored the factors influencing the price of collectible whisky, and the Repeat-sales regression, which delved into the returns on these collectibles. The hedonic approach is particularly valuable in uncovering consumers' willingness to pay for various characteristics of collectible whisky. However, it should be noted that this approach assumes constant market valuation of these characteristics over time, which may not always hold true. Cordiez categorized characteristics of collectible whisky into four groups: Bottle Characteristics, Distillery Characteristics, Cask Characteristics, and Transaction Characteristics, and highlighted that, similar to many collectibles, whisky operates in a supply-driven economy, where scarcity factors positively influence price.

The Repeat-sales regression focused on financial return and risk hedging of strictly collectible whiskey (differentiating it from the retail market). The author also emphasised that evaluating the financial performance of collectibles remains a challenging task, often relying on the construction of specialized indices that can be sensitive to the data and methods used.

Ultimately, Cordiez aimed to construct an optimized portfolio using Mean Variance Analysis, striving for the highest risk-return level, also known as the Max Sharpe Ratio Portfolio. His research provides insight into the unique price determinants and characteristics of collectible whisky and its potential for diversification in an investment portfolio. It underscores the emotional connection that investors have with these assets, emphasizing the passion that often drives their acquisition.

This thesis adopts a parallel methodology. Initially, the investigation will delve into unraveling the factors influencing rum prices, employing the hedonic approach. This exploration will be conducted utilizing a rum dataset (chapter 4). Subsequently, the study will scrutinize rums as a viable alternative investment asset, drawing data from an auction house. It aims to assess the investment potential inherent in exclusive rum offerings within this niche market segment.

3.2 Rum as an investment asset

Another aspect influencing the determination of rum prices may lie in its collectible usability. Comparable phenomena have been documented in analogous collectible asset classes, indicating a potential corollary within the rum market. Consequently, this section endeavors to provide a comprehensive review of the extant academic literature pertaining to this phenomenon.

3.2.1 Collecting with investment intentions

From a psychological standpoint, the act of collecting often stems from a pursuit of closure, completion, or perfection. Danet & Katriel (1989) outline five strategies employed by collectors to achieve this objective: 1) finalizing a series or set; 2) utilizing physical spaces, such as adorning walls within their homes; 3) curating visually appealing and cohesive displays; 4) manipulating the scale of collected objects (e.g., collecting miniatures); and 5) aspiring to perfect their acquired items.

Notably, a significant segment of collectors seek expectations of financial gains. Research by Pearman *et al.* (1983), conducted among 154 collectors of antique and popular culture memorabilia, revealed that 35 percent identified investment as their primary motive for collecting. This motive ranked higher than options such as collecting for leisure or pleasure, engaging in collecting as a hobby, or valuing the preservation of antiquities. Additionally, in a broader survey encompassing various collector categories, 22 percent of respondents cited financial investment as a key motivation for their collecting endeavors. These insights reflect the multifaceted motivations behind collecting activities, where financial aspirations form a substantial driver for many enthusiasts in their pursuit of acquiring and preserving collectible items. Such inclinations toward financial gains within collecting hobbies underscore the potential allure of rum as an alternative asset for investment consideration.

Walgreen (2010) aimed to explore the potential benefits of including collectibles in investment portfolios from a financial perspective, without considering emotional, consumption, or aesthetic value. By examining the performance of collectible indices over the past 25 years, the researchers found significant variation in performance within the collectibles group, with Vintage Bordeaux wines outperforming most other assets and stamps showing underperformance. Collectibles exhibited low correlations with traditional asset classes, indicating their potential as diversifiers and hedges against specific risks, such as inflation. However, adding collectibles to an optimal mixed asset portfolio was not found to be beneficial in terms of efficiency, except for a small number of required returns.

During periods of financial turmoil, certain collectible categories, such as European 19th-century art and Old Masters, demonstrated significant diversification benefits and recession resistance. Overall, collectibles can serve various purposes, including diversification, risk hedging, and recession-proofing. Art investments are particularly beneficial for hedging exposure to other assets or specific events like recessions, while Vintage Bordeaux wines behave more like equity and offer profit opportunities. Building on these findings, the author will now seek to identify profit opportunities within the realm of rums.

3.2.2 Measuring the return of collectibles

Burton & Jacobsen (1999) defined three basic methods for calculating returns to collectibles:

1. **Composite Index Formation**: This involves curating sets of items whose prices are measured and averaged to create a composite index. However, a notable limitation of this approach arises from the disparity in items available for sale across different periods. Consequently, the index compiler must undertake some form of quality-standardization for the basket at each point in time.

2. **Hedonic Regression for Price Index**: This method employs a "hedonic" regression wherein an item's price is regressed against its diverse characteristics, encompassing factors like age or purchase price. While this technique accommodates quality differences among items in index

calculation, it restricts the rate of return to constancy over the data timeframe within a linear regression framework. This approach is utilized in the initial phase of this thesis to discern the influential factors driving rum price creation, excluding its application for return analysis.

3. **Repeat-sales Regression for Returns**: Utilizing data from Repeat-sales, this approach involves conducting a "Repeat-sale" regression. Unlike composite indexes with varying market baskets, Repeat-sales regressions manage quality control by leveraging price changes for specific items. However, a limitation arises as this method systematically selects only those items sold at least twice during the sample period, potentially disregarding relevant information from other sales transactions.

In pursuit of comprehensively investigating the risk-return profile of rum as a potential alternative investment asset, the author has strategically opted to employ the Repeat-sales regression methodology. This methodological selection is underpinned by its inherent alignment with the core aim of the thesis, which revolves around assessing the risk and return characteristics of rum within the realm of alternative investments.

By utilizing the Repeat-sales regression approach in conjunction with historical auction data, the study seeks to delve deeper into the intricate dynamics governing rum's market behavior. This methodological choice is driven by the method's efficacy in capturing changes in prices across multiple transactions, offering a robust framework to analyze the inherent risk and return attributes associated with rum as an investment option.

Chapter 4

Data

The lack of academic papers specifically focused on the topic of rums can be attributed primarily to the scarcity or absence of comprehensive datasets within the domain of rum production. Academic research heavily relies on data-driven analysis and empirical evidence to formulate hypotheses, conduct studies, and draw substantive conclusions. The limited availability of structured and comprehensive datasets concerning various facets of rum production, such as raw material sourcing, distillation methods, aging processes, regional variations, and market trends, complicates scholars' ability to conduct in-depth and rigorous studies in this field. The absence of this fundamental data infrastructure constrains the academic community from undertaking extensive research endeavors and publishing scholarly papers dedicated explicitly to the nuanced aspects of rum production and its broader socio-economic or cultural implications.

The absence of an available dataset specifically dedicated to rums may stem from several factors within the realm of data collection and organization within the spirits industry. Firstly, the spirits industry, including rum production, might not prioritize or extensively publish detailed datasets due to proprietary concerns or competitive reasons among distilleries. Additionally, the diversity in rum production, varying from region to region, distillery techniques, aging processes, and flavor profiles, might contribute to the challenge of compiling a comprehensive and standardized dataset. Moreover, the lack of a centralized authority overseeing rum production globally could lead to fragmented data sources, making it challenging to create a unified and exhaustive dataset encompassing the nuances of rum production. Lastly, the historical tradition of craftsmanship and artisanal approaches in rum-making might contribute to limited public dissemination of detailed data, as certain information could be considered trade secrets or closely guarded proprietary knowledge within the industry.

In the absence of accessible datasets, the acquisition of requisite data involves the utilization of web scraping techniques. This method entails the extraction of information from web pages programmatically, typically through the application of specialized software tools or scripts. Web scraping serves as an expedient approach to gather structured or unstructured data, enabling the collection of relevant information directly from online sources. The process involves accessing a webpage's HTML content, parsing its elements, and extracting desired data points, providing an avenue to obtain the necessary information when predefined datasets are unavailable or inaccessible.

This thesis works with two datasets, one for the Hedonic analysis of price determinants of rum, and one for the analysis of rum's usability as an alternative investment asset.

4.1 Hedonic analysis dataset

For the Hedonic analysis dataset this thesis endeavors to pioneer the creation of the inaugural rum dataset by employing web scraping techniques on the website *www.rum-x.com*. Through the systematic application of web scraping methodologies, this study aims to extract, compile, and organize comprehensive data pertaining to various facets of rum production, including but not limited to, distillery information, rum types, aging profiles, and regional classifications. By leveraging custom-built scripts, this research seeks to navigate the website's HTML structure, meticulously parsing its content to harvest relevant data points systematically. The collected information will undergo rigorous validation and structuring processes to ensure accuracy and consistency, ultimately culminating in the formation of the initial comprehensive dataset dedicated exclusively to the multifaceted realm of rum, thereby contributing significantly to the academic understanding and analysis of this intricate industry.

The initial dataset obtained in March 2024 by scraping contained information about 14,733 individual rum bottles and consisted of following variables:

- **ABV** - Numeric variable representing the Alcohol By Volume (ABV) percentage of the rum.

- **Age** - Numeric variable representing the age of the rum in years (time the rum spent in a barrel before being bottled).
- **Average price** - Numeric variable representing the average price of the rum bottle according to the Rum-X web page in euros.
- **Bottle volume** - Categorical variable representing the volume of the rum bottle.
- **Bottled** - Numeric variable representing the bottling year of the rum.
- **Bottler** - Categorical variable representing the entity or person who bottled the rum.
- **Brand** - Categorical variable representing the brand of the rum.
- **Category** - Categorical variable representing the category or type of rum (e.g., white rum, spiced rum).
- **Closed bottles** - Numeric variable indicating the number of bottles that were purchased by the Rum-X web users and remained closed.
- **Country** - Categorical variable representing the country of origin of the rum.
- **Distillation** - Categorical variable representing the method of distillation used in producing the rum (e.g., column still, pot still).
- **Distillery** - Categorical variable representing the distillery where the rum was produced.
- **Emptied bottles** - Numeric variable indicating the number of bottles that were purchased by the Rum-X web users and were emptied.
- **Made from** - Categorical variable representing the ingredients used to make the rum (e.g., Molasses, Sugar cane juice).
- **Mark** - Categorical variable representing rum labeling.
- **Name** - Categorical variable with the official name of the bottle.
- **No. of bottles** - Numeric variable indicating the total number of bottles produced of the specific rum.

- **Opened bottles** - Numeric variable indicating the number of bottles that were purchased by the Rum-X web users and were opened but not finished.
- **Open rate** - Numeric variable representing the ratio of opened and emptied bottles to the sum of all purchased bottles (closed, opened, and emptied).
- **Price range** - Categorical variable representing the price range of the rum (e.g., under 50€, above 250€).
- **Rating** - Numeric variable measuring the average of user ratings on the Rum-X web page. The scale is from zero to ten.
- **Starting price** - Numeric variable representing the cheapest price offer for the bottle listed on the Rum-X web page.
- **Vintage** - Numeric variable representing the year the rum was produced.

Regrettably, a subset of these variables proved unsuitable for subsequent analytical endeavors due to data deficiencies. Table 4.1 offers a comprehensive overview of the missing values attributed to each respective variable within the dataset.

The examination of data quality reveals a concerning prevalence of missing values across numerous variables. Indeed, a substantial proportion of the dataset is affected by these data deficiencies. Consequently, the variables "Bottled" and "Mark" were considered unsuitable for inclusion in subsequent analytical pursuits, given that over 90% of their respective data entries were unknown values. The author proceeded with a rigorous data cleaning process, delineating distinct steps for each variable as follows:

- **ABV** - Observations with missing values were excluded from the dataset.
- **Age** - Observations with missing values were excluded from the dataset.
- **Average price** - Observations with missing values were excluded from the dataset.
- **Bottle volume** - Only bottles with volume 70cl were kept in the dataset for relevant comparison.
- **Bottled** - Variable excluded due to missing data.

Table 4.1: Table of missing values - Hedonic analysis data.

Variable	Count of missing values
Bottled	13941
Mark	13531
Starting price	10251
Brand	10069
No. of bottles	9139
Vintage	7031
Bottler	6793
Distillation	5519
Age	4800
Category	4320
Distillery	3748
Open rate	3362
Price range	3235
Emptied bottles	2441
Opened bottles	2441
Closed bottles	2441
Made from	1900
Bottle volume	938
ABV	190
Name	1
Country	0
Average price (in €)	0
Rating	0

Notes: The table presents missing values within each of the variables in the Hedonic analysis dataset. Complete dataset contained 14 733 observations.

- **Bottler** - Variable excluded due to missing data and low theoretical potential for significant effect on the price.
- **Brand** - Observations with missing values were excluded from the dataset.
- **Category** - Observations with missing values were excluded from the dataset.
- **Closed bottles** - Variable excluded because of possible multicollinearity problem with Open rate which author considered more informative.
- **Country** - Observations with missing values were excluded from the dataset.

- **Distillation** - Observations with missing values were excluded from the dataset.
- **Distillery** - Observations with missing values were excluded from the dataset.
- **Emptied bottles** - Variable excluded because of possible multicollinearity problem with Open rate which author considered more informative.
- **Made from** - Observations with missing values were excluded from the dataset.
- **Mark** - Variable excluded due to missing data.
- **Name** - Variable excluded for irrelevancy.
- **No. of bottles** - Observations with missing values were excluded from the dataset.
- **Opened bottles** - Variable excluded because of possible multicollinearity problem with Open rate which author considered more informative.
- **Open rate** - Observations with missing values were excluded from the dataset.
- **Price range** - Variable excluded due to causality problem.
- **Rating** - Observations with missing values were excluded from the dataset.
- **Starting price** - Variable excluded due to causality problem and missing values.
- **Vintage** - Variable excluded due to missing values.

Further, all numeric variables were normalized by the Min-Max-normalization so that we keep the scales consistent and stable:

$$\text{Min-Max Normalization:} \quad X_{\text{normalized}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Following the execution of these methodical procedures, the resultant dataset culminated in a modest compilation of 538 observations. While this figure may seem diminutive relative to the initial dataset's expanse, it nonetheless constitutes a substantive and, arguably, representative subset within the rum industry.

4.2 Repeat-sales analysis dataset

In the pursuit of examining Repeat-sales analysis which delves into the potential of rum as a viable alternative investment instrument, the researcher encountered the recurring scarcity of readily available and comprehensive dataset. Therefore, the author once again used the systematic web scraping methodology. This dataset created in March 2024 and originated from the renowned Rum Auctioneer[1] online auction house website, known for its specialization in rum auctions.

Compared to the dataset obtained for the Hedonic analysis, the data procured from the Rum Auctioneer website exhibited a notable enhancement in quality (missing values in table 4.2), which could be attributed to the inherent practices and standards upheld by auction houses. Furthermore, the discernible refinement in data quality was also attributable to the relatively fewer and standardized features presented on the website. As a result, the original raw dataset, meticulously collected from the Rum Auctioneer platform, encompasses a total of 50,456 records detailing rum sales transactions containing these variables:

- **Age** - Numeric variable representing the age of the rum in years (time the rum spent in a barrel before being bottled).
- **Base** - Categorical variable representing the base ingredient used in the production of the rum (e.g., Molasses, Sugar cane juice).
- **Bottle Size** - Categorical variable representing the volume of the rum bottle.
- **Bottled Strength** - Numeric variable representing the Alcohol By Volume (ABV) percentage of the rum.
- **Bottler** - Categorical variable representing the entity or person who bottled the rum.
- **Cask Type** - Categorical variable representing the type of cask used for aging the rum (e.g., Barrel, Sherry).
- **Date** - Date variable representing the date of the auction or sale.

[1] Available at *www.rumauctioneer.com*

- **Distillery** - Categorical variable representing the distillery where the rum was produced.
- **Distillery Status** - Categorical variable representing the current status of the distillery (Operational, Closed).
- **Lot** - Categorical variable representing the lot number of the rum bottle in the auction.
- **Name** - Categorical variable with the official name of the bottle.
- **Production method** - Categorical variable representing the method of production used for the rum (e.g., Pot Still, Column Still).
- **Region** - Categorical variable representing the region where the rum was produced.
- **Vintage** - Numeric variable representing the year the rum was produced.
- **Winning Price** - Numeric variable representing the winning bid price of the rum bottle.

Table 4.2: Table of missing values - Repeat-sales data.

Variable	**Count of missing values**
Cask Type	28963
Vintage	17360
Age	15413
Production method	6615
Distillery	4704
Distillery Status	2871
Bottler	1119
Base	1041
Region	401
Bottled Strength	242
Bottle Size	68
Name	0
Winning Price	0
Lot	0
Date	0
Details	0

Notes: The table presents missing values within each of the variables in the Repeat-sales analysis dataset. Complete dataset contained 50,456 observations.

Regrettably, the author found that Lot numbers failed to adequately encapsulate the uniqueness of each bottle, contrary to initial assumptions. Consequently, to rectify this discrepancy, identical bottles were identified based on uniform values across all variables, with the exception of Date, Lot, and Winning Price, which inherently varied for each bottle. Following the identification of duplicate rows, the author curated a dataset specifically tailored for the analysis of Repeat-sales, comprising the subsequent variables:

- **Name with size** - This categorical variable ensures comparability among similar bottles by considering only those with identical volumes.
- **First Price** - A numeric variable denoting the initial purchase price of the bottle.
- **Second Price** - A numeric variable indicating the subsequent purchase price of the bottle.
- **Date of First Purchase** - Date variable representing the date when the bottle was first purchased.
- **Date of Second Purchase** - Date variable representing the date when a similar bottle was resold.
- **Holding Period** - A numeric variable representing the interval, measured in months, between the initial and subsequent purchase dates.
- **First Price log** - A numeric variable representing the natural logarithm of the initial purchase price.
- **Second Price log** - A numeric variable representing the natural logarithm of the subsequent purchase price.
- **Return** - A numeric variable denoting the return on investment or the percentage change in price between the first and second transactions.

This dataset formed the basis for subsequent analyses of repeated sales, encompassing 21,457 resale observations, a notably extensive corpus for such investigations.

Chapter 5

Methodology

5.1 Hedonic analysis

This section delineates the methodological framework employed to identify the principal determinants of rum prices, and works with the Hedonic analysis dataset 4.1.

In response to the complexity and potentially high correlation among variables within the dataset (particularly emphasizing the categorical variables associated with the geographical attributes of the rum, such as Country, Distillery, and Brand), the author opted for a manual stepwise regression approach. Traditionally, stepwise regression involves iteratively adding or removing variables from the model based on predefined criteria aimed at optimizing goodness-of-fit metrics. However, due to the intricate nature of the data, relying solely on termination criteria might not offer a holistic understanding of the model's behavior.

To address this challenge, the author adopted a hands-on approach, initially constructing simple regression models and progressively augmenting their complexity. Throughout this manual process, the author carefully observed variations in model parameters and evaluated multiple measures of fit, including R-squared and adjusted R-squared. By systematically adjusting the composition of the models and closely monitoring the impact on these fit indices, the author sought to gain a comprehensive understanding of how different variables contribute to model performance.

This methodological adaptation allowed for a nuanced exploration of the dataset's complexities and facilitated the identification of meaningful predictors while mitigating the potential pitfalls of automated variable selection algorithms. Through this iterative refinement process, the author aimed to develop robust regression models that effectively captured the underlying relationships within the data.

The objective of this analysis is to discern the principal determinants influencing the price of rum. Consequently, the author adjudged that employing an Ordinary Least Squares model is the most apt methodology. However, to ensure a comprehensive exploration of the issue at hand, the author further integrated Bayesian Model Averaging (BMA) and Lasso regression techniques into the analytical framework.

The Ordinary Least Squares model is formulated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \tag{5.1}$$

where:

- Y represents the dependent variable (price of rum),
- β_0 denotes the intercept term,
- $\beta_1, \beta_2, ..., \beta_p$ are the coefficients of the predictor variables $X_1, X_2, ..., X_p$,
- $X_1, X_2, ..., X_p$ represent the independent variables,
- ε represents the error term.

Bayesian Model Averaging, as proposed by Steel (2020), incorporates model uncertainty by considering multiple regression models simultaneously. The BMA approach involves estimating the posterior model probabilities and combining the predictions from different models weighted by their posterior probabilities.

Bayesian Model Averaging involves calculating the weighted average of model-specific parameter estimates. If we denote the parameters of the i-th model as β_i, and the posterior probability of the i-th model given the data as $P(M_i|D)$. Then the combined estimate $\hat{\beta}$ using BMA can be expressed as:

$$\hat{\beta} = \sum_{i=1}^{K} P(M_i|D) \cdot \hat{\beta}_i$$

where:

- K represents the total number of models.
- $\hat{\beta}_i$ is the parameter estimate from the i-th model.

- $P(M_i|D)$ is the posterior probability of the i-th model given the observed data D.

This formula essentially combines the parameter estimates from different models, weighted by their posterior probabilities. It allows us to incorporate uncertainty about which model is the true underlying model. The comprehensive application of the BMA approach across the entirety of the dataset encounters practical limitations due to the substantial number of categories within the categorical variables. The computational complexity arising from the vast array of potential models exceeds the computational resources at the author's disposal. Consequently, the author elected to employ the BMA approach exclusively on the numerical variables, acknowledging its computational feasibility within this subset. This strategic decision enabled a comparative evaluation with the OLS approach, thereby facilitating a nuanced assessment of modeling techniques.

In response to the challenge posed by the high dimensionality of the dataset, the author endeavored to mitigate the issue by adopting a regularization technique. The objective of regularization is to impose constraints on the model parameters, thereby restraining the proliferation of effects and alleviating the burden of excessive variable inclusion.

Lasso regression is a regularization technique that involves minimizing the sum of squared residuals with an additional penalty term on the absolute size of the coefficients, effectively shrinking some coefficients to zero. The Lasso model aims to promote sparsity in the coefficient estimates, thus facilitating variable selection and reducing overfitting. Mathematically, the Lasso regression can be formulated as follows:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{5.2}$$

where:

- n represents the number of observations.
- p represents the number of predictor variables.
- y_i represents the observed value of the dependent variable for the i-th observation.
- x_{ij} represents the value of the j-th predictor variable for the i-th observation.

- β_0 represents the intercept term.
- β_j represents the coefficient of the j-th predictor variable.
- λ represents the regularization parameter, controlling the strength of the penalty on the absolute size of the coefficients.

The utilization of these complementary methodologies allows for a robust and multifaceted examination of the factors influencing rum prices, thereby enhancing the depth and reliability of the analysis.

5.2 Repeat-sales analysis

This section describes the methodology which aims to answer the research question about the viability of rums as an alternative investment asset. This segment of the thesis is predicated upon the examination of the Repeat-sales Dataset 4.2.

Following a similar methodology as Cordiez (2020), the analysis initiates with the computation of the logarithm of price differentials for each resale, a procedure delineated during dataset generation (see Section 4.2). Subsequently, the dataset undergoes transformation to adhere to the framework of the standard Repeat-sales model, as expounded by Martin J. Bailey & Nourse (1963), under the assumptions of a zero-mean error term and absence of heteroskedasticity.

$$r_i^t = ln\left(\frac{p_i^t}{p_i^s}\right) = \sum_{t=0}^{T} \beta^t D_i^t + \epsilon_i^t \tag{5.3}$$

where:

- r_i^t represents the return on the resale.
- p_i^t and p_i^s represent the prices of the second and initial purchases respectively.
- D_i^t denotes an indicator variable set to 1 during the period of resale, -1 during the first transaction involving the bottle, and 0 otherwise.
- ϵ_i^t denotes the stochastic error term.

Nevertheless, in theoretical terms, this model encounters heteroscedasticity due to the phenomenon wherein the increased duration between the first and second purchase (referred in data to as the Holding period) correlates with augmented return volatility. In response to this concern, the author employed Weighted Least Squares (WLS) regression methodology to address this inherent issue. To execute WLS, it is imperative to initially conduct a regression analysis of the Holding period against the squared residuals (ϵ_i^2) derived from the original equation 5.3.

$$\epsilon_i^2 = \delta_0 + \sum_{t=1}^{T} \delta^t HP_i^t + \mu_i^t \tag{5.4}$$

where:

- δ_0 denotes the intercept term.
- HP_i^t represents the holding period, which signifies the temporal gap in months between the initial and subsequent purchases.
- D_i^t denotes an indicator variable set to 1 during the period of resale, -1 during the first transaction involving the bottle, and 0 otherwise.
- μ_i^t denotes the stochastic error component.

Upon completing the estimation of δ^t across all time periods, the fitted values of squared residuals are calculated and subjected to square root transformation. These transformed values are subsequently utilized as weights within the framework of WLS.

$$W_i = \sqrt{\hat{\epsilon}_i^2} \tag{5.5}$$

We use these weights to divide each observation by its corresponding weight and fit the initial regression (5.3) again:

$$\tilde{r}_i^t = \sum_{t=0}^{T} \tilde{\beta}^t \tilde{D}_i^t + \tilde{\epsilon}_i^t, \tag{5.6}$$

Subsequently, we advance our analysis by applying the exponential function to the estimated betas to ensure interpretability.

$$\Pi^t = \exp\left(\tilde{\beta}^t\right) \tag{5.7}$$

And finally, creating the Repeat-sales index for each period in the dataset:

$$R^t = \frac{\Pi^t}{\Pi^{t-1}} - 1 \tag{5.8}$$

Chapter 6

Results

6.1 Hedonic analysis results

The subsequent section delves into the outcomes derived from the Hedonic analysis, dissecting the factors influencing the pricing dynamics of rum.

The Hedonic analysis was conducted with multiple variations of regressions described in the 5.1 section. Three main approaches were utilized: Bayesian Model Averaging, Stepwise regression, and Lasso regression.

6.1.1 Bayesian Model Averaging results

First, we direct our attention to the Bayesian Model Averaging procedure. It is noteworthy that solely numerical variables were incorporated into the analysis. This choice stemmed from the impracticality of conducting the averaging process on all categorical variables, a decision influenced by limitations in computational resources.

From the results presented in Table 6.1, it's evident that three variables exhibit a notably high likelihood of significance: ABV, Age and Open Rate exceed a 97% probability threshold. The observed signs of the coefficients for all three variables align with the anticipated theoretical expectations of the author. Specifically, an increase in the aging duration of rum correlates positively with higher prices, indicative of the premium placed on aged varieties. This observation is likely due to the higher costs connected to the prolonged aging process such as storage costs and evaporation loss. Furthermore, elevated alcohol volume demonstrates a positive correlation with pricing, implying a valuation premium for beverages with higher alcohol content within this context.

Table 6.1: Bayesian Model Averaging for numerical variables

Variable Name	Probability	Avg. Coefficient
Constant	0.056802	-0.000204
ABV	0.978124	0.063205
Age	1.0	0.226037
No. of bottles	0.267849	0.030704
Open rate	0.973528	-0.05038
Rating	0.055176	0.000229

Notes: The table presents descriptive statistics derived from the outcomes of the Bayesian Model Averaging regression. Within the summary, the "Probability" column denotes the likelihood of a variable exhibiting statistical significance, while the "Average Coefficient" column illustrates the mean value of the regression coefficients (represented by beta).

Conversely, a higher open rate is associated with decreased prices, suggesting that more expensive bottles tend to be collected rather than consumed.

The constant variable appears to have minimal impact, and the explanatory power of Rating for price determination appears very weak. Surprisingly, the Number of Bottles also demonstrates a relatively low probability of influence, contrary to the authors' expectations.

It is imperative to maintain a cautious perspective regarding the interpretative scope of these findings. Notably, the absence of categorical variables from the analysis underscores the incomplete statistical rigor inherent in the present results. Consequently, the current analysis may not comprehensively encapsulate the multifaceted dynamics underlying the phenomenon under investigation.

6.1.2 Stepwise regression results

Based on the outcomes derived from the Bayesian Model Averaging regression analysis, the researcher embarked upon an initial exploration of the significance and coefficient magnitudes associated with each numerical variable. Subsequently, the researcher proceeded to conduct a stepwise regression analysis.

The first step involved regressing each individual variable independently against the price variable. The results are available in the table 6.2.

The outcomes from the single variable regression provide insights into the individual contributions of each variable towards explaining the underlying data dynamics. Notably, there are differences between the outcomes of the numeric

Table 6.2: Results for single variable regressions

Variable name	Coefficient	F-statistics p-value	R-Squared	Adjusted R-Squared
ABV	0.0375	0.0931	0.005	0.003
Age	0.2236	1.13e-23	0.171	0.170
No. of Bottles	0.0565	0.369	0.002	-0.000
Open Rate	-0.0703	4.71e-05	0.030	0.029
Rating	0.1461	3.25e-06	0.040	0.038
Brand	Not Applicable	1.37e-20	0.426	0.313
Category	Not Applicable	0.0942	0.025	0.010
Country	Not Applicable	1.32e-09	0.188	0.133
Distillation	Not Applicable	0.0613	0.035	0.015
Distillery	Not Applicable	8.57e-14	0.367	0.243
Made from	Not Applicable	0.656	0.003	-0.003

Notes: The table presents the outcomes derived from individual variable regressions conducted in isolation with respect to the price variable. Each regression model was structured to include an intercept. Notably, for categorical variables, which encompassed multiple coefficients, the coefficient column is deemed inapplicable.

variables from these regressions and those derived from the Bayesian Model Averaging analysis (see table 6.1).

Specifically, the variable ABV appears to have descended below the threshold of statistical significance and shows low ability to explain the observed data patterns. Age persists as a significant predictor and shows a very similar coefficient as in the BMA results. Number of Bottles remains below of statistical significance. Remarkably, Open rate appears to possess some explanatory efficacy, although notably inferior to Age. Intriguingly, Rating emerges as a more influential predictor than Open rate in this context, a departure from the hierarchical importance depicted within the BMA results.

Regarding the categorical variables, it is noteworthy to commence by acknowledging the number of categories within each variable, as displayed in Table 6.3.

A higher count of categories within a variable possesses the potential to enhance the explanatory capacity of the model, particularly evidenced by the R-squared metric. This augmentation stems from the increased variability afforded to the model, facilitating a more nuanced portrayal of the underlying data dynamics. Consequently, it becomes imperative to scrutinize the adjusted R-squared metric, which serves as a corrective measure, accounting for the inclusion of additional variables, or in this context, categories within the model. Such an analytical approach ensures a more judicious evaluation of the model's

Table 6.3: Number of categories for categorical variables

Variable name	**Number of categories**
Brand	89
Category	9
Country	35
Distillation	12
Distillery	89
Made from	4

Notes: The table shows number of categories for each of the categorical variables.

explanatory power.

With the foregoing considerations in mind, let us now undertake a more detailed examination of the outcomes pertaining to the categorical variables.

It becomes evident that the variable "Brand" emerges as the foremost contributor to explicating the dataset. Evidenced by a commendable R-squared value of 0.426 and an adjusted R-squared of 0.313, it manifests a formidable explanatory capacity. It is noteworthy, however, that such explanatory power is intricately linked to its expansive representation, encompassing a considerable 89 categories.

Conversely, while the variable "Category" exhibits some discernible explanatory prowess, as evinced by an R-squared value of 0.025 and an adjusted R-squared of 0.010, its significance is tempered by an F-statistics p-value that falls below the threshold of statistical significance, suggesting a limited influence on the price dynamics.

In contrast, "Country" emerges as a noteworthy predictor, demonstrating an R-squared value of 0.188 and an adjusted R-squared of 0.133. Remarkably, this predictive efficacy is underscored by the realization that "Country" encompasses less than half the number of categories present in the "Brand" variable.

"Distillation," although marginally failing to surpass the threshold of statistical significance as indicated by the F-statistics, nonetheless exhibits some degree of explanatory capability, despite comprising a mere twelve categories.

"Distillery" shows great explanatory power but considering that it has the same count of categories it falls behind the Brand variable greatly.

Lastly, the "Made from" variable evinces negligible explanatory power, likely attributable to the simplicity of the model and the limited scope of its representation, with only four categories considered within the variable.

After conducting the initial simplest regressions, the subsequent step involves amalgamating the variables into a unified model to achieve the optimal fit. Nevertheless, a significant risk of collinearity arises, particularly among the categorical variables. To assess this collinearity, the author employed the Chi-squared test for independence to examine each pair of categorical variables. The result of this test is shown in table 6.4.

Table 6.4: P-values of Chi-Squared test for independence

	Country	**Distillery**	**Brand**	**Category**	**Made from**	**Distillation**
Country		0.0	0.0	3.0e-100	2.7e-95	5.0e-211
Distillery	0.0		0.0	3.5e-24	1.8e-61	2.2e-237
Brand	0.0	0.0		3.5e-78	4.4e-149	5.0e-295
Category	3.0e-100	3.5e-24	3.5e-78		1.3e-24	3.3e-53
Made from	2.7e-95	1.8e-61	4.4e-149	1.3e-24		5.4e-229
Distillation	5.0e-211	2.1e-237	5.0e-295	3.3e-53	5.4e-229	

Notes: The table shows p-values of pairwise Chi-Squared test for independence. In this scenario, the null hypothesis posits that the two variables exhibit correlation, thereby implying a correlation among all pairs of variables.

Indeed, the author expected to discover evidence of collinearity but, unexpectedly, all pairs of categorical variables exhibited highly significant correlation, highlighting the pervasive nature of this phenomenon within the rum data.

This finding significantly complicated subsequent research steps, given the limited array of viable options to address such pervasive multicollinearity. Consequently, the author proceeded with meticulous model construction, keeping multicollinearity at the forefront of considerations. Subsequently, the decision was made to explore regularization techniques, with Lasso regression being the method of choice (see 5.2).

The author thus embarked upon the integration of multiple variables into comprehensive models, consistently scrutinizing the indicators of model fit. Tables 6.5 and 6.6 present an illustrative selection of the models examined. However, it is pertinent to note that the comprehensive analysis encompassed a broader spectrum of models beyond those specifically delineated.

The findings underscore the absence of a singular model that optimally explicates the dataset. As previously noted, pervasive multicollinearity within the dataset cautions against reliance on any single model for interpretation, given the potential for considerable misinterpretation. However, a comprehensive

Table 6.5: Results for multiple variable regressions

Model	Only numerics			All variables			Numerics + Brand		
Variable	Used	Coef	P-val	Used	Coef	P-val	Used	Coef	P-val
Constant	•	-0.012	0.623	•	0.086	0.361	•	0.005	0.904
ABV	•	0.078	0.001	•	0.092	0.003	•	0.093	0.000
Age	•	0.231	0.000	•	0.221	0.000	•	0.207	0.000
No. of Bottles	•	0.132	0.027	•	-0.123	0.076	•	-0.089	0.177
Open Rate	•	-0.046	0.004	•	-0.029	0.111	•	-0.025	0.111
Rating	•	0.002	0.947	•	-0.058	0.191	•	-0.005	0.895
Brand				•	NA	NA	•	NA	NA
Category				•	NA	NA			
Country				•	NA	NA			
Distillation				•	NA	NA			
Distillery				•	NA	NA			
Made from				•	NA	NA			
Measures	R^2	AR^2	Fscore	R^2	AR^2	Fscore	R^2	AR^2	Fscore
	0.213	0.205	28.74	0.593	0.395	2.990	0.516	0.415	5.095

Notes: The table presents the outcomes derived from a subset of models employed in the examination of the determinants influencing the pricing dynamics of rum. In each instance, price is elucidated as the dependent variable. The constituent models delineate the employed variables, with numerical attributes accompanied by their respective coefficients and associated p-values, alongside measures assessing the models' goodness of fit.

Table 6.6: Results for multiple variable regressions

Model	Combination 1			Combination 2			Combination 3		
Variable	Used	Coef	P-val	Used	Coef	P-val	Used	Coef	P-val
Constant	•	-0.163	0.082	•	-0.064	0.236	•	-0.010	0.854
ABV	•	0.101	0.000	•	0.099	0.000	•	0.098	0.000
Age	•	0.221	0.000	•	0.257	0.000	•	0.273	0.000
No. of Bottles							•	0.026	0.646
Open Rate	•	-0.023	0.150	•	-0.031	0.054	•	-0.037	0.019
Rating							•	0.003	0.929
Brand	•	NA	NA						
Category									
Country							•	NA	NA
Distillation							•	NA	NA
Distillery				•	NA	NA			
Made from	•	NA	NA						
Measures	R^2	AR^2	Fscore	R^2	AR^2	Fscore	R^2	AR^2	Fscore
	0.518	0.416	5.063	0.485	0.381	4.625	0.427	0.368	7.248

Notes: The table presents the outcomes derived from a subset of models employed in the examination of the determinants influencing the pricing dynamics of rum. In each instance, price is elucidated as the dependent variable. The constituent models delineate the employed variables, with numerical attributes accompanied by their respective coefficients and associated p-values, alongside measures assessing the models' goodness of fit.

examination across multiple models can unveil discernible explanatory trends inherent within each variable.

The constant term consistently exhibited insignificance, likely attributable to the fairly large number of variables.

Across all models, the Alcohol by Volume (ABV) exerted a consistently significant positive influence on pricing, suggesting a robust association between higher alcohol content and elevated prices. This aligns closely with the Bayesian Model Averaging estimation, which similarly identified ABV as a highly probable significant determinant.

Similarly as in the BMA regression, Age emerged as one of the main drivers of price, manifesting a notably significant positive coefficient across all models, underscoring its paramount role in data explanation.

Conversely, the variable "Number of bottles" exhibited erratic behavior, with predominantly insignificant coefficients and fluctuating signs, aligning with the conclusions drawn from the BMA analysis, implying its limited impact on rum pricing dynamics.

"Open rate" consistently demonstrated a negative trend but the p-value oscillated closely around the significance threshold. We can conclude that this effect has some explanatory capacity suggesting that bottles which remain closed after being purchased tend to be the more expensive ones. While the coefficient's sign and magnitude closely resembled those of the Bayesian Model Averaging, which regarded it as highly probable for significance in the absence of categorical variables.

Rating remained highly insignificant, underscoring its subordinate role in price determination. This conclusion could be explained by the subjective nature of this variables compared to the other more objective ones.

The Brand variable emerged as a pivotal determinant, notably augmenting the explanatory power of models, evidenced by the marked increase in adjusted R-squared when incorporated, particularly evident in the "Numerics + Brand" model (see table 6.5). Notably, brands such as Don Papa and La Flibuste exhibited significant positive coefficients across various models.

While the "Category" variable marginally improved the fit when added to the "Only numerics" model, its combined effect with other categorical variables tended to diminish the adjusted R-squared. The author concluded that the effect of "Category" is most probably already included in other categorical variables.

Similarly, the "Country" variable demonstrated explanatory potential in isolation, but combined with "Brand" variable it increased only the R-Squared while adjusted R-Squared remained unchanged, indicating that the effect of "Country" is also already included in "Brand."

The impact of "Distillation" on pricing proved negligible across various model combinations, suggesting a minimal effect on price determination.

The "Distillery" variable mirrored the explanatory power of the "Brand" variable with slightly inferior model fit improvement. When integrated with "Brand," its additional explanatory value remained negligible, reinforcing the dominance of "Brand" in explaining price dynamics.

Lastly, the "Made from" variable marginally enhanced model fit upon inclusion but lacked substantial impact on price determination.

In light of the pronounced collinearity, attempts to explore interaction terms between categorical variables yielded no appreciable improvement in adjusted R-squared, prompting the decision to abstain from further reporting on these findings beyond this commentary.

6.1.3 Lasso regression results

The final methodological strategy employed by the researcher for price determination involved Lasso regression. This particular approach was chosen to mitigate issues stemming from multicollinearity.

Tables 6.7 and 6.8 delineate the outcomes of the Lasso regression utilizing identical model specifications as Tables 6.5 and 6.6.

The findings derived from the Lasso regression analysis reveal nuanced disparities when compared with preceding models while maintaining fundamental consistencies.

The persistent insignificance of the constant term underscores its negligible impact across all iterations.

An intriguing observation manifests with the ABV variable within the Lasso framework; its explanatory efficacy appears diminished upon the inclusion of the Brand variable. Nonetheless, ABV retains its salience when the Brand factor is omitted from the analysis. Once again, it appears that the "Brand" variable encompasses aggregated information from majority of other variables within the dataset.

The Age coefficient retains its positive and statistically significant nature across all model specifications with a slight diminution in coefficient's magnitude.

Table 6.7: Results for Lasso regressions

Model	Only numerics		All variables		Numerics + Brand	
Variable	Used	Coef	Used	Coef	Used	Coef
Constant	•	0	•	0	•	0
ABV	•	0.043	•	0	•	0
Age	•	0.205	•	0.147	•	0.143
No. of Bottles	•	0	•	0	•	0
Open Rate	•	-0.043	•	-0.021	•	-0.020
Rating	•	0	•	0	•	0
Brand			•	NA	•	NA
Category			•	NA		
Country			•	NA		
Distillation			•	NA		
Distillery			•	NA		
Made from			•	NA		
Measures	R^2	AR^2	R^2	AR^2	R^2	AR^2
	0.202	0.193	0.316	-0.228	0.29	0.139

Notes: The table presents the outcomes derived from a subset of models employed in the examination of the determinants influencing the pricing dynamics of rum using a Lasso regression. In each instance, price is the dependent variable. Numerical attributes are accompanied by their respective coefficients.

Table 6.8: Results for Lasso regressions

Model	Combination 1		Combination 2		Combination 3	
Variable	Used	Coef	Used	Coef	Used	Coef
Constant	•	0	•	0	•	0
ABV	•	0	•	0.020	•	0.010
Age	•	0.146	•	0.169	•	0.171
No. of Bottles					•	0
Open Rate	•	-0.022	•	-0.026	•	-0.028
Rating					•	0
Brand	•	NA				
Category						
Country					•	NA
Distillation					•	NA
Distillery			•	NA		
Made from	•	NA				
Measures	R^2	AR^2	R^2	AR^2	R^2	AR^2
	0.309	0.160	0.293	0.147	0.24	0.160

Notes: The table presents the outcomes derived from a subset of models employed in the examination of the determinants influencing the pricing dynamics of rum using a Lasso regression. In each instance, price is the dependent variable. Numerical attributes are accompanied by their respective coefficients.

Conversely, the variable denoting the number of bottles exhibits consistent insignificance across all Lasso models, suggesting its marginal relevance in explicating price variations. This observation aligns with its low significance levels in BMA.

In contrast, the Open rate variable showed significance across all Lasso regressions, characterized by a negative coefficient. This contrasts with its previous results in the unregularized models.

The Rating variable continues to evince insignificance throughout.

Regarding categorical variables, their effects remain largely similar with those observed previously. However, the overall model fit is noticeably diminished due to the constrained number of active variables, a characteristic inherent to Lasso regression's variable selection mechanism.

In the realm of categorical variables, certain categories emerge as statistically significant across multiple models. Noteworthy among these are brands such as Don Papa, La Flibuste, Plantation, and Wild Series Rum, alongside distilleries like Bleeding Heart Rum Company, Caroni, Hampden, and La Favorite. It is noteworthy to mention that Don Papa is exclusively produced by the Bleeding Heart Rum Company, and La Flibuste is produced in La Favorite, thereby amalgamating these categories into singular entities.

The variable indicating the utilization of sugar cane juice in production emerges as significantly positive. Conversely, the adoption of the Pot Still distillation method consistently exhibits a negative association with price.

In synthesizing the disparities between regularized and unregularized models, it becomes evident that Lasso regression yields diminished goodness-of-fit metrics, a consequence of its inherent penalization mechanism. While the coefficients' magnitudes experienced slight adjustments, their directional signs persisted across both frameworks. Notably, the inclusion of the Brand variable in Lasso rendered ABV insignificant, highlighting a nuanced interplay between these predictors. Conversely, the variable "No. of bottles" failed to attain statistical significance within the Lasso framework. Intriguingly, Open rate emerged as a significant predictor within the Lasso model.

Following the comprehensive analysis of the Hedonic model outcomes, several conclusions emerge. Primarily, the absence of a singular model perfectly aligning with the dataset stems predominantly from the inherent multicollinearity present within the data.

Among the variables examined, Age and Brand emerge as the most influential factors in elucidating price dynamics. The positive correlation between Age and price likely stems from heightened expenses associated with rum aging, encompassing storage costs and evaporation losses (see section 2.4).

The influence of Brand on price unveils a complexity surpassing that of Age. The study posits that the Brand variable intricately interweaves with other pertinent factors, insinuating that the brand name inherently encapsulates a plethora of information, including the country of origin, distillery, production methods, associated costs, and reputation. This theoretical proposition underscores the necessity for empirical validation utilizing more extensive and robust datasets. Such validation not only holds promise for refining our understanding of the intricate dynamics at play but also beckons as a pivotal avenue for future research endeavors, promising deeper insights into the nuanced interplay between brand identity and pricing dynamics within the rum market.

As stated above, other variables' explanatory power consistently paled in comparison to Age and Brand. However, some repeated patterns emerged in the results which could reveal some tendencies of the price dynamics. For example, higher ABV showed the tendency to increase the price of the bottle. Additionally, the analysis revealed a negative relationship between Open rate and price, indicating that bottles which are purchased but not opened exert upward pressure on pricing. This phenomenon may be explained by the collectible nature associated with unopened bottles. When a consumer purchases a bottle of rum without intending to open it, it suggests a deliberate decision to abstain from regular consumption. Instead, such acquisitions often signify an inclination towards collecting or viewing the bottle as an investment opportunity, which could lead to emotional purchases connected to increased prices, as discussed in section 3.2.1.

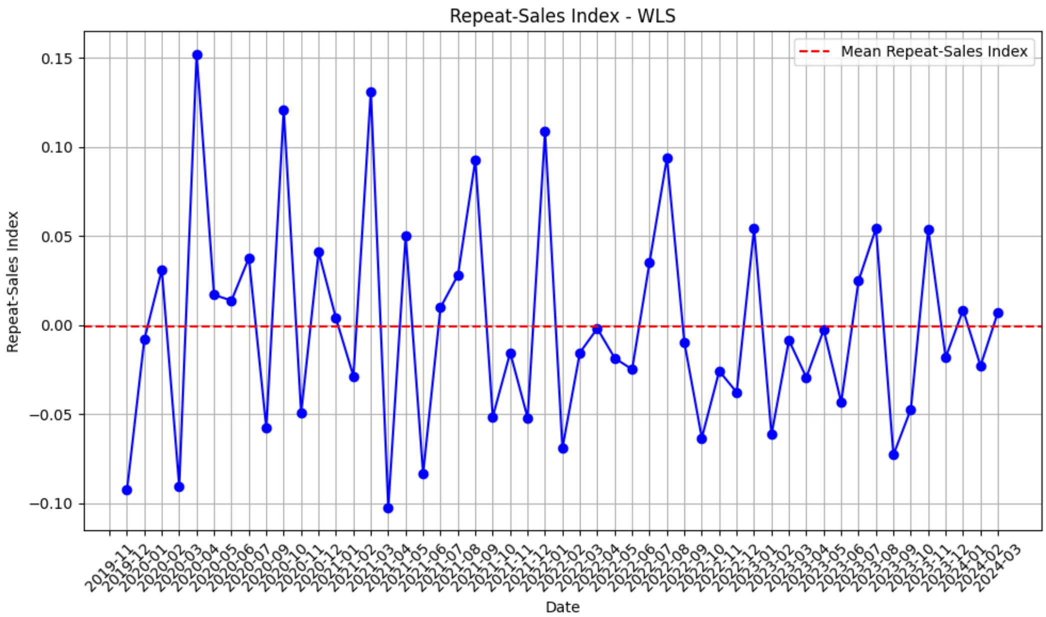
To author's surprise, the results do not share similar conclusions about the effect of scarcity (represented by the No. of bottles variable) on price, as Cordiez (2020) found in his whiskey study. Contrary to the whiskey research, the current rum analysis does not corroborate significant effects of scarcity on price.

Similarly, the variable of Rating failed to demonstrate explanatory efficacy. The author posits that this observation may be attributable to the subjective and inconsistent nature inherent to the Rating variable.

6.2 Repeat-sales analysis results

The following section thoroughly explores the findings obtained from the Repeat-sales analysis, intricately examining the various factors that shape the pricing trends of rum over time. The Repeat-sales analysis was carried out through the creation of the Repeat-sales index outlined in the methodology section (see 5.2).

Figure 6.1: Repeat-sales index



Notes: The figure depicts the Repeat-sales index created by utilization of Weighted Least Squares methodology. Red line represents the mean value.

Figure 6.1 depicts the Repeat-sales index spanning from November 2019 to March 2024 created by the WLS with weights obtained from the Holding period (see the methodology 5.4), while Table 6.9 presents a comprehensive overview of the index's statistical summary.

Table 6.9: Statistical Summary for Repeat-sales index

Measure	**Value**
Mean	-0.0006
Median	-0.0084
Standard Deviation	0.0587
Variance	0.0034
Skewness	0.6090
Kurtosis	-0.0225

The calculated mean value of -0.0006 portrays a near-zero central tendency, indicative of the index's tendency to concentrate around this central point on average. Furthermore, the median value of -0.0084, marginally lower than the mean, suggests a subtle leftward skewness in the distribution. The relatively diminutive standard deviation (0.0587) and variance (0.0034) align with these findings, underscoring a narrow dispersion of data points around the mean.

Of particular interest, the skewness coefficient of 0.6090 unveils a moderate right skewness, revealing a slight inclination towards higher values. This departure from symmetry is further underscored by the kurtosis coefficient of -0.0225, indicating a distribution with slightly flattened tails and fewer outliers compared to a standard normal distribution.

These observations offer compelling evidence that the overall trend in the auction sales of rum does not significantly appreciate over time. This is contrary to the conventional findings associated with collectible assets. While rum shares certain characteristics with whiskey, particularly in the realm of collectibility, it does not consistently exhibit the appreciating market value typically associated with such assets, as delineated by Cordiez (2020). The Repeat-sales index, which serves as a barometer of price fluctuation over time, portrays a landscape where consistent price growth is notably absent. On average, the index hovers near zero, suggesting a lack of sustained upward movement in rum prices.

The differing investment returns between rum and whiskey could stem from several interrelated factors. A crucial factor contributing to the divergence in investment returns is brand recognition and reputation. Certain whiskey brands have built a strong reputation and brand equity over decades or even centuries, commanding higher prices in the market. This brand recognition, coupled with a history of price appreciation in the whiskey market, can attract more investors and collectors, further driving up prices. While rum brands also have their loyal followings, they may not have achieved the same level of recognition or prestige in the investment market, influencing their investment returns.

The realm of spirit investment is deeply entwined with the intricate dynamics of market demand and prevailing trends, which wield substantial influence over the potential returns on such investments. The fluctuating landscape of consumer preferences and evolving market trends serves as a pivotal determinant in shaping the performance trajectories of spirits as viable investment assets. Within this context, the contrast between the whiskey and rum markets emerges as particularly noteworthy.

Notably, the whiskey market commands a significantly larger share of the global spirits landscape when compared against the realm of rum. A comprehensive analysis of market data reveals that the global revenue attributed to rum amounted to US$17.0 billion in the fiscal year of 2024[1] , whereas the corresponding figure for whiskey soared to US$92.9 billion within the same temporal ambit. This substantial disparity in revenue underscores the profound disparity in market size and commercial traction between these two spirits categories.

The discernible discrepancy in market size between whiskey and rum can be posited as a multifaceted phenomenon with far-reaching implications. Primarily, the considerable gap in revenue underscores the evidently broader consumer base and more intricate market dynamics that characterize the whiskey domain in comparison to its rum counterpart. The expansive consumer base inherently facilitates a more robust ecosystem conducive to a diverse array of investment opportunities and market transactions within the whiskey sector. Conversely, the comparatively narrower market scope of rum may imply a more contained investment landscape, potentially characterized by fewer opportunities for market participation and a lesser degree of liquidity.

6.3 Assumptions and limitations

6.3.1 Hedonic analysis limitations

The primary and conspicuous limitation of the Hedonic analysis pertains to the utilized dataset. In the absence of an extant dataset at the time of thesis development, the author initiated the creation of a pilot dataset. However, the acquired data exhibited deficiencies in quality, with a substantial portion of observations featuring missing values, thereby rendering them unsuitable for robust investigative analysis. Consequently, only a meager 3.6% of the total number of bottles possessed complete variables, despite constituting a representative sample size. This disparity underscores a potential sampling bias, as the majority of bottles remained excluded from the analytical framework.

Furthermore, certain variables, such as "Open rate" and "Rating", manifested subjective characteristics, as they were generated by members of the Rum-X community. Consequently, a degree of subjectivity is inherent in these variables, raising questions regarding their reliability and objectivity. For in-

[1] The figures were obtained from the Statista webpage in April 2024.
Available at *www.statista.com*

stance, with respect to the "Open rate" variable, it remains uncertain whether all users of the Rum-X website uniformly report their bottle as "opened" upon actual consumption.

A supplementary issue arises from the plausible incompleteness of the variable set. Empirical evidence, provided by Paroissien & Visser (2020), suggests that wines awarded in tasting competitions often command higher prices subsequently. Similarly, the consideration of production costs emerges as a pertinent factor that could significantly influence pricing dynamics. However, owing to the unavailability of these variables, they were regrettably omitted from the analytical framework.

Finally, a notable challenge encountered was the presence of collinearity within the dataset. The categorical variables exhibited intercorrelations (see 6.4), complicating the isolation of individual effects. The author endeavored to mitigate this challenge by segregating each effect and employing averaging techniques across multiple models. Nonetheless, the intricate nature of the dataset rendered this process considerably convoluted and posed challenges in interpretation.

6.3.2 Repeat-sales analysis limitations

Much like the Hedonic analysis, the Repeat-sales analysis also hinged significantly upon the quality of the dataset. Despite being arguably of superior quality, the dataset was not devoid of limitations. The foremost constraint lies in the potential for sampling bias. Originating from the Rum Auctioneer, a specific auction house, the dataset is susceptible to sampling bias inherent to the clientele of this particular platform. Moreover, as highlighted in the study by Breeden & Liang (2017), auction dynamics exert a notable influence on the pricing of wine bottles. Consequently, similarly to wine auctions, certain rum prices may not align optimally due to the unique dynamics inherent to each auction setting.

A secondary limitation arises from the absence of a unique identifier for each bottle within the dataset. Consequently, the author was compelled to identify duplicates by scrutinizing all features provided by the auction house and postulating that bottles sharing identical features were indeed the same. However, this assumption is not universally valid. Upon closer examination of a subset of resales, it was observed that bottles with identical features exhibited disparate visual designs. Nonetheless, notwithstanding the aesthetic variations, the content within these bottles remained consistent, comprising the same rum possessing identical attributes.

Chapter 7

Conclusion

This study delved into two primary inquiries: Firstly, what factors determine the pricing of rum, and secondly, can rum be regarded as a viable alternative investment asset?

To address the former query, a comprehensive Hedonic analysis was conducted utilizing the Weighted Least Squares methodology on a pivotal dataset sourced via web scraping from the Rum-X website. The endeavor of identifying the price determinants was notably intricate due to the pervasive issue of multicollinearity among variables. To contend with this challenge, the researcher adopted a multifaceted approach, employing Stepwise regression, Bayesian Model Averaging, and Lasso regression techniques. This methodological diversity allowed for a nuanced exploration of the underlying dynamics governing rum prices.

The findings underscored the absence of a singular model capable of perfectly elucidating the dataset. Instead, the analysis yielded multiple models, each offering unique insights into the determinants of rum pricing. Notably, Age and Brand emerged as standout variables, exhibiting robust explanatory power across all models and thereby warranting consideration as primary drivers of rum prices.

While other variables demonstrated some degree of explanatory effect, their influence consistently paled in comparison to Age and Brand. Overall, the evidence suggests that rum prices are predominantly shaped by the reputational impact of the Brand. It became evident that much of the information pertaining to a rum bottle was encapsulated within its brand identity, as the incorporation of additional categorical variables in conjunction with Brand yielded marginal improvements in the goodness of fit measures.

Key observations from the analysis included the positive effect of aging on price, as well as the tendency for higher alcohol volume rums to command higher prices. The higher cost of aged rums can be attributed to factors such as the significant investment of time and storage space required for aging, the development of depth and complexity of flavor achieved through the aging process, and the inevitable evaporation loss, often referred to as the "angel's share."

Additionally, the analysis revealed a negative relationship between Open rate and price, indicating that bottles which are purchased but not opened exert upward pressure on pricing. Surprisingly, neither Rating nor Number of bottles exhibited a discernible impact on pricing dynamics.

Drawing parallels with the seminal work of Oczkowski (2001) and Luigi Benfratello & Sacchetto (2009) on wine pricing, the researcher concludes, based on the statistical importance of Brand variable, that the reputation attributes play pivotal roles in shaping rum prices. On the other hand, in contrast to the findings of Cordiez (2020) regarding collectible whiskey prices, the scarcity factor appears to exert a less pronounced influence on rum pricing dynamics.

Regarding the examination of rum as an alternative investment asset, a Repeat-sales methodology was implemented on a dataset obtained by web scraping from the Rum Auctioneer online auction platform. From this meticulous analysis, it becomes evident that auctioned rums, in general, do not exhibit consistent price increases over time. The findings reveal that the average Repeat-sales index, signifying the overall price fluctuation within a given period, hovers near zero, indicating a lack of consistent price growth. However, the index's distribution displays positive skewness, suggesting the potential for significant price escalation in selected periods.

Despite sharing characteristics with whiskey, rum does not consistently demonstrate the appreciating market value delineated by Cordiez (2020) as a hallmark feature of collectible assets. Consequently, the author deduces that rum may not be optimally suited as an alternative investment asset at the present juncture.

The author suggests that these findings point to the implications of a significantly smaller market size and, consequently, a reduced consumer base. Indeed, the rum market is estimated to be only about one-fifth the size of the whiskey market[1]. This substantial difference in scale significantly curtails the demand for rum compared to whiskey, which could explain the absence of observed significant returns in the rum market.

It is imperative to acknowledge the limitations inherent in these conclusions, as they may significantly impact their interpretability. Most importantly, both datasets utilized in the analyses were susceptible to sample bias.

Particularly, the Hedonic analysis dataset suffered from low data quality, with a majority of observations featuring missing values. Secondly, the inclusion of subjective variables like "Open rate" and "Rating," reliant on user-generated inputs, introduced a layer of uncertainty which could further influence the results. Moreover, some theoretically crucial factors, such as tasting competition awards and production costs, were regrettably excluded due to unavailability. Lastly, collinearity among categorical variables added another layer of complexity, demanding sophisticated modeling techniques for interpretation.

The Repeat-sales analysis, despite its reliance on a comprehensive dataset, confronted notable constraints. Foremost among these is the apprehension regarding sampling bias, stemming from the dataset's exclusive derivation from a singular auction house, thus implying the possibility of skewed findings. Moreover, the intricate interplay of auction dynamics with rum pricing dynamics, as delineated in prior scholarly investigations, introduces additional layers of complexity. Additionally, the absence of distinctive bottle identifiers necessitates the adoption of assumptions in the identification of duplicates, a process fraught with uncertainty. This uncertainty is exemplified by instances where discrepancies in aesthetic attributes obscure consistent rum contents, underscoring the methodological intricacies inherent in the analysis.

Hence, for future research endeavors, the author advocates for the acquisition of high-quality datasets directly from reputable sources to mitigate the aforementioned limitations and enhance the robustness of subsequent analyses.

[1] The figures were obtained from the Statista webpage in April 2024.
Available at *www.statista.com*

References

ANGULO, A. M., J. M. GIL, A. GRACIA, & M. SÁNCHEZ (2000): "Hedonic prices for spanish red quality wine." *British Food Journal* **102(7)**: pp. 481–493.

BREEDEN, J. L. & S. LIANG (2017): "Auction-price dynamics for fine wines from age-period-cohort models." *Journal of Wine Economics* **12(2)**: p. 173â€“202.

BURTON, B. J. & J. P. JACOBSEN (1999): "Measuring returns on investments in collectibles." *The Journal of Economic Perspectives* **13(4)**: pp. 193–212.

CHANEL, O., L.-A. GÉRARD-VARET, & V. GINSBURGH (1992): *The relevance of hedonic price indices: the case of paintings.* Groupe de Recherche en Economie Quantitative et Econométrie, UA CNRS.

CORDIEZ, T. (2020): *That's the Spirit: An Analysis of Investment in Collectible Whisky.* Master's thesis.

DANET, B. & T. KATRIEL (1989): "No two alike: Play and aesthetics in collecting." *Play & Culture* **2**: pp. 253–277.

GOLAN, A. & H. SHALIT (1993): "Wine quality differentials in hedonic grape pricing." *Journal of Agricultural Economics* **44(2)**: pp. 311–321.

LANDON, S. & C. E. SMITH (1997): "The use of quality and reputation indicators by consumers: the case of bordeaux wine." *Journal of consumer Policy* **20**: pp. 289–323.

LANDON, S. & C. E. SMITH (1998): "Quality expectations, reputation, and price." *Southern Economic Journal* **64(3)**: pp. 628–647.

LUIGI BENFRATELLO, M. P. & S. SACCHETTO (2009): "Taste or reputation: what drives market prices in the wine industry? estimation of a hedonic model for italian premium wines." *Applied Economics* **41(17)**: pp. 2197–2209.

Martin J. Bailey, R. F. M. & H. O. Nourse (1963): "A regression method for real estate price index construction." *Journal of the American Statistical Association* **58(304)**: pp. 933–942.

Oczkowski, E. (1994): "A hedonic price function for australian premium table wine." *Australian Journal of Agricultural Economics* **38(1)**: pp. 93–110.

Oczkowski, E. (2001): "Hedonic wine price functions and measurement error." *Economic Record* **77(239)**: pp. 374–382.

Paroissien, E. & M. Visser (2020): "The causal impact of medals on wine producers' prices and the gains from participating in contests." *American Journal of Agricultural Economics* **102(4)**: pp. 1135–1153.

Pearman, W. A., J. Schnable, & A. K. Tomeh (1983): "Rationalization and antique collecting." *Free Inquiry in Creative Sociology* **11(1)**: pp. 55–58.

Pierre Combris, S. L. & M. Visser (2000): "Estimation of a hedonic price equation for burgundy wine." *Applied Economics* **32(8)**: pp. 961–967.

Schamel, H. (1998): "Phase space vortices in unbunched beams in synchrotrons." *Physica Scripta* **1998(T75)**: p. 23.

Steel, M. F. J. (2020): "Model averaging and its use in economics." *Journal of Economic Literature* **58(3)**: pp. pp. 644–719.

Taylor, L. O. (2003): *The Hedonic Method*, pp. 331–393. Dordrecht: Springer Netherlands.

Wade, C. (1999): "Reputation and its effect on the price of australian wine." *Australian and New Zealand Wine Industry Journal* **14(4)**: pp. 82–84.

Walgreen, D. (2010): "Investing in collectibles." *Erasmus University* .