Universitat Autònoma de Barcelona

Facultat d'Economia i Empresa UAB

Treball de Final de Grau

Facultat d'Economia i Empresa

TÍTOL:
DOES BANK GUIDANCE STATISTICALLY ENHANCE PORTFOLIO OPTIMALITY? AN EMPIRICAL STUDY.

AUTOR:
MARCEL GIBERT SÁNCHEZ

TUTOR:
RAMÓN PRAT CASANOVAS

GRAU:
BUSINESS MANAGEMENT AND ADMINISTRATION

DATA:
28/05/2024

Abstract: The following paper investigates empirically the associated relationship between bank guidance and portfolio optimization in Catalonia. In other words, whether this set of recommendations and instructions offered by financial institutions does statistically enhance the optimality of Catalan retail investors' financial portfolios. Despite having access to valid empirical investment theories, individuals exhibit a tendency to deviate from these strategies and thus, incurring the tenancy of suboptimal financial portfolios. Within this unideal context, it conveys that financial institutions are capable of providing meaningful value to their investment clients by assorting wealth management recommendations. Ergo, the purpose of this essay is to represent the distribution of optimal and suboptimal portfolios across individuals, likewise understand and quantify the critical variables directing toward inefficient portfolios, and fundamentally, probe the nexus between adhering to bank instructions and holding a well-balanced portfolio. Furthermore, this thesis intends to give investor's empirical insights to effectively manage and control their financial allocation.

Key words: portfolio optimization, empirical study, hypothesis testing, data-driven

Short abbreviations:

EMH – Efficient-Market Hypothesis

MPT – Modern Portfolio Theory

ETF – Exchange Traded Funds

SEC - Securities and Exchange Commission

CI – Confidence Interval

SAFS – Self-Awareness Financial Score

TABLE OF CONTENTS

TABLE OF FIGURES

1. INTRODUCTION

Navigating the sophisticated arena of investment entails not only wise and rational decision-making, but also consistent strategic guidance, making it truly challenging to operate efficiently the financial markets. Moreover, retail investors' decisions are usually influenced by cognitive dissonances, and the establishment of anchor points. Despite the existence of empirically grounded investment methodologies to maximize returns while adhering to risk, a significant proportion of subjects tend to construct suboptimal investment portfolios.

Within this complex landscape, where the anomalies and inefficiencies are expected, banks appear to be a trustworthy alternative to play the role of offering valuable insights to their customers. These entities are perceived as reliable partners with deep understanding and know-how in the matter. In this advantageous environment, they serve as a reliable source of information for a vast majority of investors, who rely on the skilfulness and competence of bank employees. Essentially, contributing and assisting stakeholders in the pursuit of economic well-being and financial success.

Given this framework, this thesis endeavors to statistically confirm the perceived effectiveness of financial entities in optimizing their customers' portfolios in Catalonia. All efforts thereby, aim towards quantifying this impact so that retail investors can possess empirical evidence to support their financial decisions. This provision of data-centric information is a valuable resource and a pivotal element of the thesis; empowering subjects to engage in wiser-informed financial decisions.

2. OBJECTIVES AND RESEARCH QUESTIONS

At the forefront of the thesis' aspiration and as a starting point for uncovering patterns across the data set, rests the principle of descriptive statistics. First, it is imperative to provide sufficient context regarding sociodemographic variables such as sex, age, or job position. Besides, the objective is to provide frequency and distribution tables to enumerate counts and percentages regarding the following aspects: percentage of respondents actually following their financial entities, self-awareness financial score (SAFS) concerning biases and transactions, and ultimately delving into the descriptive statistics concerning risk tolerance, investment horizon, and diversification allocation among sample participants.

Once the sample is contextualized, another objective is to cross-examine relevant variables to expose correlations and patterns regarding the subsequent aspects: determine if gender, age, or job position statistically influences portfolio optimization, if there is a significant relationship between age and risk exposure, and quantify the interrelation between self-awareness financial score (SAFS) and portfolio optimization.

Beyond the earlier overview, the main standing point for the thesis is to determine if there is a statistically significant relationship between following banks' guidance and holding an optimal portfolio. In this sense, a distribution table is constructed to depict visually these percentages. Furthermore, the thesis aims not only to not merely tabulate this table but also to construct a multiple logistic regression model to determine which variables influence the most the likelihood of possessing a suboptimal portfolio.

Finally, it is crucial to provide readers and investors with data-driven insights so that they can effectively improve the general performance of their portfolios. At the end, empirical evidence is the foundation and tool for thoughtful financial decision-making.

3. METHODOLOGY

Towards understanding the efficient construction of portfolios, a solid foundation with a literature review and examination is primordial, as these acknowledgments are essential to determine if a given portfolio is efficient or not, based on theory-proven strategies.

The questionnaire design must be extensive enough to collect all the data necessary to address the specific research questions while remaining agile to focus solely on functional data. Additionally, the investigation involves processes related to the cleaning and verification of the dataset as variables may be structured, grouped or codified, before analysis initiation.

Descriptive analysis, the exploration of data through basic statistical measures, is a mandatory process in the early-stage analysis to provide sufficient context and insights regarding variables, employing the following techniques: distribution tables, means, quartiles, confidence intervals (CI), or percentages. Then, variables are crossed to reveal patterns and identify significant relationships.

To address the main topic of whether bank guidance enhances portfolio optimization, a contingency table and a logit regression model will be constructed to test the hypothesis, catering the main drivers and catalysts influencing the subjects' portfolio optimality.

Despite the use of advanced statistical techniques, the methodology and objective of this thesis remain clear: be sophisticated and technical, as well as visually represent insights in a comprehensive manner so readers and investors can easily follow. This cornerstone is key for the research; while statistics enable us to correctly project results to the larger population, data visualization is also vital to grasp and understand outputs. At the end, data is not just about numbers, it is about exposing the story behind them.

4. LITERATURE REVIEW

4.1 FINANCIAL ASSETS BACKROUND

Given this context, where securities are those items that compose a financial portfolio, the thesis aims to define and list the most commonly held securities among retail investors. For that purpose, note that the term “asset” possesses different implications depending in the context in which it is used, as it refers to any resource, whether tangible or intangible, that possesses intrinsic value and economic benefits for its owner.

Despite its broad meaning, the study aims to focus on financial assets since these are those securities that most commonly compose retail investors’ portfolios. These financial securities contain a unique trait of a dual nature: while they represent an asset for the investor, they are a liability for their issuer at the same time. To illustrate this duality, consider an individual who holds the right to receive periodic interest payments from bonds, while an institution or entity is obliged to make these interest payments happen.

Globally, there exists a wide, diverse and contrasting selection of investment vehicles, each characterized by its own risk and return possibilities. In fact, the exact number is hopeless to determine as new and modified investment products are continuously introduced. That is the main reason why this paper only focuses on the main investment assets that the average stakeholder capitalizes on, which usually are: checking accounts, pension plans, stocks, mutual funds, exchange-traded-funds (ETFs), bonds, real estate investments, commodities, cryptocurrencies, or any other instruments held in pursuit of economic benefits.

4.2 THE BASICS OF PORTOLIO AND OPTMIMALITY

In the following paragraphs, the definitions and implications of portfolio and optimality are investigated since both elements are essential to understand investment practices. As previously commented, they constitute crucial pillars on which the presented hypothesis is analysed and therefore, having a clear understanding helps to provide the context in which risk-return, time horizon, and diversification principles are deployed.

4.2.1 Portfolio definition and formulation

Informed by the study of (Gunjan & Bhattacharyya, 2023), "*portfolio in literal terms means a collection of financial investments which tend to give some form of return in future*" and mathematically represented by the same authors as follows:

$$P(t) = \sum_{j=1}^{n} w_j(t) s_j \quad (1)$$

where s represents each security, w is the weight of each one of it, and t is the specific period of time

This formulation gives rise to understand that financial portfolios are merely a combined effect produced by the individual list of assets held in that composition. As weighting is a main component here, although a given financial product yields substantial returns, if it does only represent a small proportion of the portfolio, its impact may be minimal. To the extent that it may not be sufficient to outweigh the impact of a small negative return with substantial weighting. Hence, it is emphasized the importance of achieving a balanced solution, an equilibrated allocation that ensures investors their overall performance is not adversely affected by a single security. Ultimately, portfolio revolves around diversification, a point further elaborated in page number 15.

4.2.2 Optimality definition

The concept of optimality encompasses a broad range of interpretations and connotations. To fully comprehend it, it must be understood that optimality encompasses more than efforts to achieve the maximum return as possible. In fact, optimality is all about trade-offs, carefully determining the most efficient solution given certain constraints such as the associated level of risk.

4.2.3 Mathematical formulations

At the end, optimality is grounded in mathematical problem-solving. This involves considering several concepts related to covariance (Equation 2) and correlation (Equation 3). Then, the construction of portfolio return (Equation 4) and portfolio risk (Equation 5) expressions follows:

$$\sigma_{ij} = p_{ij} * \sigma_i * \sigma_j \tag{2}$$

where σ is the standard deviation and p_{ij} is the correlation between the assets (Equation 3)

$$p_{ij} = \frac{\sigma_{ij}}{\sigma_i * \sigma_i} \tag{3}$$

where $p_{ij} \in [-1, 1]$ indicating that results closer to -1 show that assets vary inversely and results closer to 1 show that they move in tandem

Given these expressions related to how assets tend to move relative each other, it is presented that:

$$Portfolio\ return = \sum_{i=1}^{n} w_i * \mu_i \tag{4}$$

where w_i is the individual weight of each asset in the portfolio and μ_i is the expected return of each asset

And that,

$$Portfolio\ risk = \sum_{i=1}^{n} {w_i}^2 * {\sigma_i}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i * w_j * \sigma_{ij} \tag{5}$$

where w_i represents the individual weight of each asset, σ_i is the standard deviation of returns of each asset, and σ_{ij} is the covariance between assets.

In conclusion, by measuring (covariance) and standardizing (correlation) the movement of assets relative to each other, the objective is to maximize the weighted sum and expected return of individual assets (portfolio return), subject to minimizing its variance of returns and considering the implicit covariance between assets (portfolio risk). To assess whether this optimality holds in practice, a Cambridge study by (Kopa & Post, 2009) demonstrates that using their mathematically formulated optimality measure, "*the U.S. stock market portfolio is significantly nonoptimal relative to benchmark portfolios formed on market capitalization and book-to-market equity ratios*" implying that in practice stocks selected by investors U.S. are not efficiently allocated. Let us investigate if this insight holds in Catalonia.

4.3 WHAT IS AN OPTIMAL PORTFOLIO THEN?

Beyond these mathematical expressions, optimality entails understanding the factors contributing to portfolio success and efficiently selecting those investment products that best align with your financial goals. Hence, a clear and attainable objective is fundamental for making wise decisions. At the end, optimality is about strategy. Each option carries its own level of risk, potential return, maturities, and countless other factors that may impact outcomes. Thus, this thesis focuses on the subsequent elements:

4.3.1 Risk-return relationship

Contrary to the presumed belief that risk and return move in tandem, an analysis of the U.S. stock market exchange (*The Empirical Risk–Return Relation: A Factor Analysis Approach - ScienceDirect*, n.d.) concludes that this relationship is not empirically confirmed. These findings underscore the importance and complexity of evaluating not only returns, which are random variables, but also its potential risk and associated likelihood. In certain circumstances, individuals may erroneously perceive slow growth assets as highly secure, steady, and consequently predictable while other investors may dismiss growth potential opportunities due to its risk.

Investors should adopt an equilibrated approach to align potential upsides and drawbacks of assets with their particular investor profile and financial aspirations. In line with this idea, it would be illogical to heavily capitalize in assets such as

cryptocurrencies if your portfolio is set to be conservative and stable over time, and vice versa. This approach not only increases the likelihood of achieving ultimate objectives but also builds resilience against unforeseen market changes.

Particularly, the total risk arising from an asset arises from multiple variables, including factors specific to the particularities of the asset itself, and others from external influences. Within this framework, the concepts related to systematic and specific risk are investigated.

Systematic risk (Equation 6) pertains to factors influencing arbitrarily an entire segment of the financial market. Some examples may include interest rate policies, inflation, economic cycles, and macroeconomic or political elements. Thus, by having influence across diverse investments and as it is presented in figure (1), it cannot be mitigated through diversification.

$$Systematic\ risk = \left(\beta_p * \sigma_m\right)^2 \tag{6}$$

where β_p is the portfolio sensitivity to market returns and σ_m is the standard deviation of market returns

On the other hand, specific risk (Equation 7) relates to the uncertainty directly linked to the stock-specific events, management decisions, and other factors merely related to the given asset. Therefore, this type of risk can be effectively mitigated through diversification and its correspondent formula would be the following:

$$Specific\ risk = (\sigma_e)^2 \tag{7}$$

where σ_e is the variance of specific returns of the asset

Ultimately, both concepts construct the expression to represent the total risk of a portfolio (Equations 8&9), denoted as:

$$Total\ risk = Systematic\ risk + \ Specific\ risk \tag{8}$$

$$\left(\sigma_p\right)^2 = \left(\beta_p * \sigma_m\right)^2 + (\sigma_e)^2 \tag{9}$$

please note that standard deviations are not additive

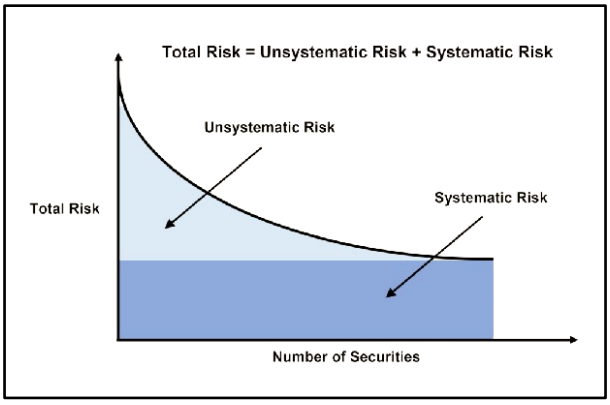


Figure 1: Total risk representation. (Stock Market Risk — Passive Investing Australia, 2020)

4.3.2 Investment horizon plan

Secondly, the intended investment time frame plays a critical role, as it should determine an individual's risk-tolerance. To illustrate this point, consider a young, healthy, and long-term oriented student. With a considerable future span ahead, it is predictable for such an individual to engage in riskier investments. Even with fluctuations, the holder has the ability to recover from potential or major drawbacks. For the same reason, short-oriented individuals are advised to prioritize capital preservation and security.

Taking the same example of the young individual, this investor may have the ability and tolerance to face instability and volatility in the short-term since the time value of money and compound interest (Equation 10) act in his benefit.

$$FV = PV * (1 + r)^n \quad (10)$$

where FV is the future value of the portfolio, PV is the present value of the initial investment, r is the weighted average return of individual assets, and n is the number of periods

Now, let us consider an identical scenario with the same initial investment, contributions, and rate of return, but differing only in the time horizon. By making use of the compound interest calculator[1] provided by the Securities and Exchange Commission (SEC), it is observed the power of compounding with the comparison showcased in Figure (2), representing the same investment held for 50 years versus 25 years.

[1] https://www.investor.gov/financial-tools-calculators/calculators/compound-interest-calculator

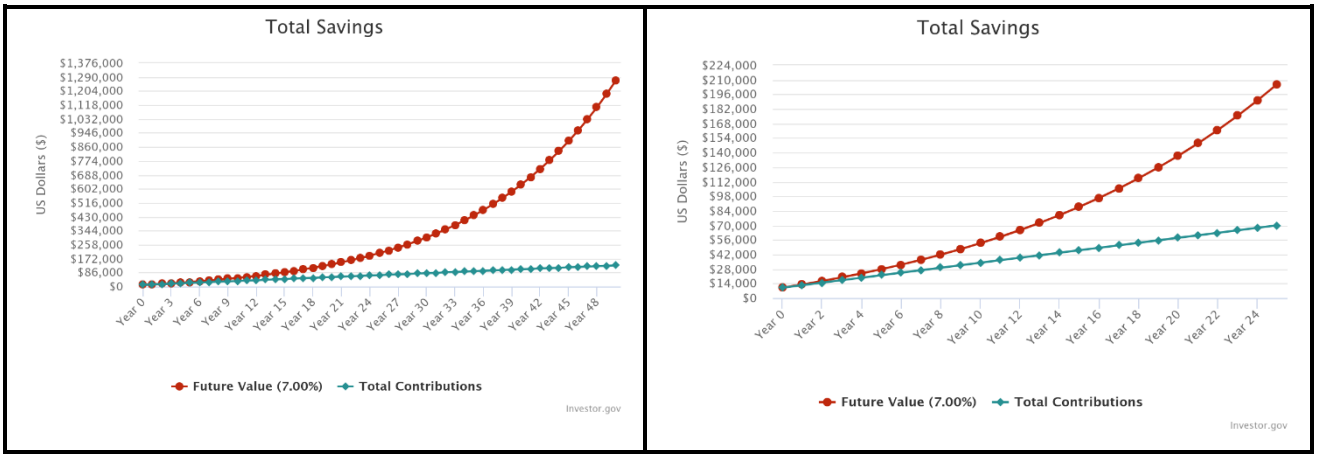


Figure 2: The power of compounding comparison. (Self-elaborated, 2024)

4.3.3 Diversification strategy

Thirdly, in conformity with mathematical formulations, the correct allocation of assets composing the portfolio is key. As observed regarding specific risk, by means of spreading investments individuals actually reduce part of the risk associated with the portfolio. To implement this point efficiently, diversification not only implies having a wide variety of assets but also ensuring that these assets are not positively correlated.

Indeed, true diversification relies on constructing a financial portfolio that encompasses a wide range of asset types, each with its inherent yields and uncertainties; acquiring financial products from different regions and countries across the map, mitigating the risk of geopolitical factors; maintaining assets with different levels of liquidity, managing efficiently cash-flow demands; investing in different sectors and industries, spreading the risk of market trends and economic cycles; capitalize on distinct investment styles, such as growth or value investing; but foremost it requires avoiding significant correlation between individual assets.

This last point is illustrated by understanding the following numerical example: Suppose three different portfolios all containing two different individual assets, with a correlation $p_{1,2} \in [1, 0, 1]$. Each individual asset is equally allocated: $w_1 = w_2 = 0{,}5$. The expected return and volatility for asset one are 0,1 and 0,2 respectively. And for asset number two these are 0,07 and 0,1.

Making use of the (Equations 2&5), the portfolio risk is computed for the three cases, and the following results are obtained:

$$Portfolio\ risk = \sum_{i=1}^{n} {w_i}^2 * {\sigma_i}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i * w_j * \sigma_{ij}$$

Correlation equal to positive one:

$$\sigma_{ij} = p_{ij} * \sigma_i * \sigma_j = 1 * 0{,}07 * 0{,}1 = 0{,}007$$

$${\sigma_p}^2 = 0{,}5^2 * 0{,}07^2 + 0{,}5^2 * 0{,}1^2 + 2 * 0{,}5 * 0{,}5 * 0{,}007 = 0{,}007225$$

$$\sigma_p = 0{,}085$$

Correlation equal to zero:

$$\sigma_{ij} = p_{ij} * \sigma_i * \sigma_j = 0 * 0{,}07 * 0{,}1 = 0$$

$${\sigma_p}^2 = 0{,}5^2 * 0{,}07^2 + 0{,}5^2 * 0{,}1^2 + 2 * 0{,}5 * 0{,}5 * 0 = 0{,}003725$$

$$\sigma_p = 0{,}061$$

Correlation equal to negative one:

$$\sigma_{ij} = p_{ij} * \sigma_i * \sigma_j = -1 * 0{,}07 * 0{,}1 = -0{,}007$$

$${\sigma_p}^2 = 0{,}5^2 * 0{,}07^2 + 0{,}5^2 * 0{,}1^2 + 2 * 0{,}5 * 0{,}5 * (-0{,}007) = 0{,}000225$$

$$\sigma_p = 0{,}015$$

The conclusion is empirical, please denote that as the correlation between individual assets decrease, the associated portfolio risk tends to be decreased as well. This confirms that diversification, considering not only all the factors previously mentioned but also correlation, mitigates a significant portion of the risk inherent in a given portfolio.

4.4 BANKS ADVICE AND COLLUSION OF INTERESTS

Following the findings of (*The Performance of Individual Investors in Structured Financial Products | Review of Quantitative Finance and Accounting*, n.d.); it is argued that investors lack fundamental skills, expertise, and time regarding investment practices. It is also concluded that individuals perceive the selection and management of financial portfolios as a complicated and time-consuming process. Given this laborious context, it might be reasonable to assume that banks are credible and savvy corporations that can enhance investment management and financial success. However conversely, the same study empirically concludes that "*banks are apparently able to exploit the investors' inability to calculate fair prices for complex payoff profiles by selling products at high premiums*".

Parallelly, it invites the exploration into the suggestion that perhaps the advice offered by financial institutions is negatively influenced by conflicting interests and biases. Indeed, banks are unambiguously profit-seeking entities, much like other companies, evoking that they are inherently attained to results and profits. Gains that may diverge from those of individuals. That is the main reason why (*Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study | The Review of Financial Studies | Oxford Academic*, n.d.) conducted a large and empirical study to determine if unbiased financial advice would actually steer investors to efficient portfolios or not. Their conclusions reveal that the average client who adheres to bank recommendations does not experiment increase in portfolio efficiency, implying that even theoretically sound advice is not a sufficient element for benefiting retail investors.

Likewise, this paper aims not only to quantify the percentage of subjects who follow bank recommendations but also determine empirically the relationship with portfolio optimality. A relationship, which in parallel to the previous study should be negative, especially considering that their work was conducted under unbiased guidelines. Note that real-life scenarios, the ones intended to represent in this analysis, tend to deviate from this ideal circumstance; apparently intensifying the coincidence with their work.

5. QUESTIONNAIRE'S DESIGN

5.1 THEORETICAL SAMPLE SIZE & REPRESENTATION

The ultimate objective of statistical research is to develop robust inferences about a population using a given sample set. Consequently, it entails obtaining valid outputs and findings that accurately represent this population. Nevertheless, collecting data from the entire population poses a of limitations and constraints related to time, data quality, biases, and budget. That is the main reason why researchers rely on sampling.

5.1.1 Sampling importance

Sampling is nothing else than the statistical technique applied when the researcher selects a subset of the population to represent it. Thus, not only the size but also the methodology must be adequate to draw reliable and valid conclusions with a given degree of confidence. This premise is underscored by (Singh & Masuku, 2014) stating that "*the selection of sampling methods and determination of sample size are extremely important in applied statistics research problems to draw correct conclusions. If the sample size is too small, it may estimate those impacts or associations too imprecisely. If the sample size is too large, the study would be more complex and may even lead to inaccuracy in results".*

5.1.2 Sampling method and limitations

When gathering information about a population, various techniques and strategies are employed depending on the methodology, objectives, and constraints of the study. Given the specific academic nature of this final degree thesis, constraints related to the research itself are apparent. It is evident that the recruitment and data collection processes are negatively influenced by the budget, time, and accessibility limitations for a bachelor's student.

Following the study of (Singh & Masuku, 2014), the most suitable sampling method for this study is Purposive Sampling. This technique involves selecting individuals according to a specific purpose or criteria. In this particular study, participants are selected based on availability and connectivity with the student and then, they voluntary

respond via web to the set of questions constructed in the survey encountered in page 42.

Nevertheless, it is imperative to highlight that while "*purposive sampling might provide biased estimate and it is not statistically recognized; random sampling, in which each individual has a pre-assigned chance of inclusion, provides better estimate of parameters".* At the end, as the individuals selected may share a close connection with the student, it is important to acknowledge the potential bias in the analysis and findings, implying that they may not be representative of the entire population.

5.1.3 Sampling size and limitations

By definition, according to (Noordzij et al., 2011), "*sample size is the number of patients that need to be included in a study to answer the research question*". In other words, it is the process of ensuring statistical validity when choosing a subset of the population to represent the entire population itself. For this thesis, the entire group of individuals to draw conclusions about are those adult inhabitants from Catalonia aged 18 years old and above. According to Idescat (Institut d'Estadística de Catalunya)[2] the total and updated number for this population segment in the first semester of 2023 is 7.901.963 Catalans.

With the intention of calculating this subset numerically, it is essential to adhere to valid statistical criteria with regard to sample size calculation, rather than simply computing it on a webpage. Building upon the research by (*Determining Sample Size by Glen Israel.Pdf*, n.d.), (Equation 11) for large population cases is developed:

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 * p * q}{e^2} \tag{11}$$

Let us briefly break down each of the items composing the formula in the following paragraphs:

[2] https://www.idescat.cat/pub/?id=ep&n=9123

Significance level (α):

The significance level or alpha value, often denoted as Type I Error, as outlined in the research of (*Hypothesis Testing, Type I and Type II Errors : Industrial Psychiatry Journal*, n.d.), signifies the maximum probability for a false-positive that the researcher is willing to accept. Typically, alpha is set at 0,05 implying that this thesis does not contemplate more than a 5% chance of incorrectly concluding that following banks' guidance has a statistically significant relationship with withholding an optimal portfolio, when in fact, no such relationship exists. Note that this value is pivotal in hypothesis testing as it determines the region for rejecting the null hypothesis.

Z-score (z):

The z-score represents the boundaries beyond which the null hypothesis is rejected based on the standard normal distribution. Given the significance level of 0,05, the critical value of 1,96 is extracted using the table in Figure (3).

	hundredths									
tenths	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	1.00000	0.99202	0.98404	0.97607	0.96809	0.96012	0.95216	0.94419	0.93624	0.92829
0.1	0.92034	0.91241	0.90448	0.89657	0.88866	0.88076	0.87288	0.86501	0.85715	0.84931
0.2	0.84148	0.83367	0.82587	0.81809	0.81033	0.80259	0.79486	0.78716	0.77948	0.77182
0.3	0.76418	0.75656	0.74897	0.74140	0.73386	0.72634	0.71885	0.71138	0.70395	0.69654
0.4	0.68916	0.68181	0.67449	0.66720	0.65994	0.65271	0.64552	0.63836	0.63123	0.62413
0.5	0.61708	0.61005	0.60306	0.59611	0.58920	0.58232	0.57548	0.56868	0.56191	0.55519
0.6	0.54851	0.54186	0.53526	0.52869	0.52217	0.51569	0.50925	0.50286	0.49650	0.49019
0.7	0.48393	0.47770	0.47152	0.46539	0.45930	0.45325	0.44725	0.44130	0.43539	0.42953
0.8	0.42371	0.41794	0.41222	0.40654	0.40091	0.39533	0.38979	0.38430	0.37886	0.37347
0.9	0.36812	0.36282	0.35757	0.35237	0.34722	0.34211	0.33706	0.33205	0.32709	0.32217
1.0	0.31731	0.31250	0.30773	0.30301	0.29834	0.29372	0.28914	0.28462	0.28014	0.27571
1.1	0.27133	0.26700	0.26271	0.25848	0.25429	0.25014	0.24605	0.24200	0.23800	0.23405
1.2	0.23014	0.22628	0.22246	0.21870	0.21498	0.21130	0.20767	0.20408	0.20055	0.19705
1.3	0.19360	0.19020	0.18684	0.18352	0.18025	0.17702	0.17383	0.17069	0.16759	0.16453
1.4	0.16151	0.15854	0.15561	0.15272	0.14987	0.14706	0.14429	0.14156	0.13887	0.13622
1.5	0.13361	0.13104	0.12851	0.12602	0.12356	0.12114	0.11876	0.11642	0.11411	0.11183
1.6	0.10960	0.10740	0.10523	0.10310	0.10101	0.09894	0.09691	0.09492	0.09296	0.09103
1.7	0.08913	0.08727	0.08543	0.08363	0.08186	0.08012	0.07841	0.07673	0.07508	0.07345
1.8	0.07186	0.07030	0.06876	0.06725	0.06577	0.06431	0.06289	0.06148	0.06011	0.05876
1.9	0.05743	0.05613	0.05486	0.05361	0.05238	0.05118	0.05000	0.04884	0.04770	0.04659
2.0	0.04550	0.04443	0.04338	0.04236	0.04135	0.04036	0.03940	0.03845	0.03753	0.03662

*Figure 3: Z-Score table (*Z-Two-Tails.Pdf, *n.d.)*

Estimated proportion (p,q**):**

The determination of components p and q is influenced by prior knowledge regarding the variables under study. In this thesis, to maintain neutral scenarios and avoid assumptions, both p and q are specified to have a value of 0,5. This selection reflects the assumption that the proportion of investors following bank's guidance or holding an optimal portfolio is equally likely, regardless of any prior information.

Margin of error (e):

Finally, the margin of error represents the degree of accuracy in an estimate, indicating to what extent the sample estimate may deviate from the population parameter. Reducing this value diminished the discrepancy but it also amplifies the sample size requirements.

Having provided adequate background information for the calculation of sample size, it proceeds to input the values into the formula and ascertain the optimal number of respondents required (n).

- Optimal value e = 5%

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 * p * q}{e^2} = \frac{(1{,}96)^2 * 0{,}5 * 0{,}5}{0{,}05^2} = 384{,}16$$

$$n = 385$$

The optimal sample size required to draw meaningful conclusions about the relationship between bank's guidance and optimality is determined to be 385 individuals. Nonetheless, considering the nature and scope of this thesis, as well as the apparent budget and time constraints inherited; achieving this number is exceptionally difficult. Thus, to address these practical limitations and available resources, the formula is recalculated by increasing the margin of error while still ensuring the reliability and validity of the research.

- Optimal value e = 10%

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 * p * q}{e^2} = \frac{(1{,}96)^2 * 0{,}5 * 0{,}5}{0{,}1^2} = 96{,}04$$

$$n = 97$$

The total number of individuals needed to complete the survey is 97. Although achieving this goal may present challenges in terms of effort and time, the objective for the thesis is settled at 100 completed and valid surveys by the scheduled time. Upon the survey's completion, the total number of responses obtained amounts to 102.

5.2 QUESTIONNAIRE CONSTRUCTION

Constructing a well-designed questionnaire stands as a cornerstone in data collection since it serves as the basic tool for researchers to filter, modify, clean, and analyse data with the precision required for the Process stage outlined in the Annexes section in point 1.3. Each question incorporated within the survey is directly related to a variable of study, emphasizing the importance of agility and relevance in asking only those specific questions intended to be studied in depth. Please note that the final delivered questionnaire[3] is accessible through the following link.

With a total of 13 questions, the processed and final data table output[4] contains a total of 21 columns and 102 observations. The extensive process to obtain this final data table has been meticulously completed according to the instructions provided in pages 41-43. Observations and variables have been added, deleted, modified, grouped, and cleaned to meet the specific objective and research questions outlined for this specific thesis, ensuring that the dataset is tailored specifically to the requirements of the researcher.

6. CATEGORAZATION RELATED TO OPTIMALITY

In this section, the challenges addressed by directly asking respondents whether they have an optimal portfolio or not are posed. This approach may yield biases responses since individuals often struggle to assess their own investment decisions, leading to a propensity to claim their portfolio is optimal even when it may not be. As the research revolves around optimality, a key variable to be analyse, it has been conducted an analysis of various survey variables and responses.

Through careful examination of the 102 observations included in the dataset, it has been added a binary dummy variable to assess whether the portfolio of each particular individual is optimal or suboptimal. The coherence between the following variables has been employed for the categorization process: the respondent's age, their risk tolerance,

[3] https://forms.office.com/e/cw1cnHbKzp
[4] https://drive.google.com/file/d/13YcEdARAkMEcDQhUG1vMY92ofalOnH1/view?usp=embed_facebook

the planned investment horizon, their diversification level, their SAFS, and other key aspects related to the theoretical framework of this study.

The dataset revealed cases where discrepancies and incongruencies arose among responses; for instance, consider observation number 50 in the data set. A 19-year-old individual characterized with an aggressive investment profile, focusing on innovative and variable investments, but planning to withhold these investments for a period inferior to a year, thereby ignoring the inherent market fluctuations that the portfolio may face. Other observations such as number 58 or 95, include conservative profiles seeking security and minimum risk, but holding portfolios mainly composed of variable securities. Not only incongruences between risk tolerance and time have been evaluated, but the critical importance of diversification has also been addressed. Thus, observations lacking adequate spreading allocation were automatically categorized as suboptimal since the theoretical framework has mathematically acknowledged the importance of diversification in mitigating risk and enhancing returns.

7. STATISTICAL ANALYSIS

Before analysis initiation, please bear in mind that in the following paragraphs, every single output for all the explored analyses are encountered, by order of appearance, in the Annexes section of the thesis. Additionally, there are provided complementary analyses, hypothesis, tables, translations, and figures supporting each of the findings.

7.1 DESCRIPTIVE STATISTICS

The analysis starts by extracting general information regarding the dataset. Indeed, an initial comprehension of the dataset is fundamental to understand the sample of study and its overall insights. Therefore, in the subsequent paragraphs a statistical overview is employed to provide background regarding sociodemographic variables, to quantify the number of participants related to following bank suggestions, the risk tolerance level, the planned investment horizon, the diversification level, and other key factors related to withholding optimal portfolios. Ergo, the thesis presents histograms, counts, percentages, tables, averages, among other statistical tools to support this analysis.

7.1.1 Sociodemographic variables

Starting with gender, in the sample consists of 56 male respondents and 46 female respondents. From the 102 total participants the relative frequencies are 55% and 45% respectively. Hence, it is observed that the sample proportion of men and women is almost equal. In comparison with the population of study, the proportion according to (*Idescat. Estimacions de Població. Població. Per Sexe i Edat. Catalunya*, n.d.) of inhabitants from Catalonia aged 18 years old and above is 49% for men and 51% for women.

Subsequently, Figure (4) illustrates the age distribution of the sample individuals with a mean around 31 years old. As visually depicted, this proportion exhibits a clear tendency to include a much younger population compared with the real figures from Idescat. The statistical measures supporting this youth population demographic are the quartiles. Indeed, the third quartile indicates that the 75% of respondents are 34 years old or younger.

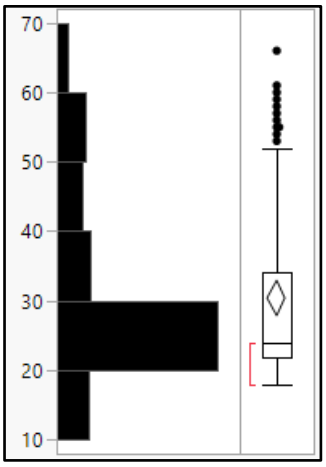


Figure 4: Age distribution (Self-elaborated, 2024)

Statistics	Figure
Mean	30,38
Standard Deviation	13,04
First Quartile	21,75
Median	24
Third Quartile	34
Outliers' exclusion	>52

Ergo, it is imperative to mention that biases previously discussed in Sampling method and limitations including that "*as the individuals selected may share a close connection with the student, it is important to acknowledge the potential bias in the analysis and findings; implying that it may not be representative of the entire population"* are encountered in this age distribution. Hence, these specific figures are not representative of the population under study.

To provide background information related to the job position, Figure (5) depicts the job sector mainly occupied by sample individuals across years.

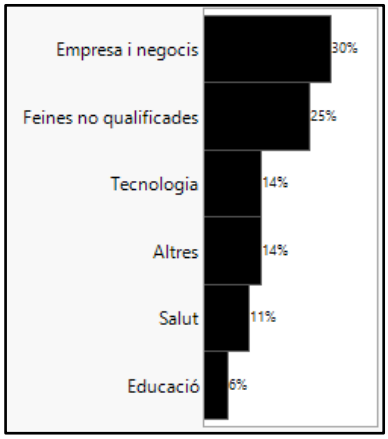


Figure 5: Job position distribution (Self-elaborated, 2024)

Almost one-third of the sample holds a job position associated to the Business sector, whereas exactly one-fourth are employed in unskilled positions. With 45% of the sample falling into these categories, it prompts the opportunity in page 31, to statistically prove a potential relationship between the investor's occupation background with the construction of an optimal portfolio.

7.1.2 Following bank suggestions

One main objectives of this thesis is to quantify the percentage of investors who prefer to be guided and rely on banks for their investment decisions. This insight is developed by using a dummy variable to assign those who do not follow bank suggestions with a value of zero, and a value of one assigned to those who do. The distribution of these categories is presented in Figure (6) showcasing the relative number for each group.

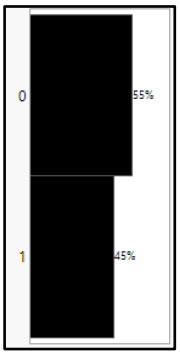


Figure 6: Following bank's suggestions distribution (Self-elaborated, 2024)

Although holders from the sample choose not to follow the recommendations provided by financial entities in a major part, the distribution is remarkably alike (55-45%). Hence, at first glance, there is not a significant tendency to follow or not follow bank suggestions. A conjecture statistically confirmed by analysing whether the observed proportion of investors relying on banks (55%) differs significantly from the hypothesized proportion of 50%. Employing Pearson's chi-square test, the null hypothesis, which states that the observed proportion equals the hypothesized proportion, is accepted. Therefore, as the calculated test statistic is greater than the p-value (0,32 > 0,05), it is statistically confirmed that the 55% of individuals not following bank suggestions do not significantly deviate from the hypothesized proportion of 50%.

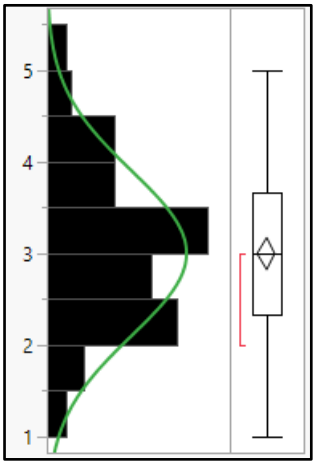
In the realm of inferential statistics, Confidence Intervals (CI) are generated to extrapolate insights from the sample data to the broader population. Due to implementing a larger margin of error (10%) wider CIs are expected, it is important to note that under the assumption of a 90% confidence level, all the Cis encountered in the thesis remain statistically reliable. In this sense, the true proportion of investors in Catalonia not following bank suggestions falls between 47-63%. Contrarily, the actual proportion of individuals who follow bank suggestions ranges from 37-53%.

All these findings estimate that although a larger proportion of investors do not follow bank recommendations, there is a balanced investor behaviour in Catalonia when it comes to follow financial entities' advice.

7.1.3 Self-Awareness Financial Score (SAFS)

A single construct has been developed, labelled as the "self-awareness financial score", to assess the collective financial acumen of investors. This construct concentrates three fundamental aspects for optimal portfolio, which are: the influence of emotions and cognitive biases on investment decision-making, the consideration of associated costs in portfolio construction (such as transaction costs, fees, tax implications), and the awareness of immediate liquidity needs and future income/expenses when making investment choices. These dimensions evaluated using a Likert Scale ranging from 1 to 5, briefly evaluate the awareness and expertise in managing portfolios effectively.

Hence, as the SAFS is calculated being the average of these three dimensions, interpret results close to 1 as favourable and results near 5 as unfavourable.



Statistics	Figure
Mean	3
Standard Deviation	0,9
First Quartile	2,3
Median	3
Third Quartile	3,7
CI	[2,8 – 3,1]

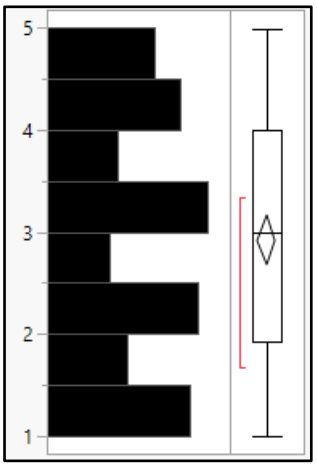
Figure 7: Self-awareness financial score distribution (Self-elaborated, 2024)

As visually depicted in Figure (7) and complemented with its correspondents statistics table, the output for sample individuals represent a clear central tendency to have a self-awareness financial score of exactly 3 out of 5, with a standard deviation of 0,9. It is worth mentioning how the histogram distribution of scores follow a similar pattern to that of a normal distribution represented by the green line. Likewise, the inference of the population is represented by the CI, expressing that Catalan investors have a punctuation between 2,8 and 3,1 points with a 90% confidence. Expanding further the analysis on the Annexes section, the component with a major negative effect on this score is the lack of consideration in immediate liquidity needs.

These insights underscore that Catalan investors exhibit a moderate level of financial acumen in managing portfolios effectively. However, findings also suggest room for enhancement to further improve their capabilities in investment decision-making, specifically on encouraging investors to consider their liquidity requirements and cash needs in the future.

7.1.4 Risk tolerance

By applying the same methodology, a unified construct has been formulated to assess the risk tolerance within their financial portfolio. In this case, the components encompass information regarding: a self-assessed risk profile (conservative-aggressive), the preferences in risk-return trade-offs (security-risk), and the allocation preference in financial assets (fixed or variable). Through this assessment, investor's risk profile and preferences are approached on a scale ranging from 1-5, indicating low risk - high risk respectively.



Statistics	Figure
Mean	2,9
Standard Deviation	1,2
First Quartile	1,9
Median	3
Third Quartile	4
CI	[2,7 – 3,1]

Figure 8: Risk tolerance distribution (Self-elaborated, 2024)

Figure (8) represents the distribution of risk across individuals for the sample and shows a central tendency with a mean of 2,9 and a standard deviation of 1,2. Correspondingly, with a 90% confidence the risk tolerance assumed by investors in Catalonia falls within the range of 2,7-3,1 in a scale of 5.

Therefore, the average investor in Catalonia tends towards a moderate level of risk, managing the risk-return trade-off in an equilibrated way. Please note in the following table that despite this central tendency, individuals are equally distributed across all levels of risk. This insight suggests that while the average stakeholder tends towards moderation, there exists a diversity of risk preferences ranging between all levels.

Risk tolerance	Count	Probability
[1-2)	25	24%
[2-3)	24	23%
[3-4)	26	26%
[4-5)	27	27%
TOTAL	**102**	**100%**

7.1.5 Investment horizon

Figure (9) visually represents the investment holding time that investors plan for their portfolio of assets. The distribution table indicates that 31% of sample investors plan to hold their assets for a time period larger than 10 years; suggesting a preference for futuristic and long-term investment strategies. In fact, the calculated CI reveals that between a 25-40% of Catalan investors plan to maintain their assets within this time period, with a 90% confidence. Under the same assumptions, between 20-35% of stakeholders have shorter-term aspirations with a range of 1-3 years. That is why, despite the predominance for long-term expectations, it exists a combination of preferences among investors, aligning with the variety of risk tolerance previously discussed.

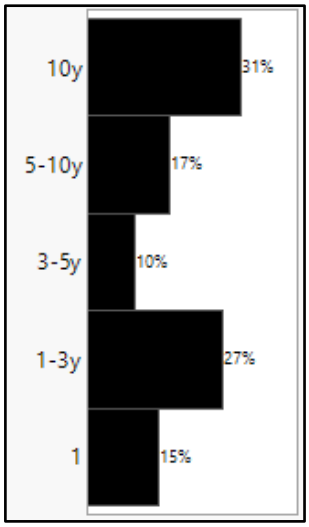


Figure 9: Holding time distribution (Self-elaborated, 2024)

7.1.6 Diversification allocation

A unified construct for evaluating the diversification among investors has also been formulated by employing the following key variables: the degree of correlation between assets (negative-positive), the breadth and diversity of asset classes (low-high), and the geographic investment extent across regions (low-high). These variables have been categorized employing a Likert Scale from 1 to 5 and thus, the diversification allocation construct ranges from a poorly concentration of investments to a well-spread allocation across classes and regions.

The histogram in Figure (10) portrays a mean of 2,8 with a standard deviation of 1. In a like manner, the true population parameter ranges from 2,6 - 2,9 with a 90% confidence level.

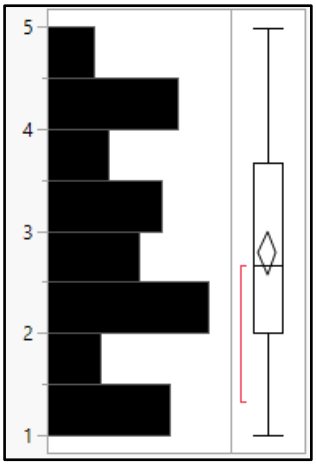


Figure 10: Diversification distribution (Self-elaborated, 2024)

Nevertheless, 23% of responses fall within the range of [1-2) and corresponding to a minimal diversification due to the combination of highly correlated assets, a low variety of classes, and all concentrated in a single region. In fact, the primarily factor contributing negatively to the diversification level is that 43% of surveyed investors indicate that their investment strategies are concentrated on a single country, emphasizing that individuals do not consider the vulnerabilities and exposure associated to country-specific downturns and risks. Hence, the findings conclude that the average investor in Catalonia exhibits a moderate level of diversification with a significant room for improvement with regard to spreading investments across various regions.

7.2 CROSSING KEY VARIABLES

To comprehensively understand investment behaviours of investors in Catalonia, the analysis continues to explore the interplay between some key variables. Specifically, the insights presented focus on examining relationships with sociodemographic factors, portfolio optimization, or risk exposure. Thereby, keep in mind that for every single relationship examined, a Null Hypothesis (HO) and an Alternative Hypothesis (H1) are tested. These hypotheses are explicitly formulated, per each relationship, in the Annexes section.

7.2.1 Sociodemographic factors on portfolio optimization

- Gender – Portfolio optimization (Contingency analysis)

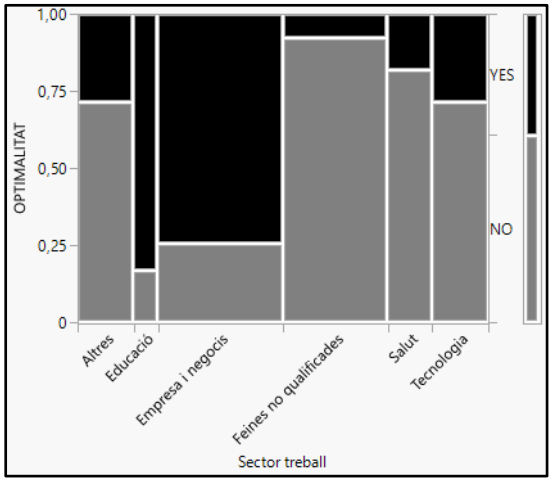
Since the Pearson chi-square test has a value greater than the significance level (0,42 > 0,05), the null hypothesis is not rejected. In practical terms, gender does not statistically influence on the likelihood of achieving an optimal portfolio or not.

- Age – Portfolio optimization (Logistic fit)

By taking a look at the output, the very high p-value (0,8) alongside that the chi-square statistic is very small (0,01), the findings conclude that there is not a statistically significant relationship between age and the success on portfolio optimization.

- Job position – Portfolio optimization (Contingency analysis)

Considering visually and statistically the output for Figure (11), the Pearson chi-square test being below the alpha value implies to reject the null hypothesis. In other words, the job sector majorly occupied by sample individuals do significantly influence on the likelihood of holding an optimal portfolio. Indeed, the following table reveals a clear disparity between individuals in business positions and those in unskilled positions, since the proportion of investors with optimal portfolios is much higher in the first case; and vice versa.



	Optimal	Subotimal	TOTAL
Business	23	8	31
Unskilled	2	24	26
Technology	4	10	14
Others	4	10	14
Health	2	9	11
Education	5	1	6
TOTAL	40	62	102

Figure 11: Optimality across job positions (Self-elaborated, 2024)

7.2.2 Correlation between age and risk tolerance

To examine the relationship between these both quantitative continuous variables, the statistical tools include a correlation analysis and a regression analysis. The Pearson's correlation coefficient, representing the strength and direction between variables, equal to -0,29 indicates a moderate negative relationship between age and risk tolerance. This tendency is confirmed by its low p-value (0,003 < 0,05) indicating that the relationship is statistically significant. Thus, as age increases, risk tolerance tends to decrease, and vice versa. Figure (12) visually represents this association being represented in red with the regression (Equation 12):

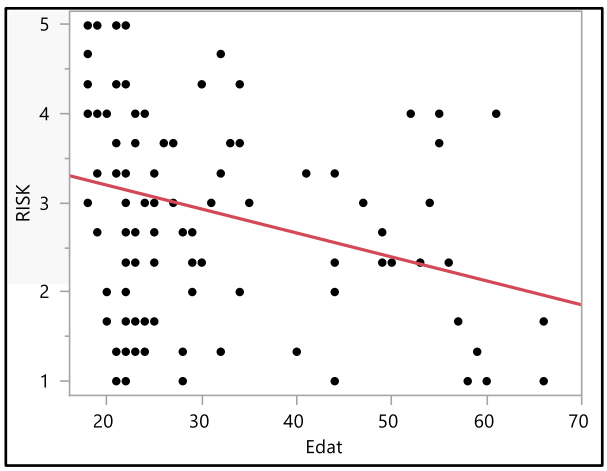


Figure 12: Linear fit for age and risk tolerance (Self-elaborated, 2024)

$$Risk\ tolerance = 3{,}74 - 0{,}027 * Age \qquad (12)$$

This mathematical presentation of how age predicts risk tolerance supports the inverse relationship between variables. Moreover, it is interpreted that over a period of 10 years, the risk tolerance is estimated to decrease by approximately 27%.

These findings indicate that age is a significant predictor of investor's risk tolerance, therefore as individuals grow older, they become less likely to take financial risks and disregard strategies focused on growth. Likewise, younger individuals tend to assume a higher risk.

7.2.3 Association between SAFS and portfolio optimization

Through a logistic regression model, it is examined how variations in the SAFS may impact the likelihood of maintaining either optimal or suboptimal portfolios across the sample set. Bear in mind that SAFS is reversed, implying that higher scores indicate a negative assessment on financial awareness.

The blue line in Figure (13) represents the prediction of the probability of portfolio optimality based on individual's SAFS hence as the score increases (negative assessment) it is predicted a decrease in the likelihood of maintaining an optimal portfolio.

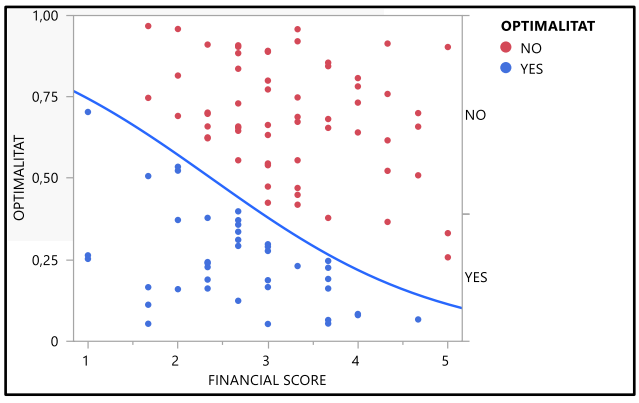


Figure 13: Logistic fit for SAFS and portfolio optimality (Self-elaborated, 2024)

This relationship is confirmed by the whole model test indicating that as the significance value is lower than the p-value (0,001 < 0,05), the null hypothesis stating that there is not a significant relationship between variables is rejected. Indeed, the analysis underscore that individuals with a lower financial acumen are less likely to construct and maintain optimal portfolios. Particularly, the odds ratio of 0,04 indicates

that for each one-unit increase in the financial score (for instance by moving from punctuations of 1 to 2), the odds of having an optimal portfolio decrease by approximately 96%.

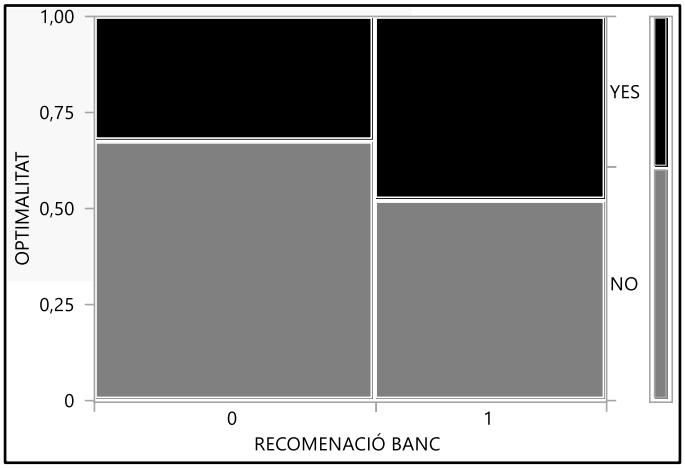
This insight emphasizes the relevance and need for education and initiatives aiming at enhancing individual's awareness on emotions and biases, on associated costs when investing, and on immediate liquidity needs.

7.3 DOES BANK GUIDANCE ENHANCE OPTIMALITY?

The following analyses aim to resolve the pivotal question of this thesis: whether there is a significant relationship between following bank guidance and having an optimal portfolio. To examine this relationship the following paragraphs include the contingency table with the absolute and relative counts of individuals within each category, its correspondent chi-square test of independence, and a multiple logistic regression model incorporating every single variable used in the research to identify the predictors most influencing portfolio optimality, offering empirical evidence on the impact of bank guidance and other factors on optimality.

7.3.1 Contingency tables

Figure (14) and its correspondent table; showcases that in the dataset approximately 61% of individuals maintain a suboptimal portfolio, whereas around 39% have an optimal portfolio so, there is a larger proportion in the sample of suboptimal portfolios than optimal ones. From these 62 individuals with a suboptimal portfolio, 38 do not follow bank guidance while 24 do. On the other hand, from those 40 individuals maintaining an optimal portfolio, the distribution between not following and following bank recommendations is equally similar, 18-22 individuals respectively.



Portfolio optimality	Followed bank guidance			
		NO	**YES**	**TOTAL**
	SUBOPTIMAL	38 37%	24 24%	62 61%
	OPTIMAL	18 18%	22 21%	40 39%
	TOTAL	56 55%	46 45%	**102** **100%**

Figure 14: Contingency table for bank guidance and portfolio optimality (Self-elaborated, 2024)

With these numbers, it is wise stating that there is a positive association between following bank guidance and portfolio optimality due to: from the suboptimal portfolios there are more individuals not following financial entities, while from the optimal ones there are more individuals actually following banks. However, although there is a positive association, this relationship is explored to test whether it is statistically significant or not by conducting a chi-square test. Indeed, the null hypothesis stating that there is not a significant relationship between both variables, is accepted since the Pearson's coefficient is greater than the significance value (0,1 > 0,05).

Hence, the relationship between following bank guidance and portfolio optimality is not statistically significant, implying that financial entities do not ensure a well-designed and optimal portfolio construction and suggesting that other alternatives, such as independent research, might be equally effective for investors.

7.3.2 Variables most influencing optimality

The analysis extends to evaluate the impact of each variable gathered during the survey, aiming to construct a multiple logistic regression model that identifies the main predictors for portfolio optimality. By considering the combined effect of all variables, a comprehensive and empirically grounded understanding of the factors significantly influencing the development of an optimal portfolio is attained.

Initially, the logistic regression model constructed in this thesis, incorporating all available variables, demonstrates statistical significance to evaluate the relationship with portfolio optimality. This statement is corroborated by the whole model test, which yields a p-value lower than the alpha value (0,0001 < 0,05). Furthermore, the coefficient of determination $(R^2 = 0{,}48)$ denotes a moderate level of explanatory power, suggesting that 48% of the variability in portfolio optimality (Y) is accounted for the model's predictor variables (X).

The following table represents the statistical significance of each variable included in the study for predicting the likelihood of maintaining an optimal portfolio. The model indicates that the job sector majorly occupied by individuals, the allocation strategy for diversification, and the assessment score for financial acumen emerge as statistically significant predictors since exhibit a p-value lower than the significance value.

Predictor	**p-value**
Job position	0,00045
Diversification construct	0,02420
Self-awareness financial score construct	0,03423
Source of information	0,23626
Following bank suggestions	0,33151
Gender	0,40824
Risk construct	0,67179
Investment horizon plan	0,86507
Age	0,99895

Therefore, the findings suggest that investors' professional background, diversification methodologies, and financial acumen assessment significantly influences on achieving wise and optimal investment strategies. Other factors such as gender, age, risk tolerance, investment time horizon, adherence to bank recommendations or the source of information alone are not individually substantial enough to impact.

8. FINAL CONCLUSIONS

The sociodemographic analysis indicates a balanced gender representation within the sample (55% men – 45% women), a skewed younger population (with 75% respondents aged 34 years or younger), and a notable portion of individuals employed in the business sector and unskilled positions (30-25% respectively). The distribution of investors following and not following bank recommendations is statistically confirmed to be balanced (55-45% respectively), signalling that investors display balanced behaviour in this regard. The self-awareness financial score of investors display a central tendency with a histogram distribution closely resembling a normal distribution. This suggests a moderate financial acumen among investors, although negatively influenced by a lack of consideration for immediate liquidity needs.

The average investor in Catalonia tends to assume a moderate level of risk (with a mean of 3 out of 5), has a predominance to hold these assets for a period longer than 10 years (31% of sample), and exhibits a moderate diversification level (with a mean of 2,8 out of 5). Despite this, 23% of sample respondents show a minimal diversification across all fields, indicating room for improvement in investment allocation.

Analyses reveal that gender and age do not significantly influence portfolio optimality, whereas job position does, with business professionals more likely to maintain optimal portfolios compared to those in unskilled positions. Furthermore, age and risk tolerance are negatively correlated implying that over a period of 10 years, risk tolerance is estimated to decrease by 27%. Similarly, for each one-unit increase in the SAFS (note the inverse scoring), the odds of having an optimal portfolio decrease by 96%.

Finally, the analysis underscore that although following bank suggestions shows a positive association with portfolio optimality, this relationship is not statistically significant. Hence, the multiple logistic regression model estimates that professional background, diversification level, and financial acumen are only significant predictors influencing the likelihood of achieving an optimal portfolio.

9. REFERENCES

Certificado profesional de Google Data Analytics | Coursera. (n.d.). Retrieved 11 March 2024, from https://www.coursera.org/professional-certificates/google-data-analytics

Compound Interest Calculator | Investor.gov. (n.d.). Retrieved 14 March 2024, from https://www.investor.gov/financial-tools-calculators/calculators/compound-interest-calculator

Determining sample size by Glen Israel.pdf. (n.d.). Retrieved 01 May 2024, from https://www.psycholosphere.com/Determining%20sample%20size%20by%20Glen%20Israel.pdf

DOES BANKS' GUIDANCE ENHANCE PORTFOLIO OPTIMIZATION? (n.d.). Retrieved 07 March 2024, from https://forms.office.com/pages/responsepage.aspx?id=KUxRa5EjMUi3dITzXEW_AQ72G4nv2VZDu7uhX1BtVN5UQVBUWjlVNTcwSFBLV1VVMDVRRlhNVU9MSi4u

Final Datatable.pdf. (n.d.). Google Docs. Retrieved 27 May 2024, from https://drive.google.com/file/d/13YcEdARAkMEcDQhUCG1vMY92ofaLOnH1/view?usp=embed_facebook

Gunjan, A., & Bhattacharyya, S. (2023). A brief review of portfolio optimization techniques. *Artificial Intelligence Review*, *56*(5), 3847–3886. https://doi.org/10.1007/s10462-022-10273-7

Hypothesis testing, type I and type II errors: Industrial Psychiatry Journal. (n.d.). Retrieved 05 May 2024, from https://journals.lww.com/inpj/fulltext/2009/18020/Hypothesis_testing,_type_I_and_type_II_errors.13.aspx

Idescat. Estimacions de població. Població. Per sexe i edat. Catalunya. (n.d.). Retrieved 19 May 2024, from https://www.idescat.cat/pub/?id=ep&n=9123

Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study | The Review of Financial Studies | Oxford Academic. (n.d.). Retrieved 10 March 2024, from https://academic.oup.com/rfs/article-abstract/25/4/975/1579400?redirectedFrom=fulltext

Kopa, M., & Post, T. (2009). A Portfolio Optimality Test Based on the First-Order Stochastic Dominance Criterion. *Journal of Financial and Quantitative Analysis*, *44*(5), 1103–1124. https://doi.org/10.1017/S0022109009990251

Noordzij, M., Dekker, F. W., Zoccali, C., & Jager, K. J. (2011). Sample Size Calculations. *Nephron Clinical Practice*, *118*(4), c319–c323. https://doi.org/10.1159/000322830

Singh, A., & Masuku, M. (2014). Sampling Techniques and Determination of Sample Size in Applied Statistics Research: An Overview. *International Journal of Commerce and Management*, *2*, 1–22.

Software estadístico. (n.d.). Retrieved 14 March 2024, from https://www.jmp.com/es_es/home.html

Stock market risk—Passive Investing Australia. (2020, February 16). https://passiveinvestingaustralia.com/stock-market-risk/

The empirical risk–return relation: A factor analysis approach—ScienceDirect. (n.d.). Retrieved 14 March 2024, from https://www.sciencedirect.com/science/article/abs/pii/S0304405X06001413?via%3Dihub

The performance of individual investors in structured financial products | Review of Quantitative Finance and Accounting. (n.d.). Retrieved 27 March 2024, from https://link.springer.com/article/10.1007/s11156-014-0479-8

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co. http://www.gbv.de/dms/bowker/toc/9780201076165.pdf

Z-two-tails.pdf. (n.d.). Retrieved 24 May 2024, from https://www.sjsu.edu/faculty/gerstman/StatPrimer/z-two-tails.pdf

10. ANNEXES

1. DATA ANALYSIS PROCEDURE

As the ultimate objective revolves around extracting meaningful information from data, it is essential to ensure that the process is meticulously planned. This systematic process enables researchers to analyse data with the required precision and rigor. For this purpose and in line with Google Data Analytics Program[5], all data analysis study must contain six fundamental steps. Let us briefly break down each of the stages in the following paragraphs and how these are being implemented in the study.

1.1 Ask

To start revolving a certain topic, effective questions need to be formulated to understand and define the problem at hand. At the end, researchers must comprehend the goal of the analysis to present insights cohesively. After providing sufficient context, the hypothesis is clear: to clarify if bank guidance enhances portfolio optimality or not.

1.2 Prepare

Then, researchers must deliberate with regard to the nature and requirements of data. To successfully respond to the problem at hand, it must be critically decided upon: the source and location of data to be used, the type of data needed, the most suitable survey questions to formulate, the verification that data is unbiased and credible, among other considerations related to the collection of data observations.

In this particular thesis, information is collected through asking questions to a representative subset of individuals via web, thereby these are online surveys. Given that this survey instrument has been developed from origin and for the explicit purpose at hand, it is classified as a primary data source. Its advantages encompass: to directly address the research questions specific for this study, to verify data quality and integrity, to ensure that data is non-obsolete in time, etc.

[5]https://www.coursera.org/professional-certificates/google-data-analytics

1.3 Process

After the survey has been conducted via Microsoft Forms[6], this software offers the possibility to export the collected data into a spreadsheet file, enabling researchers to manage and organize information in a familiar format. Thus, while Microsoft Forms provides storage and security systems, Excel facilitates manual tasks employed in this research associated with data integrity, such as: finding incorrectly entered data, excluding repeated values, adding/removing rows or columns for structure purposes, clean possible errors, among other. All aiming to export this data table to JMP[7], the statistical software employed to statistically analyse and visually represent the insights on portfolio optimality. Then, it must be ensured that the automatic classification by JMP of the type of variables is being classified accordingly to its nature and measurement scale. Please note that variables can be nominal if they represent categories with no natural order or ranking; ordinal if categories are naturally ordered; or continuous if variables can take any value within a range.

Furthermore, there are certain variables that might require to be reorganized, constructed or codified to facilitate its proper analysis and management. For instance, when examining the variable age, it is helpful to group it into intervals rather than treating it as a continuous variable. More data organizational tasks include to categorize multiple variables into a single construct. At the end, the process stage of data analysis, elaborated through JMP, is crucial to prepare and ensure the dataset's readiness is ready for analysis and statistical techniques.

1.4 Analyse

By implementing statistical techniques such as descriptive statistics when providing context regarding variables, cross-tabulation when unveiling patterns, logit regression when testing the thesis hypothesis, the dataset at hand is being tested and studied empirically. This data-driven study enables the correct identification and unveiling of the patterns and trends that are hiding behind opinions, behaviours and numbers. This bring us to the following discussion related to reporting these findings in a comprehensive way.

[6] https://forms.office.com/e/cw1cnHbKzp
[7] https://www.jmp.com/es_es/home.html

1.5 Share

As outlined in the research of (Tukey, 1977), "*effective visualizations will allow a scientist both to understand their own data and communicate their insights to others*". Hence, this step plays a significant importance within the data analysis process since it focuses on summarizing and depicting with graphical representations the outputs previously elaborated in the analysis. Ultimately, it is not just about employing fancy statistical techniques, but successfully reporting these findings. To achieve this, the thesis provides for each of the analyses, the creation of effective visuals and a clear communication; ensuring that investors and readers can easily follow up. Some examples included in the thesis are: distribution tables, histograms, box plots, contingency tables, regression plots, etc.

1.6 Act

Finally, this stage is tailored for those investors who after being informed in an empirical way, they must carefully consider the specific findings and opportunities of the study to make more wisely and data-driven decisions. Decisions aiming at enhancing their investment strategies but ultimately, to attain their desired financial goals.

2. CATALAN TO ENGLISH TRANSLATIONS

Catalan	**English**
Empresa i negocis	Business and Management
Feines no qualificades	Unskilled positions
Tecnologia	Technology
Altres	Others
Salut	Health
Educació	Education

3. STATISTICAL OUTPUTS OBTAINED FOR:

3.1 Gender distribution

Gender	Count	Probability
Male	56	55%
Female	46	45%
TOTAL	102	100%

3.2 Age interval distribution

Age interval	Count	Probability
18-22	40	39%
23-27	21	20%
28-32	12	11%
33-37	6	6%
38-42	2	2%
43-47	5	5%
48-52	4	4%
53-57	6	6%
58-62	4	4%
63-67	2	2%
TOTAL	**102**	**100%**

3.3 Job position distribution

Age interval	Count	Probability
Business and Management	31	30%
Unskilled positions	26	25%
Technology	14	14%
Others	14	14%
Health	11	11%
Education	6	6%
TOTAL	**102**	**100%**

3.4 Following bank suggestions

Test Probabilities

Level	Estim Prob	Hypoth Prob
1	0,45098	0,5
0	0,54902	0,5

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	0,9820	1	0,3217
Pearson	0,9804	1	0,3221

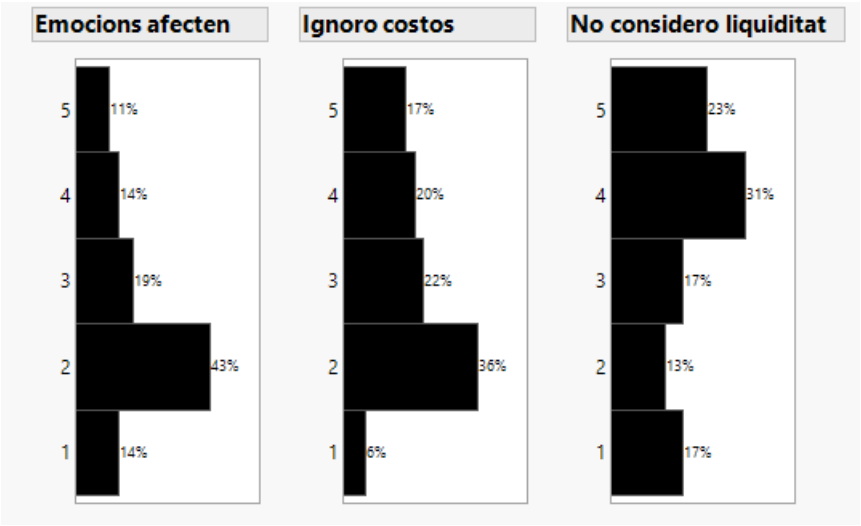
Method: Fix hypothesized values, rescale omitted

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
1	46	0,45098	0,372251	0,532243	0,900
0	56	0,54902	0,467757	0,627749	0,900
Total	102				

Note: Computed using score confidence intervals.

3.5 SAFS component decomposition



3.6 CIs for the investment horizon plan

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
1	15	0,14706	0,098522	0,213835	0,900
1-3y	28	0,27451	0,208364	0,352309	0,900
3-5y	10	0,09804	0,059509	0,157342	0,900
5-10y	17	0,16667	0,114757	0,235803	0,900
10y	32	0,31373	0,243796	0,393281	0,900
Total	102				

Note: Computed using score confidence intervals.

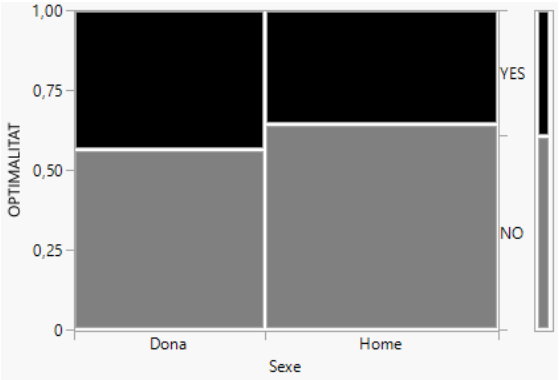
3.7 Geo-investment allocation table

Regional diversification	Count	Probability
Low	44	43%
Neutral	34	33%
High	24	24%
TOTAL	**102**	**100%**

3.8 Hypothesis testing table

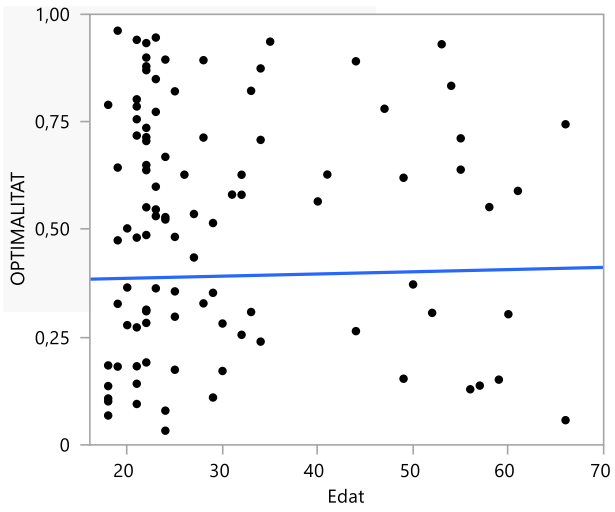
Section	Null Hypothesis (H0)	Alternative Hypothesis (H1)
Sociodemographic factors on portfolio optimization	Gender or Age or Job position exert no significant influence on portfolio optimization	Gender or Age or Job position exert a significant influence on portfolio optimization
Correlation between age and risk tolerance	There is no notable correlation between age and risk tolerance	There is a notable correlation between age and risk tolerance
Correlation between self-awareness financial score and portfolio optimization	Self-awareness score does not exhibit a significant relationship with portfolio optimization	Self-awareness score exhibits a significant relationship with portfolio optimization

3.9 Gender and Optimality relationship



Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0,638	0,4244
Pearson	0,639	0,4242

3.10 Age and Optimality relationship



Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0,009232	1	0,018463	0,8919
Full	68,300485			
Reduced	68,309717			

3.11 Age and Risk tolerance relationship

Linear Fit

RISK = 3,7364235 - 0,0268587*Edat

Summary of Fit

RSquare	0,083529
RSquare Adj	0,074364
Root Mean Square Error	1,166112
Mean of Response	2,920392
Observations (or Sum Wgts)	102

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	12,39356	12,3936	9,1141
Error	100	135,98162	1,3598	Prob > F
C. Total	101	148,37518		0,0032*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3,7364235	0,29393	12,71	<,0001*
Edat	-0,026859	0,008897	-3,02	0,0032*

Summary Statistics

	Value	Lower 95%	Upper 95%	Signif. Prob
Correlation	-0,28901	-0,45776	-0,10017	0,0032*
Covariance	-4,56867			
Count	102			

Variable	Mean	Std Dev
Edat	30,38235	13,04223
RISK	2,920392	1,212048

3.12 SAFS and Optimality relationship

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	5,049255	1	10,09851	0,0015*
Full	63,260462			
Reduced	68,309717			

RSquare (U)	0,0739
AICc	130,642
BIC	135,771
Observations (or Sum Wgts)	102

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Unit Odds Ratio	Odds Ratio
Intercept	1,85385111	0,79081	5,50	0,0191*	.	.
FINANCIAL SCORE	-0,7815455	0,2653237	8,68	0,0032*	0,4576981	0,04388503

3.13 Following bank suggestions and optimality relationship

N	DF	-LogLike	RSquare (U)
102	1	1,3037096	0,0191

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	2,607	0,1064
Pearson	2,606	0,1065

3.14 Multiple logistic regression model

Effect Summary

Source	Logworth	PValue
Sector treball	3,351	0,00045
DIVERSF.	1,616	0,02420
FINANCIAL SCORE	1,466	0,03423
Font d'info	0,627	0,23626
RECOMENACIÓ BANC	0,480	0,33151
Sexe	0,389	0,40824
RISK	0,173	0,67179
TIME	0,063	0,86507
Edat	0,000	0,99895

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Sexe	1	1	0,6839128	0,4082
Edat	1	1	1,71765e-6	0,9990
Sector treball	5	5	22,3665688	0,0004*
RISK	1	1	0,179515	0,6718
TIME	1	1	0,02887422	0,8651
DIVERSF.	1	1	5,08028504	0,0242*
FINANCIAL SCORE	1	1	4,48341719	0,0342*
RECOMENACIÓ BANC	1	1	0,94300518	0,3315
Font d'info	6	6	8,02525142	0,2363

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	32,913294	18	65,82659	<,0001*
Full	35,396423			
Reduced	68,309717			

RSquare (U)	0,4818
AICc	118,061
BIC	158,667
Observations (or Sum Wgts)	102