POLITECNICO DI TORINO

DIGEP

Master Degree in

“Engineering and Management”

Finance Field

Master’s Degree Thesis

The study of Households' Financial Liabilities: households' choice on mortgages, considering the different composition of families, during the period 2012-2019

Supervisor
Prof. Riccardo Calcagno

Candidate
Alberto Caforio

Academic Year 2023/2024

ACKNOWLEDGEMENTS

I feel obliged to say a few words to take stock of these two exciting years now that I have reached the end of my course.

I would like to thank Professor Riccardo Calcagno, thanks to whom, first of all, I have been able to broaden my horizons in new disciplines; I would like to thank him for his availability and the help he has given me in writing my thesis, and for allowing me to delve into one of the subjects I had found interesting during my studies.

I would like to thank the entire corps of professors with whom I have had the pleasure of learning, interacting and improving myself, both as a person and in terms of knowledge, during these years of study.

I would also like to thank my colleagues with whom I have shared this journey between lectures, projects and afternoons of studying.

Thanks to all the flat mates and amiconi of Link, current and not, with whom we have shared so much, from the Watergate to the coffee pot, beautiful experiences that are definitely worth telling.

Thank you to all my friends. Thank you all for being an indispensable and unforgettable part of this journey.

Grazie ai miei parenti, zie, zii, cugini e cugine, nonna. Grazie per l'attenzione, la disponibilità e la dolcezza che mi avete dato fin dal giorno zero.

Grazie a mio fratello, a mia madre e a mio padre, grazie perché vi devo tutto; grazie per avermi tirato su quando ero giù, per avermi motivato quando ero a terra, per avermi incoraggiato quando ero a pezzi e per aver festeggiato con me ogni traguardo, grande o piccolo che fosse. Vi voglio bene.

Now that the time to say thank you is over, it is time to get on with the world of mortgages. Have fun.

ABSTRACT

This study, based on a survey conducted by the Bank of Italy, analyses the propensity of Italian households to choose between fixed-rate and variable-rate mortgages in a context of relative economic stability between 2012 and 2019. .

The methodological approach adopted for the selection of participating households aims at ensuring a meaningful representativeness through inclusion criteria that take into account different socio-economic and cultural strata. The following sections examine the extent to which the choice of one type of tariff over another can be influenced by socio-economic and cultural variables. The aim of this review is to provide an understanding of the dynamics driving households' financial choices in a period of economic stability, in order to identify possible patterns of behavior and to contribute new insights and analysis to the existing literature.

Our empirical results suggests that while certain economic and housing variables, such as annual income and house value, have a potential impact on mortgage type preference, their effects are not necessary crucial and need more voices in the analysis to confirm their own importance. This suggests that, within the observed period of relative economic stability, Italian households' choice between fixed and variable rate mortgages may be less sensitive to fluctuations in these financial factors than hypothesized. The socio-economic strata and cultural factors could underlie the nuances in these financial decisions.

These open-ended conclusions call for further research into the cultural and behavioral factors that influence household financial decisions in stable economic climates.

SUMMARY

1. INTRODUCTION 1

1.1 HOUSEHOLD FINANCE: AN OVERVIEW 1

1.2 OVERVIEW: TRENDS IN THE ITALIAN ECONOMY BETWEEN 2012 AND 2019 7

2. LITERATURE REVIEW 9

2.1 OPTIMAL MORTGAGE CHOICE: OVERVIEW & NUMBERS 9

2.2 MORTGAGES: AN OVERVIEW 12

2.3 MORTGAGES: DEFINITION 12

2.4 “OPTIMAL MORTGAGE CHOICE “: AN OVERVIEW 14

2.5 “CHOICE OF MORTGAGE CONTRACTS?”: AN OVERVIEW 16

2.6 “ARE ADJUSTABLE-RATE MORTGAGE BORROWERS BORROWING CONSTRAINED?”- AN OVERVIEW 17

3. METHODOLOGY 19

3.1 DATABASE OBSERVATION: AN OVERVIEW 19

3.2 VARIABLES : AN OVERVIEW 20

3.3 PAIRWASE CORRELATION TABLE 24

3.4 INTEREST RATES: COLLECTION OF DATA 27

3.5 CATEGORISATION OF THE MORTGAGE BY " RATA_AR " 29

3.6 REGRESSION MODELS 33

4. RESULTS OF THE EMPIRICAL RESEARCH 37

4.1 EMPIRICAL RESULTS: PROBIT REGRESSION WITH ALL THE CHOSEN VARIABLES 38

4.2 EMPIRICAL RESULTS: FOCUS ON CATEGORICAL VARIABLES (ETA5, AREA3, ANNO) 43

4.3 EMPIRICAL RESULTS: FOCUS ON ECONOMIC VARIABLES (DEBI12A, RATA_AR, Y) 45

4.4 EMPIRICAL RESULTS: FOCUS ON HOUSING PROPERTY CHARACTERISTICS 47

4.5 COMMENTS ABOUT THE EMPYRICAL RESULTS: PROBIT REGRESSIONS ISSUES 49

4.6 COMMENTS ABOUT THE EMPYRICAL RESULTS: COMPARISON WITH THE LITERATURE GUIDELINES 51

5 IMPROVEMENTS: AN OVERVIEW 53

6 CONCLUSION 54

7 BIBLIOGRAPHY 55

APPENDIX A 57

DATABASE ANALYSIS ON SOCIAL AND DEMOGRAPHIC BASE. 57

APPENDIX B 71

INTRODUCTION AND OVERVIEW ON DEBT FINANCING VARIABLES 71

1. INTRODUCTION

1.1 HOUSEHOLD FINANCE: AN OVERVIEW

The definition (1) of **Household Finance** has been given its birth in the recent years, and it's associated with the study of how households use the financial instruments to achieve their personal goals.

Household Finance has been relatively **under-researched** and little considered (1) as a financial sector in its own right. However, since the beginning of the new millennium, thanks in part to the greater fluidity of information and knowledge of the key players, it has experienced a boom in terms of insights.

Household Finance has **two macro branches**, focusing on two different aspects of the discipline (1):

- **Normative household Finance** studies, through purely theoretical models, the typical behaviour that individuals should adopt in their consumption, saving, investment and non-investment decisions in order to maximise their utility function. However, as we have studied and have been able to learn in various finance courses, the theory does not always correspond to reality - and this is where Behavioural Finance comes into play;

- **Positive Household Finance**, more precisely, highlights what are the real choices made by the actors in question and compares them with what are the theories of the subject matter mentioned above. As we know, it is not always the household that makes decisions to maximize its utility function, but the financial actor as a whole. These discrepancies may be the result of simple mistakes caused by "ignorance" of the issue, or by social, cultural and economic factors in each family.

As expected, until a few decades ago, household finance was considered a branch of corporate finance, whereas recently it has gained a place and a new field of its own.

The separation from pure asset pricing and corporate finance is primarily due to the very **nature of the study (1)**: the subject whose financial behaviour is being studied has different reasons and priorities than a subject in the economic sector: he has to make different choices in the area of how to pay for consumer goods, for everyday needs.

A family's main indebtedness is also the pure mortgage debt (the subject of our research), with all that this entails, such as choosing between the various options available. A family's main income is, in the vast majority of cases, the sum of wages earned in everyday work, an asset that is difficult to insure, risky and far removed from the specificities of pure finance.

Moreover, the **specificity of the subject itself (1)** lies in the heterogeneity of family capabilities, which can vary according to different economic, social and cultural contexts.

In fact, according to a **study carried out by the Bank of Italy (2)**, the financial literacy of households was measured in 2023 by means of a public survey and the result obtained, through estimations and rounding, is that the overall average financial **literacy of the respondents is about 50%,** a figure that can certainly be improved and which, among other things, does not reflect the homogeneous preparation of the respondents.

The survey (2) shows that, on average, the level of general knowledge varies according to the level of education attained, with an increase (albeit slight) as the level of education rises, and that it also varies according to the gender of the person interviewed: female respondents, in fact, report an average level that is about 2 per cent lower than that of their male counterparts.

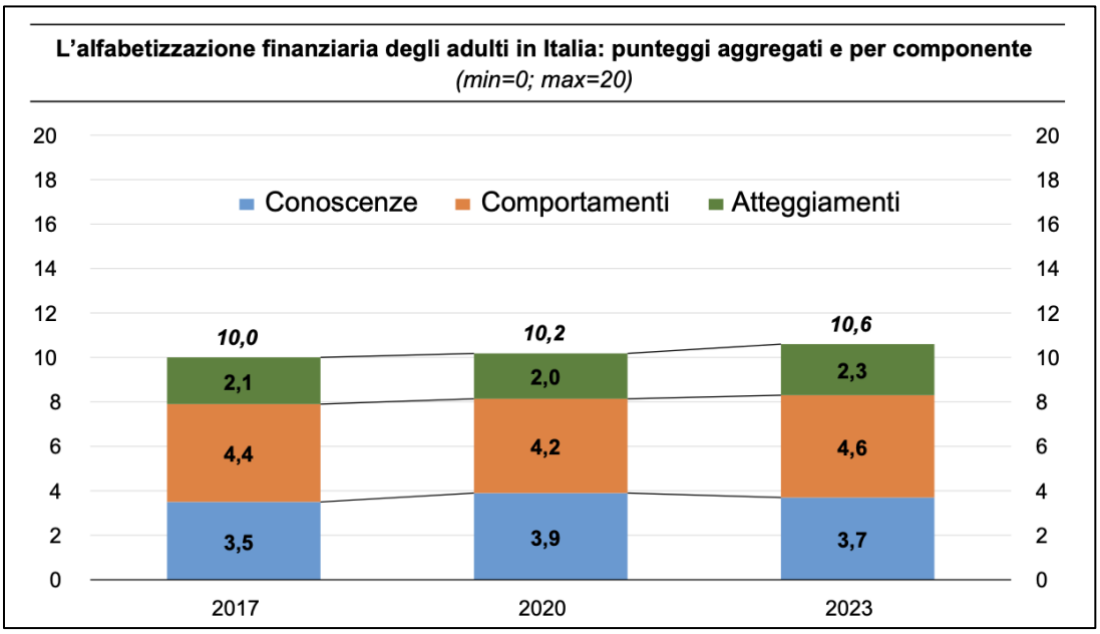


Figure 1 - Financial Education of Italian Hoseholds.

As can be seen in the figure above, knowledge has improved slightly compared to previous years, but there is still a long way to go.

A further reason why household finance needs to be studied as a separate branch of the financial sector is also due to the **laws that may apply to certain financial instruments** (1) that are aimed at households. While regulations that are already specific to corporate finance may be associated with some available financial instruments, others that are exclusively related to households can and should require a different regulatory interpretation.

Indeed, the financial decisions of households have in the past been, and may in the future be, influenced by the activities of governments in different countries; the dynamics of governments must indeed be taken into account; and, last but not least, there is the question of the extent to which the activities of governments in different countries may have an impact on the financial decisions of households.

Last but not least, there are the **figures relating to the instruments targeted at the family**, which indicate the extent to which this sector needs to have its own autonomy.

Once again, we can see that the **financial assets and liabilities of households represent a large part of the Italian economy**, thanks to a survey conducted by the Bank of Italy (3).

At the end of 2021, for example, the **net worth of Italian households**, adjusted in this way for their financial liabilities, **amounted to around EUR 10,010 billion**. This is around 8.7 times their disposable income. The most important financial instruments in this calculation are, of course, property values, but also other instruments such as policies and insurance (3). There has been a slight year-on-year increase in household net wealth, the start of which was of course the recovery from the 2008-2012 crisis in Europe.

Financial liabilities (3), which are deducted from net wealth, **amount to around 967 billion**. They also account for a large part of the debt in our country. In contrast to net wealth, the figures for liabilities appear to have remained constant over the last few years, suggesting a slight increase in both financial and non-financial assets.

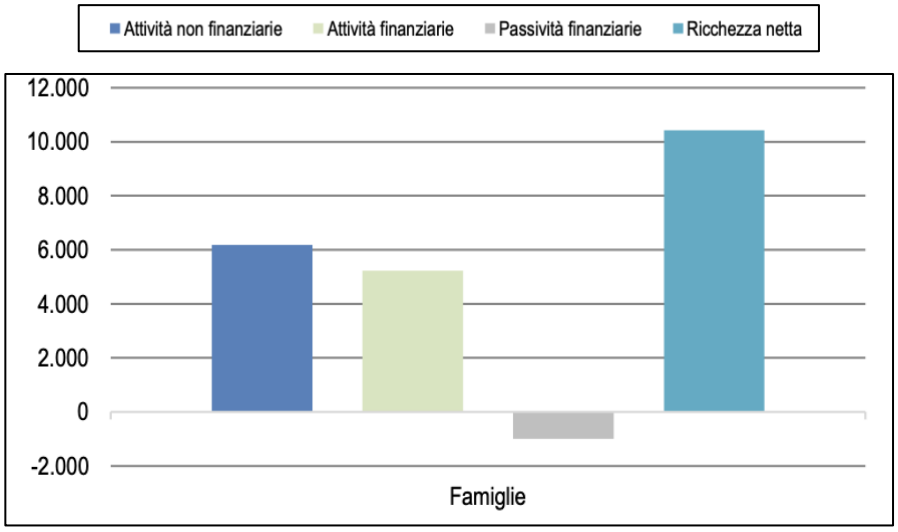


Figure 2 - Financial Activities & Liabilities of Italian Households.

Once we have understood and defined the reasons why the study of household finance has finally managed to gain independence from commercial and corporate finance, we need to **mention the factors** that have led, especially in the last two decades, to raise the issue of the lack of study of this subject and to **give it a boost in popularity** (1).

The most obvious reason is common to all the new industries that have been the subject of growth and development over the past two decades. **Digitization and technological innovation** have had the effect that more data on the different types of financial instruments in use by households have been able to be collected in the first place. Before the internet exploded, this information was collected only by surveying and interviewing households:

the **surveys turned out to be inaccurate** in a large number of cases, regardless of whether the households were more or less wealthy. As a result, the low usefulness of the sampled interviews led to a lack of interest in these analyses, and the few surveys that were conducted produced unsatisfactory results.

With the advent of digitalization, there has been a shift to a more **truthful circulation of data (1)**. The various financial institutions store this information in their private databases, which analysts can then search. Undoubtedly, the movement and development of household finance has been greatly stimulated by this change in the storage and collection of information.

What is more, the **household** is becoming more **active** and more **involved** (1) in the financial decisions that need to be taken. It is no longer a passive actor in decisions that in the past were entrusted almost exclusively to experts in the sector or to financial institutions.

The aspects and phenomena described above are just some of the reasons why household finance is a key area of study, both for experts, analysts and those working in the field, and for the households themselves, who are directly affected by the phenomena described.

However, our work will not focus on issues that can be traced back to the entire field of household finance. Instead, **we will concentrate** on financial liabilities and, in particular, on the **choice of mortgage types by households and the factors that may have intervened in their choice**.

The period we are going to analyze is going to be the choices made by households in the sample survey carried out by the Bank of Italy between 2012 and 2019. We'll discuss in the next paragraph the relevance of this period in terms and why this 7-years span can be considered "anormal".

1.2 OVERVIEW: TRENDS IN THE ITALIAN ECONOMY BETWEEN 2012 AND 2019

From 2012 to 2019, Italy has enjoyed a period of relative economic stability, which is **quite exceptional** given its past turbulence, in particular its 2008-09 global financial crisis and subsequent European sovereign debt crisis.

Several factors, including the expansionary monetary policy of the **European Central Bank (ECB)**, domestic structural reforms and a favorable global environment, have contributed to this period of stability.

Through various monetary policy measures, the European Central Bank has played a crucial role in ensuring the stability of the eurozone and, by extension, Italy. Since 2012, the ECB, led by Mario Draghi, has pursued a policy of extremely low interest rates, accompanied by **programmes to purchase assets** (4), known as **Quantitative Easing (QE)**. The aim of this policy has been to stimulate inflation to levels close to, but below, 2%. This has supported the eurozone economy.

With the economic recovery in the US and the continued growth of emerging economies such as China, the period also benefited from a relatively favourable global economic environment. These conditions favoured foreign demand for Italian goods, which in turn supported exports and thus the Italian economy.

In response to previous crises, Italy has implemented several structural reforms with the aim of improving the competitiveness and flexibility of the economy. These reforms have involved the labour market, the pension system and the banking sector, among others. An important example is the 2012 "Fornero reform" (5), which introduced **changes to the Italian pension system** with the aim of improving its sustainability.

The period of recovery and monetary policy following the recent global crisis is clearly having an **impact on the two interest rates** that lenders usually use as the **basis for determining the fixed and floating rates for mortgages**.

- **EURIRS (6)**: The interbank reference rate is used as the indexation parameter for mortgages with a fixed interest rate. It is published daily by the European Banking Federation and is equal to a weighted average of the rates at which banks operating in the European Union execute interest rate swaps;

Figure 3 - EURIRS interbank rate between 2014 and 2019

- **EURIBOR (7):** Interbank reference rate, published daily by the European Banking Federation as a weighted average of the interest rates at which banks operating in the European Union are lending deposits. It is used to index variable rate mortgages.

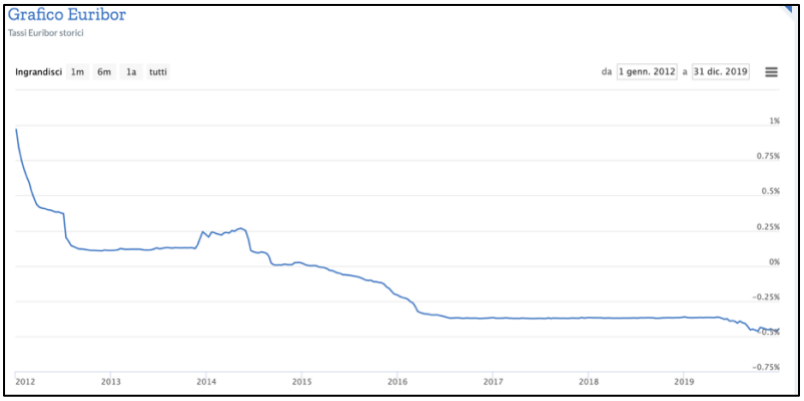


Figure 4 - EURIBOR interbank rate between 2012 and 2019

2. LITERATURE REVIEW

In the previous chapter, we dealt with and introduced the discipline of household finance, without going into the specifics of the two macro areas studied, namely household assets and household liabilities.

In this work of analysis, we will mainly deal with a specific branch of the second macro area just described, namely the **optimal choice of mortgage**, which is present in the study of the broader topic of households' borrowing decisions.

2.1 OPTIMAL MORTGAGE CHOICE: OVERVIEW & NUMBERS

The choice of the right mortgage is undoubtedly one of the most important aspects of household finance: the **vast majority of average families**, in terms of income and composition, tend to have only one home in their lifetime, and the **most important decision to be taken is whether or not to take out a mortgage** to buy or renovate the property that will serve as their main home.

Moreover, it is very important to consider that the opening of a mortgage generally **represents the pivot of all the debts accumulated by a family**, and this phenomenon is valid both outside our country and in Italy.

According to a survey conducted by **QuiFinanza (9)**, the pandemic intercession triggered a revival of the real estate sector, culminating in an escalation of both the amount of domestic aggregates taking out mortgage loans to purchase houses and the total volume of credit disbursed. In April, as attested by the Federazione Autonoma Bancari Italiani and the Bank of Italy, a **total of EUR 425 billion was recorded**, an **increase of 13.7% compared with the end of 2017**.

This economic phase was not only helped by a fall in property prices, but also by a series of relief measures announced by the executive during the pandemic, such as the introduction of the Bonus Prima Casa for people under 36 years of age and, in the longer term, by an increase in the liquidity available to Italian citizens in the 2020-2022 biennium. **As a result, at present 3.5 million households, or 13.6 percent of the total, are indebted to a mortgage**. This form of indebtedness appears to be the most widespread (9), given that the total number of households indebted in any way, including personal loans, rises to 6.8 million, of which a significant proportion, namely 3.5 million, took out mortgages to buy their homes.

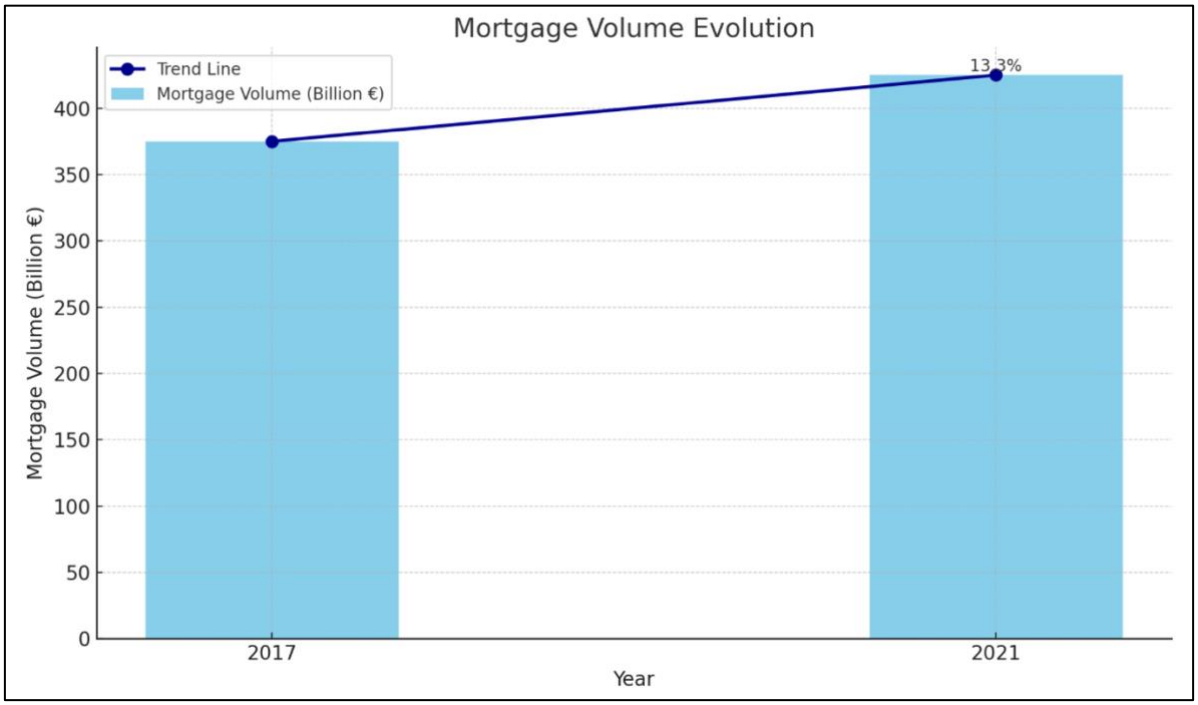


Figure 5 - Mortgage Volume Evolution

Thus, there is an urgent need to study and deepen the issue of mortgage underwriters, both from the point of view of figures (which we have already discussed) and from the point of view of possible benefits for the household itself.

For the household that has not yet taken out a mortgage and will have to do so in the near future, and which will therefore be able to recognize the information provided by financial institutions, filter it and decide on the basis of the knowledge it has acquired, it is **useful to know what the variables are that may determine the choice of one type of mortgage rather than another (1)**.

In addition, the study of the behaviour of families in the choice of a mortgage provides a measure of the level of "**Financial Education**" discussed in the opening chapter of this report:

Will the family have acted efficiently, following the guidelines? Have they managed to maximise their 'utility function'? Instead, are there biases in the family's behaviour that do not allow for an optimal choice, that condition the outcome?

Despite the fact that this is such an important and useful topic for learning, we are surprised at **how little research has been done** (1) on the topic of 'optimal choice of mortgage' in the past and even today.

The reasons why there is a lack of analytical models that can be studied in a general way to make the choice of a mortgage more efficient for households can be many and varied:

The main obstacle over the years has certainly been the **difficulties of finding specific data** on the type of mortgage selected for each household, together with the household's own characteristics.

These data, which are in fact sensitive material, are hardly ever disclosed by the financial institutions themselves and are not made public in order to study possible patterns or trends.

As a result, most of the information available is based on phenomena that can be traced back to surveys carried out for **purely informative purposes**, such as the one carried out by the **Bank of Italy** when it drew up the SHIW.

It should be borne in mind, however, that the majority of surveys are carried out by telephone or physically, through direct questioning of the head of the household or a family member, and it would therefore be a good idea to have a cursory look at the information at our disposal and validate it before proceeding with the analysis.

2.2 MORTGAGES: AN OVERVIEW

In the two papers dealing with the problem of the optimal choice of mortgage, in order **not to complicate the work** and the subsequent considerations, reference is made exclusively to **two types of mortgages available** to the household: **fixed-rate mortgages** (**FRMs**) or **adjustable-rate mortgages** (**ARMs**).

Although we know that there are several viable options outside these two, in order to better understand the work done by the various authors of the papers, we thought of opening a small mirror in which we refer to the definition of a mortgage, to the 2 types chosen and to the possible methods of repaying the debt (in short, the main ones).

This mirror will be useful, as expected, both to outline a general picture of what the mortgage situation is today and to make more coherent the results that will be presented below.

2.3 MORTGAGES: DEFINITION

A **mortgage** (10) is a **long-term loan** that allows an individual or family to **buy a property, usually a house**, by paying for its value over time. The bank or credit institution providing the loan has a lien on the property until the borrower has repaid the debt in full.

The two main categories of mortgage are distinguished by the type of interest rate applied: **fixed rate mortgage** (**FRM**) e **Adjustable-rate mortgage** (**ARM**).

- **FIXED RATE MORTGAGE** (**FRM)**: It provides certainty for the financial planning of the borrower, who undertakes to repay the amount borrowed in **constant instalments**, with an **interest rate** that remains **unchanged (10)** throughout the duration of the contract. This type of loan is particularly suitable in contexts of low interest rate volatility or when the borrower prefers to avoid the risk of possible credit cost increases over time;

- **ADJUSTABLE-RATE MORTGAGE** (**ARM**): It is characterized by an **interest rate** that **may vary (10)** according to the evolution of certain reference indices, such as the Euribor rate or the reference rate of the Central Bank. In this case, the **repayment instalments may vary over time**, which makes the variable rate mortgage more exposed to the fluctuations of the financial markets. On the other hand, this type of mortgage can be advantageous in a scenario of falling interest rates, as it allows the borrower to benefit from more favourable repayment conditions.

As far as the repayment terms are concerned, the amortization plan of the loan defines the way in which the borrower undertakes to repay the capital and pay the interest on the amount lent. The **most common plan** is the "**French plan**", which involves constant deferred repayments composed of a capital component and an interest component, with the interest component gradually diminishing over time and the capital component increasing. This mechanism allows the Bank to reduce risk over time and the borrower to become more debt sustainable.

2.4 “OPTIMAL MORTGAGE CHOICE “: AN OVERVIEW

As expected, one of the few studies on the subject of 'optimal mortgage choice' is the work of **Campbell and Cocco** in their article entitled **'Optimal Mortgage Choice' (11).**

The paper presented by the two authors thoroughly examines the **problematic issue** of "**choosing a mortgage**" and, as expected, emphasises that choosing a mortgage is not trivial. There is an inherent tension between the accessibility of the problem, given its almost universal relevance, and its inherent complexity, given the need to deal with financial uncertainty, credit constraints and uninsurable risks. Indeed, providing tools for financial economists to provide evidence-based financial advice to prevent homeowners from falling prey to potentially harmful advice seems to be the ultimate goal of the two researchers.

The paper begins this work (11) by assessing the impact that the two different types of mortgage contracts (introduced above) can have on household welfare. It distinguishes between fixed and variable rate mortgages.

The **paper highlights (11)** how the nature of the mortgage contract chosen, whether it is a fixed-rate mortgage (FRM) or an adjustable-rate mortgage (ARM), plays a **crucial role** in shaping the **financial risk profile** to which a household may be exposed. In particular, **FRM** contracts tend to transfer **equity risk** to the borrower (11). This stabilizes payments but potentially exposes the borrower to losses in the event of adverse house price movements. On the other hand, **ARMs** introduce significant **income risk**, as interest rate fluctuations can directly affect the sustainability of mortgage payments, particularly in adverse economic scenarios where low incomes and house values may coincide with high interest rates (11).

In fact, according to Campbell and Cocco, the **difference in risk between the two types of mortgage** (ARM and FRM) is **crucial for understanding the dynamics of households' mortgage choice**, as it highlights a trade-off between payment

stability and flexibility in the face of economic changes. Research suggests that households with certain characteristics, such as **higher income stability**, l**ower risk aversion**, more favorable terms in the event of default and a **higher probability of transfer**, may find **ARMs more advantageous** despite the income risk they entail (11). This perception is consistent with empirical evidence confirming the propensity of certain demographic segments, such as **married couples** or co-borrowers with presumably **more stable incomes**, to **use ARMs**.

The link between individual income risk and corporate risk management strategies offers an interesting perspective that extends beyond personal finance, suggesting that lessons learned in the context of residential mortgages may be relevant in the broader context of financial risk management.

This broad analysis highlights how important it is to view **mortgage choice** as more than an individual financial decision, but as an important part of wider discussions about **risk management** and **household financial security** (11). This approach reflects an important paradigm in financial economics that recognizes the need to integrate well-established theoretical knowledge with a practical understanding of market dynamics and human behavior (the behavioral finance that we have been discussing in our academic studies).

The overall picture emerging from the analysis provides fruitful ground for future research, focusing on international differences in mortgage markets and the potential applications of these dynamics in corporate and risk management settings. The depth of the analysis (11) reflects the need for a holistic and comprehensive approach to the assessment of financial decisions, taking into account both the current market conditions and the emerging socio-demographic trends.

2.5 “CHOICE OF MORTGAGE CONTRACTS?”: AN OVERVIEW

We are now going to take a look at the article written by the two authors, **Coulibaly & Li (12)**, who have proposed an analysis of the following of the **determinants of household mortgage choice**, using detailed data from the Survey of Consumer Finances (SCF). This approach **overcomes** some of the **limitations of previous studies** and thus provides a **more comprehensive** and reliable perspective on borrowers' behavior.

The **results** (12) from the use of a simple logit model underline that not only price variables and contract terms are key aspects in the choice of a mortgage, but that **affordability** and **financial stress** also play an **important role**. This implies that mortgage decisions are not just the result of cold economic calculations but are also influenced by personal considerations and the current financial situation of borrowers.

A particularly interesting finding is the **importance of risk factors** in tilting borrowers' preference **towards fixed-rate mortgages** (FRMs), especially for those with **higher risk aversion**, **uncertain income** or low likelihood of moving house in the short term (12). This highlights how risk perception and personal risk appetite are key determinants in the choice of mortgage type.

Furthermore, it is striking that attitudes towards risk influence mortgage choice along two distinct dimensions. This underlines the **importance of individual attitudes to risk** in guiding financial decisions. This is further evidence of the importance of individual attitudes to risk as a guide to financial choices.

To sum up, the analysis suggests (12) that the decision to opt for a particular type of mortgage contract is not only an economic one but is also deeply rooted in individual perceptions of risk, personal financial circumstances, and future prospects.

These findings provide a strong empirical foundation highlighting the crucial role of risk perceptions in mortgage choice, **adding to the debate on Behavioral Finance** and providing valuable insights for policy interventions designed to guide consumers towards more informed financial choices appropriate to their risk profile.

2.6 "ARE ADJUSTABLE-RATE MORTGAGE BORROWERS BORROWING CONSTRAINED?"- AN OVERVIEW

In this article, the authors, specifically **Johnson & Li** (13) examine how the **financial** and **demographic characteristics** of borrowers **with adjustable-rate mortgages** (ARMs) compare with those of other borrowers. The focus is on credit constraints. Surprisingly, **ARM borrowers** do **not appear to be more constrained** in their access to credit than other borrowers, based on conventional characteristics such as age, education and wealth status. But several other indicators, such as being more likely to have been refused credit in the past five years, being more likely to default on credit card payments, and using available credit limits more heavily, suggest that ARM borrowers may in fact be experiencing greater financial difficulty.

The paper (13) also shows that the **sensitivity of consumption** to past income is particularly pronounced among ARM borrowers, a dynamic that highlights how changes in **mortgage payments** can have a significant impact on the spending behavior of ARM borrowers. This finding is **consistent** with theories suggesting that **ARM borrowers may have to reduce consumption** in response to **increases in mortgage payments**, potentially reinforcing the depressive effect of interest rates on consumption at the macro level. The article (13) tends to suggest that in *periods of relative economic stability*, or even when *interest rates tend to fall slightly*, households prefer to take out an **adjustable-rate mortgage.**

One of the **unresolved questions** in this analysis (13) is whether it is the nature of the ARM mortgage that causes the greater sensitivity of consumption, or whether it is some third, unobserved factor that determines both the choice of the ARM mortgage and the variability of consumption. These ambiguities open the way for further research to explore more deeply the causal and underlying mechanisms of these financial dynamics.

The authors' contribution to this area of research is significant, as it provides a more mature understanding of the relationship between mortgage types and household financial behaviour and highlights the importance of taking into account a wider range of factors when assessing mortgage decisions and their economic impact.

We have found three different articles dealing with the same subject, but sometimes touching on different points of the subject itself.

All that remains is for us to start the empirical analysis, with the data and information that we will be able to obtain, in order to compare and understand whether the concepts of the models that we have just commented on really fit the empirical analysis that we will carry out in the following chapters.

Here you will find a synoptic table that can summarise the most important pieces of evidence for each of the articles that we have summarised above.

ARTICLE	EMPIRICAL PREDICTION #1	EMPIRICAL PREDICTION #2	EMPIRICAL PREDICTION #3
Campbell & Cocco	Households with higher Income Stability are more likely to choose ARMs	Households with lower risk aversion are more likely to choose ARMs	Households with higher probability to move are more likely to choose ARMs
Coulibaly & Li	Households with lower Income Stability are more likely to choose FRMs	Households with higher risk aversion are more likely to choose FRMs	Younger Households tends are more likely to choose FRMs
Johnson & Li	Households who choose ARMs more likely to save on consumption in response to increasing payments	Households tend to choose ARMs in a period of constant interest rates or even decreasing ones	

Table 1 - Synoptic Table on the Empirical Findings

3. METHODOLOGY

It is now time to describe the **main methods** used in our empirical analysis, listing the variables for each dataset used.

The first clarification concerns the main variables initially considered: they are presented according to their presence in the **different datasets of the SHIW**.

There were many variables in the Bank of Italy's database, some of which were redundant, while others were not useful for our study. In the following, therefore, I enclose a **list of the variables** considered in this part of the work.

Let us recall that the variables in the database are in turn divided into sub-datasets, grouped by relevance. For the sake of simplicity, we will **summarise the main variables considered in a single table**.

3.1 DATABASE OBSERVATION: AN OVERVIEW

As can be seen in more detail in section 'Annex B', the necessary work that led to the inclusion of all households started with the first essential cut-off: to take into account, of course, only the NQUEST codes associated with the reference period of our analysis, making the two variables attributable to the opening or payment of a mortgage: 'DEBI12A' and 'RATA_AR'.

Only unique household codes (NQUEST) where both variables were positive and obviously non-zero were taken into account.

Once the numbers in the SHIW database had been massively skimmed, we proceeded to divide the NQUEST family codes into two macro-groups: the first group formed by the unique codes present only once in the database, and the second group formed by some codes present several times in the database (see Appendix B). In conclusion, there were 277 unique codes with the characteristics listed.

3.2 **VARIABLES : AN OVERVIEW**

Here is the **list of variables** included in our analysis: the first eight describe the main characteristics of household composition, the next eight describe household liabilities to financial institutions, and finally we focus on variables relating to ownership, type, value, ownership type and year of ownership:

VARIABLE	DESCRIPTION	CODING
ANNO	*Survey reference period (year).*	/
NQUEST	*Questionnaire number – household ID*	/
ETA5	*Age class*	1 = up to 30 years; 2 = 31 to 40 years; 3 = 41 to 50 years; 4 = 51 to 65 years; 5 = over 65 years.
AREA3	*Breakdown of the territory into three geographical areas.*	1 = North; 2 = Centre; 3 = South and Islands.
STUDIO	*Title of study*	1 = none; 2 = primary school certificate; 3 = junior high school; 4 = senior high school; 5 = university degree; 6 = postgraduate degree.
Y	*Net disposable income. Variable calcula at household level.*	
FASCIARED DITO	*Net disposable income divided into 3 ba*	1 = up to 28,000eur; 2= from 28000eur to

		50000eur; 3 = from 50000eur
QUALP7N	*Main occupation: professional status*	"EMPLOYEE EMPLOYMENT: 1 = manual worker or similar position; 2 = clerk or teacher; 3 = managerial/executive employee; 4 = executive INDEPENDENT WORK: 5 = freelancer; 6 = entrepreneur, self-employed and co-employed; 7 = NON-PROFESSIONAL'.
DEB12A	*Amount owed to banks or finance companies at the end of the year for the acquisition or improvement of immovable property.*	/
RATA_AR	*Total instalments paid during the year for the acquisition or improvement of immovable property*	/
FASCIAREDDITO	*Net disposable income divided into 3 ba*	1 = up to 28,000eur; 2= from 28000eur to 50000eur; 3 = from 50000eur

TIPOIMM	*Type of property in ownership.*	1 = Dwelling; 2 = Other building; 3 = Agricultural land; 4 = Non-agricultural land
ANCOSTR	*Year of construction of the dwelling or building.*	/
USOIMM	*Main use of the dwelling or building.*	1 = Principal residence; 2 = Holiday residence; 3 = Professional or commercial use; 4 = Rented year-round to individuals or families; 5 = Rented year-round to companies; 6 = Rented part of the year to individuals or families; 7 = Rented part of the year to companies; 8 = Not rented; 9 = Other use.

VALABIT	*Value of the property.*	/
POSS3	*Manner of acquiring the property.*	1 = Purchased; 2 = Inherited; 3 = Purpose-built
ANPOS	*Year in which ownership started*	/

Table 2 - Variable used in the analysis

Once the inputs and variables used have been defined, a pairwise correlation table is created, including the variables shown in the table just shown. This excludes purely input variables, such as the family identification code (NQUEST).

The **correlation coefficients** were obtained using the specific function =**CORRELATION(x,y)** in Excel. The qualitative conclusions seem to be in line with the coefficients found.

3.3 PAIRWASE CORRELATION TABLE

The **correlation matrix (14)** shows the correlation coefficients between two different variables: the **correlation index** evaluates their relationship, the relation. Apart from the correlation coefficient of a variable with itself (which is obviously equal to 1), the index that examines the correlation between two variables must be studied in order to **highlight** any **variables** that are **excessively correlated** with each other, i.e. **dependent on each other**, so that one of them can be discarded before the next analysis. The **sign of the coefficient** naturally indicates the **direction of the relationship** between two variables: a negative coefficient, for example, would imply an inverse relationship between two items (14).

	FASCIA REDD	VALABIT	POSS3	ANPOS	DEB12A	RATA_AR	ETA5	AREA3	STUDIO	QUALP7I	Y
FASCIA REDDITO	1,000										
VALABIT	0,242	1,000									
POSS3	0,070	0,024	1,000								
ANPOS	0,028	0,072	-0,188	1,000							
DEB12A	-0,005	-0,002	0,005	-0,042	1,000						
RATA_AR	-0,042	-0,019	-0,010	0,016	-0,006	1,000					
ETA5	0,202	0,074	0,130	-0,337	0,041	-0,008	1,000				
AREA3	-0,125	-0,079	0,062	0,062	-0,037	0,044	0,010	1,000			
STUDIO	0,436	0,192	-0,032	0,205	-0,037	0,004	-0,121	0,014	1,000		
QUALP7N	0,051	0,103	0,060	-0,164	0,003	0,004	0,303	0,067	-0,088	1,000	
Y	0,469	0,372	0,004	0,058	-0,031	-0,019	0,142	-0,104	0,290	0,120	1,000

Table 3 - Pairwise correlation table

As can be seen, some of the above **variables are missing**, i.e. **TYPEIMM** and **USOIMM**: this choice was made because for Hp we excluded in the study debts to financial institutions that were not for residential property. If the study excludes purely agricultural land, the variable USOIMM has the value =1 in all rows, i.e. main residence. These two variables were therefore discarded before drawing any conclusions from the table.

Only a few variables appear to be correlated in some way, perhaps only one with a very strong correlation if we take a quick look at the results in the table. **Let's comment the results** in the general pairwise correlation table, by inspecting the correlations value **higher than |0.20|**:

- The **correlation** between **STUDIO** and **FASCIAREDDITO**, which represents income, is **0.436**, indicating a moderate positive association between educational attainment and household income. From an engineering perspective, this relationship can be interpreted as demonstrating how **human capital**, as measured by educational attainment, can have a **direct economic impact** on an **individual's income**. This concept is consistent with human capital theories which postulate that investment in education improves productivity and hence potential earnings;
- As for **ETA5** and **FASCIAREDDITO**, we find a **correlation of 0.202**. This could indicate that there is a slight **tendency towards higher earnings with increasing age**. This can be interpreted as a possible indicator of career progression or the accumulation of experience that could lead to higher salaries. However, the correlation is not strong, suggesting that factors other than age play a more important role in determining income;
- The correlation between **Y** (income) and **VALABIT** (monetary value of dwelling) is **0.372**, indicating a positive correlation. A correlation coefficient of 0.372 suggests that there is a **positive relationship** between a household's **income** and the **value of its home**. However, the correlation is not extremely strong. This could imply that as income increases there is a tendency to own higher value homes, but there are other significant variables at play.
- The correlation between **STUDIO** and **Y** is **0.290**, confirming the relationship between education and income from another angle. This reinforces the idea that **human capital** is a **key determinant of income**.
- The **QUALP7N** variable is positively correlated with the **ETA5** variable (**0.303**). This may be quite simple, as the two extremes of the two variables correspond to people over 65 and unemployed respectively, which tends to be a correct correlation. In addition, the relationship between the two variables could be an indication of a **higher proportion** of **self-employed** and managerial occupations **in the older age groups**.

- Analyzing the negative correlation of **-0.337** between **ETA5** (age group) and **ANNPOSS** (years of home ownership), it can be interpreted that there is a moderate **inverse trend** between the **age** of the respondent and the **number of years of home ownership**. This could indicate that younger age groups tend to have a shorter period of ownership than older age groups. One explanation for this correlation could be that younger people are less likely to have owned their home for a longer period, due to factors such as having entered the labour market more recently or having moved due to career opportunities. On the other hand, older age groups may have had more time and more opportunities to settle down in a property for a longer period of time;
- The positive correlation of **0.205** between **STUDIO** (level of education) and **ANNPOSS** (year of home ownership) can provide interesting insights in the context of the housing market and social stratification. This relationship suggests that there is a **tendency**, although not a strong one, for **individuals with higher levels of education to own their home for longer.**

It should be remembered that **DEB12A** does **not correspond to the total value of the debt requested from the financial institution**, but to the amount of the debt **remaining** to be settled with the bank **in the year of the interview**. The study of the correlation between these variables may therefore be biased.

If we have another look at the correlation coefficients, then we can conclude that most of the variables are independent of each other and can therefore be used in our model for our conclusions.

FASCIAREDDITO and **Y**, on the other hand, are, as you might expect, **highly correlated**. It was easy to conclude that FASCIAREDDITO was an artificial variable that I created to make my life easier during the analysis.

To put it simply, we can only choose one of the two, and the **choice fell on Y**(Income), which, being a purely numerical data, could give us more exhaustive answers further down the line.

3.4 **INTEREST RATES: COLLECTION OF DATA**

Interest rates have been included in this section (15) because they will be very useful in the next step we will make towards completing the restructuring of our data.

They will be very useful because they will be used to construct hypothetical scenarios (we will see in the following pages) that will help us categorize household debt in the two areas of our interest: **fixed** (FRM) or **adjustable mortgages** (ARM).

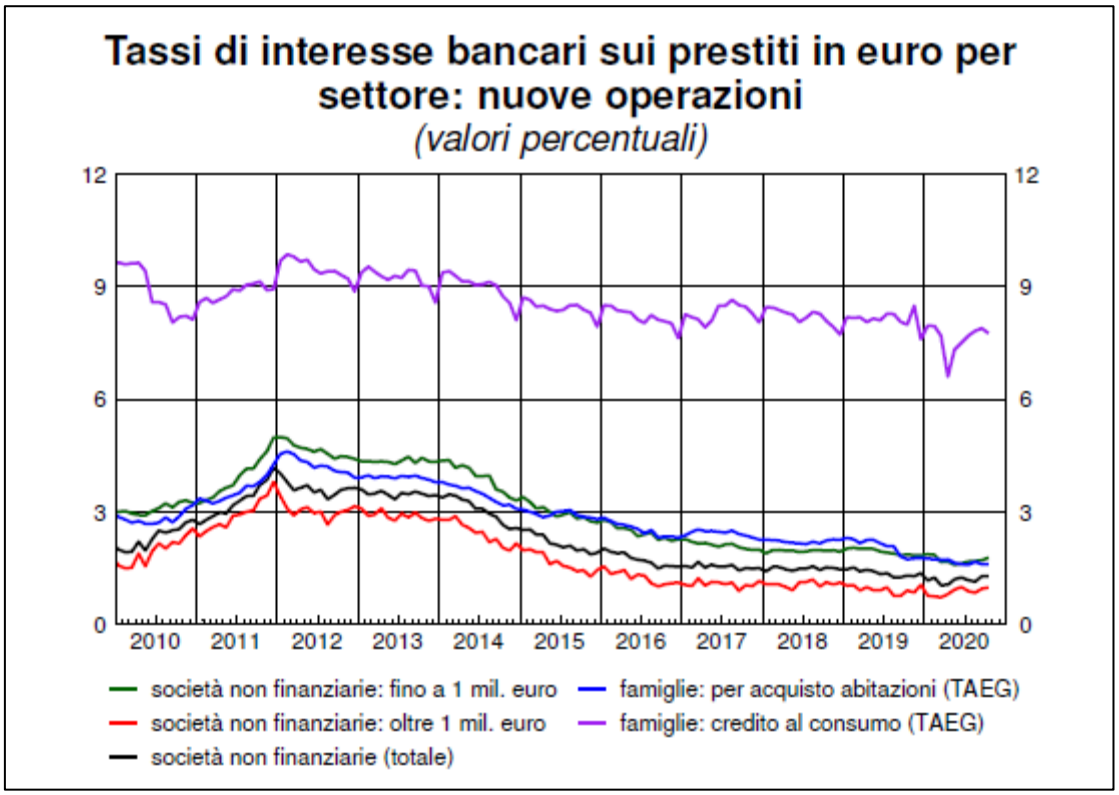


Figure 6 - Banks Interest rates by type

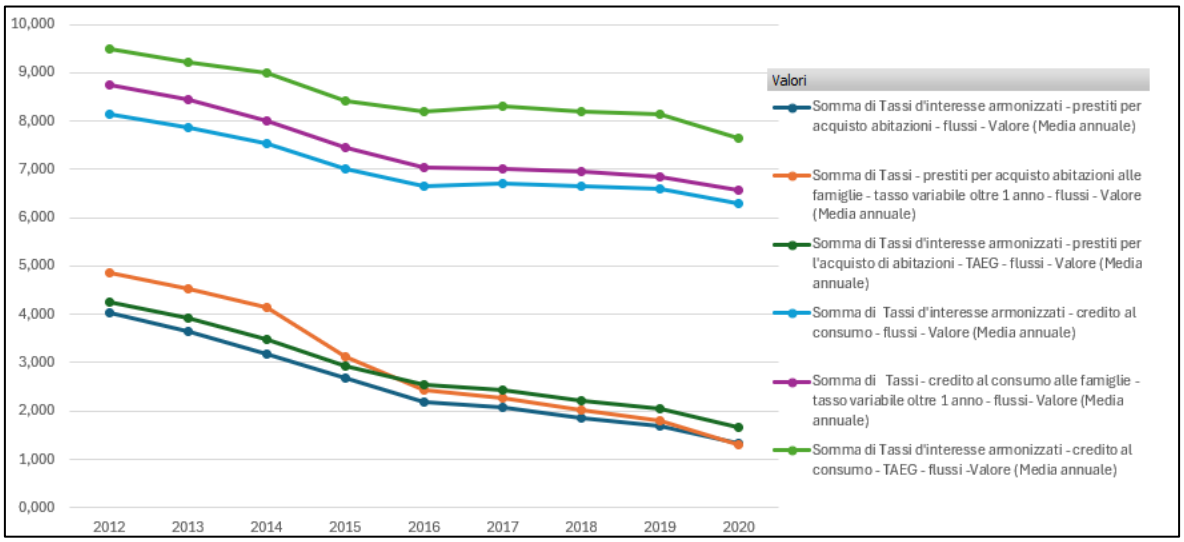


Figure 7 - Yearly avg interest rates, by type

ANNO	Somma di Tassi d'interesse armonizzati - prestiti per acquisto abitazioni - flussi - Valore (Media annuale)	Somma di Tassi - prestiti per acquisto abitazioni alle famiglie - tasso variabile oltre 1 anno - flussi - Valore (Media annuale)	Somma di Tassi d'interesse armonizzati - prestiti per l'acquisto di abitazioni - TAEG - flussi - Valore (Media annuale)	Somma di Tassi d'interesse armonizzati - credito al consumo - flussi - Valore (Media annuale)	Somma di Tassi - credito al consumo alle famiglie - tasso variabile oltre 1 anno - flussi- Valore (Media annuale)	Somma di Tassi d'interesse armonizzati - credito al consumo - TAEG - flussi - Valore (Media annuale)
2012	4,020	4,863	4,262	8,135	8,737	9,479
2013	3,655	4,536	3,917	7,866	8,431	9,229
2014	3,186	4,146	3,470	7,532	8,014	9,002
2015	2,687	3,133	2,943	7,002	7,440	8,424
2016	2,196	2,426	2,533	6,661	7,034	8,187
2017	2,064	2,269	2,431	6,708	7,014	8,293
2018	1,858	2,023	2,216	6,660	6,942	8,199
2019	1,692	1,809	2,059	6,587	6,841	8,128
2020	1,321	1,314	1,663	6,290	6,576	7,658

Table 4 - Yearly average Bank interest rates, by type

Please keep in mind that the rates presented here (15) are primarily **harmonized interest rates**, taking into account the various durations of the rate and the various banks involved in the survey. The harmonized data in the database published by the **Bank of Italy** (15) are on a **monthly basis**. For clarity, an **arithmetic average** has been used to represent the rates on an **annual basis**.

From the Bank of Italy survey, we have extrapolated 2 main different types of interest rates, crucial for our study:

- Annual Effective Annual Rate (**TAEG**) for home purchases;
- **Variable** TAEG on home purchases;

3.5 CATEGORISATION OF THE MORTGAGE BY " RATA_AR "

The procedure summarized in the title requires new variables to be created. Please note that the variables presented below will not be used later. They are needed on the spot to identify the type of debt contracted by the consumer. **Therefore, the table of variables at the beginning of the chapter does not include them.**

Let us start by defining all the variables used:

VARIABLE	DESCRIPTION	CODING
ANNO	*Survey reference period (year).*	
NQUEST	*Questionnaire number – household ID*	
RATA_AR	*Age class*	
80%VALIMM	80% *of Value of the property.*	0.80*VALIMM
DURATAPOSS	*#years owning the property*	
20ANNIFIXEDMORT	*Theor. yearly installment for a 2(– rate mortgage.*	-RATA(i;20; 0.8*VALIMM)
20YVARMORTG	*Theor. yearly installment for a 2(– rate mortgage.*	-RATA(i;20; 0.8*VALIMM)
25YFIXEDMORT	*Theor. yearly installment for a 2[– rate mortgage.*	-RATA(i;20; 0.8*VALIMM)
25YVARMORT	*Theor. yearly installment for a 2[– rate mortgage*	-RATA(i;20; 0.8*VALIMM)
30YFIXEDMORT	*Theor. yearly installment for a 3(– rate mortgage.*	-RATA(i;20; 0.8*VALIMM)

30YVAREDMORT	*Theor. yearly installment for a 30 – rate mortgage*	-RATA(i;20; 0.8*VALIMM)
CONVALIDA	(*see below*)	

Table 5 - Fictitious variables used in computations

The **CONVALIDA variable** contains a **nested logic function** which returns a text string with the name of one of the suggested loan types based on the presence or absence of RATA_AR in one of the suggested ranges. If it is absent from all the proposed ranges, the logic function returns the string "VERIFICARE".

```
SE(E(RATA_AR<=-5% 20YFIXED;RATA_AR>=+5% 20YFIXED);20YFIXMORTGAGE;SE(E(RATA_AR<=-5% 20YVAR;RATA_AR>=+5% 20YVAR);20YVARMORTGAGE;

SE(E(RATA_AR<=-5% 25YFIXED;RATA_AR>=+5% 25YFIXED);25YFIXMORTGAGE;SE(E(RATA_AR<=-5% 25YVAR;RATA_AR>=+5% 25YVAR);25YVARMORTGAGE;

SE(E(RATA_AR<=-5% 30YFIXED;RATA_AR>=+5% 30YFIXED);30YFIXMORTGAGE;SE(E(RATA_AR<=-5% 30YVAR;RATA_AR>=+5% 30VAR);30YVARMORTGAGE;"VERIFICARE"))))))
```

Figure 8 - CONVALIDA nested function

The temporary variables obtained are the **hypothesized repayments** for each household, based on **80% of the property value of the home**, using the correct interest rate.

We have calculated the instalments for **20-year**, **25-year** and a **30-year** plan mortgage, both in terms of **fixed** and **adjustable mortgage**. It is important that each theoretical instalment calculated is assigned the correct average interest rate for the reference year.

For the **calculations relating to fixed-rate instalments**, the **average fixed rate** of the year in which the mortgage was taken out (not the reference year, mind you!) was used in the formula for each family interviewed. e.g. for a family interviewed in 2016 with a property purchased in 2012, the average APR of 2012 was used in the calculation, and so on.

On the other hand, for the **calculation of variable rate instalments**, the cases are less complex: the average variable rate associated with the year of the interview was applied to each household, regardless of whether the years matched or not.

The **results** obtained are much more encouraging than the first test carried out.

Putting the results and numbers together using a pivot, we can see that the number of mortgages identified is growing exponentially:

TYPE	Conteggio di NQUEST
FIXEDMORTG	58
VARMORTGAGE	86
Totale complessivo	144

Table 6 - Division on Mortgage Type

We obtained no less than 144 compatibilities out of 277. It is worth noting that, for ease of reading, we have grouped all scenarios of the same type into the two macro variables '**FIXEDMORTG**' and '**VARMORTGAGE**'.

A surprising result, however, is that the initial hypothesis, i.e. that we expected most households to have taken out a fixed-rate mortgage, was refuted.

3.6 **REGRESSION MODELS**

In order to understand the study carried out, it is necessary to define the **multivariate analysis model** used. Given the research question, the first step is to understand the dependent variable: it is a binary variable that can only take the values 0 or 1. The dichotomous nature of the dependent variable limits the scope of the analysis to the use of three models: the **linear probability model** and the **probit model**.

LINEAR PROBABILITY MODEL

LPM is a **binary dependent variable regression model (16)**, where the relationship between the dependent and independent variables is written:

$$Y = f(X1, X2, X3, \dots\dots, Xn) + \varepsilon = f(\boldsymbol{X}) + \varepsilon = \boldsymbol{\beta X} + \varepsilon$$

By making the linear function explicit:

$$Yi = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad i = 1 \dots . n$$

- Y is the dependent variable;
- X_1 , X_2 , … . , X_n are the independent (or explanatory, regressor) variables;
- β_0 is the intercept of the unknown population;
- β_k is the effect on Y of a variable X_k , holding the other variables constant;
- u_i is the regression error.

The **expected value** of the variable Y is:

$$E(Y) = X\beta$$

and given the definition of the dependent variable, the expected value that Y=1 is P(Y=1):

$$E(Y) = 1 * P(Y = 1) + 0 * P(Y = 0) = P(Y = 1) = P(Y = 1) = X\beta$$

The **regression coefficients** are calculated by minimising the sum of ordinary squares (OLS method), i.e. minimising the mean square difference between the current values of Yi and the predicted value based on the estimated line.

The OLS estimator (in the case of two regressors) is given by:

$$min_{b_0 b_1 b_2} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i})]^2$$

Using the linear model facilitates the estimation and interpretation of the β-coefficients of the regression, which represent the marginal effects. Furthermore, the coefficients and predictions seem to be sufficiently reliable. This may result in prediction probabilities that are lower than 0 or higher than 1. However, this model is less than optimal when there is a dichotomous dependent variable, which may result in prediction probabilities that are lower than 0 or higher than 1.

That's why we'll now introduce the concept of Probit and Logit Models, with the main focus on the first one, used to extract the results of our analysis.

PROBIT MODELS

Probit model (17) is non-linear regression model that is particularly useful for **binary dependent variables**. It allows you to estimate the probability of having one value or another of the dependent variable as a non-linear function of the independent variables.

$$P(Y = 1) = G(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k) = G(X\beta)$$

where G is a non-linear function allowing βX to take values between 0 and 1. The range of this function is C=[0,1], which is typical of distribution functions, and therefore the first derivative is not constant, which means it depends on the regressors. The **dependence of the regressors** has the effect of varying the slope of the curve in such a way that it remains within the defined code range, in accordance with the limitation of the dependent variable.

$$P(Y = 1) = \Phi(X\beta) = \int_{-\infty}^{+\infty} \Phi(z)dz$$

The standard normal distribution (cumulative density function) is used as a non-linear function in the probit model.

We now define the **likelihood function** (17) as:

$$P(Y_i = 1)^{y_i} * P(Y = 0)^{1-y_i}$$

taking its logarithm and summing all observations i to obtain the log likelihood function:

$$\sum_{i=1}^{n}\{[Y_i * log(P(Y_i = 1))] + [(1 - Y_i) * \log(P(Y_i = 0))]\}$$

and replacing $P(Y = 1)$ by $G(\beta X)$, the regression coefficients can be obtained by log-likelihood function maximization.

The **maximum similarity estimator** is:

$$max \sum_{i=1}^{n}\left[Y_i * \log\left(P(Y = 1)\right) + (1 - Y_i) * \log(P(Y = 0))\right]$$

This means that the effect of an independent variable on the dependent variable does not depend on the size of the dependent variable, but rather on the size of the effect of the independent variable on the dependent variable. When analysing, it was sufficient to enter the marginal command dx(*) in the statistical program used.

Probit models may have problems with the **phenomenon of Heteroscedasticity** (18) : if the variance of the error distribution conditional on the independent variable does not vary with the independent variable, the error u will be homoscedastic. To avoid underestimating the errors or drawing erroneous conclusions, the standard errors need to be robust to heteroskedasticity (for the reference tools, you need to enter robust into the regression command).

In conclusion, the choice of the model was mainly based on "disciplinary preference": the **probit was better suited to econometric analysis**.: these models are more effective for models with random effects, where the effects differ from one individual to another and for **medium-sized samples**, as is the case **here** (**19**).

4. RESULTS OF THE EMPIRICAL RESEARCH

The results of the analysis were obtained using the 'probit' function in the STATA software. The dependent variable considered in the analysis, the variable **TYPE**, is defined as a dummy variable with two possible values (=**0 if FIXEDRATEMORTGAGE** and =**1 if ADJUSTEDRATEMORTGAGE**).

In the following pages, we will pay attention to different focuses, starting with the generic probit regression with all variables in play, and then intersecting different variables with each other.

It is important to note that normally one should comment on the **statistically significant variables,** while mentioning the remaining non-significant ones. As we do not have enough variables of this type (for reasons **explained later in the comments**), we will also give a more detailed overview of those that are **not statistically significant**, but which may nevertheless provide us with some interesting insights.

4.1 EMPIRICAL RESULTS: PROBIT REGRESSION WITH ALL THE CHOSEN VARIABLES

```
14 . probit TYPE ANNO RATA_AR Y DEBI12A VALABIT ETA5 STUDIO AREA3 QUALP7N

   Iteration 0:  Log likelihood = -97.073553
   Iteration 1:  Log likelihood = -82.078933
   Iteration 2:  Log likelihood = -81.768249
   Iteration 3:  Log likelihood = -81.767243
   Iteration 4:  Log likelihood = -81.767243

   Probit regression                                 Number of obs =    144
                                                     LR chi2(9)    =  30.61
                                                     Prob > chi2   = 0.0003
   Log likelihood = -81.767243                       Pseudo R2     = 0.1577

------------------------------------------------------------------------------
        TYPE | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ANNO |    .163118   .0516037     3.16   0.002     .0619766    .2642595
    RATA_AR |   -.0001113   .0000546    -2.04   0.042    -.0002183   -4.22e-06
           Y |   -3.08e-07   6.49e-06    -0.05   0.962     -.000013    .0000124
    DEBI12A |    2.43e-06   2.56e-06     0.95   0.343    -2.59e-06    7.46e-06
    VALABIT |    4.47e-06   1.91e-06     2.34   0.019     7.27e-07    8.22e-06
       ETA5 |   -.0974488   .1287203    -0.76   0.449    -.3497361    .1548384
    STUDIO  |    .155288    .148531     1.05   0.296    -.1358274    .4464034
      AREA3 |    .2284423   .1421388     1.61   0.108    -.0501447    .5070292
    QUALP7N |   -.0220792   .0618639    -0.36   0.721    -.1433302    .0991719
      _cons |   -330.1686    104.081    -3.17   0.002    -534.1637   -126.1735
------------------------------------------------------------------------------
```

Table 7 - General Probit Regression

```
Marginal effects after probit
      y  = Pr(TYPE) (predict)
         =  .61731427
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]       X
---------+--------------------------------------------------------------------
    ANNO |    .0622404      .01979     3.14   0.002    .023443   .101038    2018.5
RATA_AR  |   -.0000425      .00002    -2.04   0.041   -.000083  -1.6e-06    8424.31
       Y |   -1.18e-07      .00000    -0.05   0.962   -5.0e-06   4.7e-06    47735.7
DEBI12A  |    9.28e-07      .00000     0.95   0.344   -9.9e-07   2.8e-06     132329
VALABIT  |    1.71e-06      .00000     2.36   0.018    2.9e-07   3.1e-06     243472
   ETA5  |   -.0371832      .04907    -0.76   0.449   -.133355   .058989     2.6875
STUDIO   |    .0592527       .0568     1.04   0.297   -.052072   .170577    4.49306
  AREA3  |    .0871659      .05425     1.61   0.108   -.019161   .193493    1.71528
QUALP7N  |   -.0084247       .0236    -0.36   0.721    -.05468    .03783    3.32639
------------------------------------------------------------------------------
```

Table 8 - Marginal Effects on General Probit Regression

The model's **log likelihood** of **-81.767243** and a statistically significant **LR chi-squared** statistic of **30.61** (p-value = 0.0003) suggest that the predictors collectively make a meaningful contribution to our understanding of the factors influencing mortgage type choice, with a **pseudo R-squared** of **0.1577** indicating a modest fit.

After reviewing briefly the generic results in terms of accuracy of the probit regression, let's now focus on each variable, keeping an eye, respectively on the coefficients and on their respective signs, on their statistical significance and on their marginal effects.

Normally, we should be able to describe mainly the effect of the statistically significant variables, with an indication of the non-statistically significant variables. Since, as we can see, there are only a few significant variables, we will also focus on the others, which may still give some indication of possible correlations between our dependent and independent variables:

- **Year (ANNO)**: The coefficient (0.163118) is positive and **statistically significant** (p = 0.002), indicating that the propensity to choose an adjustable-rate mortgage has increased significantly over time. The **marginal effect** (0.0622404) suggests that with each passing year, the probability of choosing an adjustable rate mortgage increases by about 6.22%, all other things being equal;

- **Annual instalment (RATA_AR)**: A negative coefficient (-0.0001113) with **statistical significance** (p = 0.042) suggests that higher annual instalments are associated with a lower probability of choosing an adjustable-rate mortgage, possibly due to the increased financial burden they represent. The **marginal effect** (-0.0000425) confirms this negative relationship, although the effect size is small.

- **House value** (**VALABIT**): The **positive coefficient** (4.47e-06) is **statistically significant** (p = 0.019), suggesting that households with higher-valued properties are more likely to opt for an interest-only mortgage, possibly due to their better financial position. The corresponding **marginal effect** (1.71e-06) supports this finding.

- **Income (Y)**: Despite a negative coefficient, it is not statistical significant (p = 0.962), which means that income doesn't have a clear influence on the choice of mortgage type within this data set. The marginal effect is also small and not statistically significant;

- **Residual mortgage debt** (**DEBI12A**): The coefficient here is **positive** (2.43e-06) but **not statistically significant** (p = 0.343), leaving the relationship between remaining debt and mortgage type preference unclear. The **marginal effect** (9.28e-07) indicates a small positive change in the likelihood of choosing an adjustable-rate mortgage with increased debt, but again without statistical significance;

- **ETA5** (age group): The **coefficient** on the age group variable, ETA5, indicates a **negative direction** (-0.0974488), although it is **not statistically significant** (p = 0.449). This suggests a tendency for the probability of choosing an adjustable rate mortgage to decrease with age, but given the lack of significance this cannot be confirmed as a clear pattern within the dataset. The **marginal effect** of -0.0371832 reinforces this trend, suggesting that the likelihood of choosing an adjustable rate mortgage decreases with age, but confidence in this effect is limited;

- **STUDIO** (level of education): The **positive coefficient** of 0.155288 (p = 0.296) for STUDIO suggests a relationship where higher levels of education may be associated with a greater likelihood of choosing an adjustable rate mortgage. However, the **statistical insignificance** suggests that educational attainment alone does not significantly influence the choice of mortgage type. The **marginal effect** of 0.0592527 adds a dimension to this finding, indicating a possible but uncertain increase in the probability of selecting an adjustable-rate mortgage with higher educational attainment;

- **AREA3** (geographical area): The **coefficient** for AREA3 is 0.2284423 and, although **not statistically significant** at the 10% level (p = 0.108), it suggests a regional influence on mortgage type preference. Specifically, the positive coefficient suggests that households in the 'Centre' or 'South and Islands' regions may have a higher propensity to take out an adjustable-rate mortgage compared to the 'North'. This could reflect regional economic differences or cultural

differences in financial behavior. The **marginal effect** of 0.0871659, although not conclusive, suggests that regional factors may play a role in mortgage choice, with households in these regions almost 9% more likely to choose an adjustable rate mortgage;

- **QUALP7N** (employment situation): The employment position captured by QUALP7N has a **negative coefficient** of -0.0220792, but with a p-value of 0.721 the relationship is **not statistically significant**. This suggests that there isn't a uniform effect on the type of mortgage chosen across the different occupational categories, from manual workers to managers. The **marginal effect** of -0.0084247 also suggests a marginal decrease in the likelihood of choosing an adjustable-rate mortgage associated with these employment categories, but given the statistical insignificance this effect is weak and does not provide a strong basis for inference.

18 . lstat

Probit model for TYPE

Classified	True D	True ~D	Total
+	**74**	**27**	**101**
−	**12**	**31**	**43**
Total	**86**	**58**	**144**

Classified + if predicted Pr(D) >= .5
True D defined as TYPE != 0

Sensitivity	Pr(+| D)	**86.05%**
Specificity	Pr(−|~D)	**53.45%**
Positive predictive value	Pr(D| +)	**73.27%**
Negative predictive value	Pr(~D| −)	**72.09%**
False + rate for true ~D	Pr(+|~D)	**46.55%**
False − rate for true D	Pr(−| D)	**13.95%**
False + rate for classified +	Pr(~D| +)	**26.73%**
False − rate for classified −	Pr(D| −)	**27.91%**
Correctly classified		**72.92%**

Table 9 - Good of Fit General Probit Regression

Talking about **good of fits** of the regression built, we can say that the classification performance of the model is quite good, with a **sensitivity of 86.05%,** meaning that it correctly identifies a high percentage of those who would choose an adjustable rate mortgage. The **specificity is moderate at 53.45%**, suggesting that the model is less adept at correctly identifying those who would not choose an adjustable rate mortgage. These figures suggest that the model is better at predicting the acceptance of adjustable rate mortgages than their rejection. The balance between positive predictive value (73.27%) and negative predictive value (72.09%) suggests a model that **performs well** across different scenarios. The overall accuracy rate of 72.92% is substantial, demonstrating the model's robust predictive ability within the given dataset.

Although the focus of our work is on the probit just described, we can try to focus on certain variables to see if they can give us some other indication, in the case of a probit with narrower variables.

Can we find statistically significant variables that can be associated with purely social and family background aspects, or with real estate factors? We will find out in the next focus on categorical and property variables.

4.2 EMPIRICAL RESULTS: FOCUS ON CATEGORICAL VARIABLES (ETA5, AREA3, ANNO)

This iteration of the probit regression focuses on the categoric variables, such as the influence of time, represented by the year (ANNO), and demographic factors, such as age (ETAS) and geographic area (AREA3), on the likelihood of having an adjustable rate mortgage (ARM) versus a fixed rate mortgage.

Probit regression — Number of obs = 144
LR chi2(9) = 24.06
Prob > chi2 = 0.0042
Log likelihood = -85.045786 — Pseudo R2 = 0.1239

TYPE	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
ANNO						
2014	-.4511678	.808984	-0.56	0.577	-2.036747	1.134412
2016	.9407886	.7582205	1.24	0.215	-.5452962	2.426873
2020	1.152382	.7207368	1.60	0.110	-.2602357	2.565
ETA5						
2 = 31 to 40 years;	-.8172341	.4456413	-1.83	0.067	-1.690675	.0562069
3 = 41 to 50 years	-.6432941	.4617187	-1.39	0.164	-1.548246	.261658
51 to 65 years	-.915652	.4773764	-1.92	0.055	-1.851293	.0199885
over 65 years	-1.03347	.7661045	-1.35	0.177	-2.535007	.4680676
AREA3						
Centre	.3100518	.2824257	1.10	0.272	-.2434924	.863596
South and Islands	.413413	.2864164	1.44	0.149	-.1479529	.9747789
_cons	-.1189126	.8408118	-0.14	0.888	-1.766873	1.529048

. mfx
no independent variables
r(102);

Table 10 - Probit Regression on Categoric Variables

The **log likelihood** is -85.045786, indicating a *good fit* of the model to the data. With a lower LR chi-squared statistic of 24.06 compared to the previous model, and a corresponding p value of 0.0042, the *model is still statistically significant*.

What we notice immediately is that all these three independent variables, combined with our dummy dependent variable, have not particularly statistical significance. Let's now highlights the key findings:

- **ETA5**(age groups): The **negative coefficients** for the age groups, with '31 to 40' being the most significant (p = 0.067), suggest a trend where younger age groups are less likely to choose an ARM. Again here, the numbers are there, but we take this conclusion with the relative caution;
- **ANNO** (year): The **coefficients** for the years 2014, 2016 and 2020 are **not statistically significant** at conventional levels, suggesting that these points in time do not have a clear impact on the choice between an ARM and a fixed-rate mortgage within the scope of these data. The trend appears to be increasing over time, but we don't take this conclusion as true in general;
- **AREA3** (geographical area): The **positive coefficients** for 'Centre' and 'South and Islands' indicate a higher probability of choosing an ARM in these regions compared to the North, with 'South and Islands' being more pronounced.

Since there are not particular statistical significance in the variables taken into account, STATA **doesn't let us find the marginal effects** on the probit regression analyzed above.

4.3 EMPIRICAL RESULTS: FOCUS ON ECONOMIC VARIABLES (DEBI12A, RATA_AR, Y)

```
23 . probit TYPE RATA_AR Y DEBI12A

   Iteration 0:  Log likelihood = -97.073553
   Iteration 1:  Log likelihood = -91.078091
   Iteration 2:  Log likelihood = -91.059071
   Iteration 3:  Log likelihood = -91.059069

   Probit regression                                 Number of obs =    144
                                                     LR chi2(3)    =  12.03
                                                     Prob > chi2   = 0.0073
   Log likelihood = -91.059069                       Pseudo R2     = 0.0620

------------------------------------------------------------------------------
        TYPE | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     RATA_AR |  -.0000154   .0000371    -0.42   0.678    -.0000881    .0000573
           Y |   6.68e-06   5.13e-06     1.30   0.193    -3.37e-06    .0000167
     DEBI12A |   4.72e-06   2.10e-06     2.24   0.025     5.97e-07    8.84e-06
       _cons |  -.5426267   .2788687    -1.95   0.052    -1.089199    .0039458
------------------------------------------------------------------------------

24 . MFX
   command MFX not defined by MFX.ado
   r(199);

25 . mfx

   Marginal effects after probit
         y  = Pr(TYPE) (predict)
            =  .60692206
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]       X
---------+--------------------------------------------------------------------
 RATA_AR |  -5.92e-06      .00001   -0.42   0.678  -.000034  .000022   8424.31
       Y |   2.57e-06      .00000    1.31   0.192  -1.3e-06  6.4e-06   47735.7
 DEBI12A |   1.82e-06      .00000    2.25   0.024   2.3e-07  3.4e-06    132329
------------------------------------------------------------------------------
```

Table 11 - Probit Regression on Economic Variables

The model has a **log likelihood** of -91.059069 and an **LR chi-squared value** of 12.03, which is statistically significant (p = 0.0073), albeit indicating a moderate relationship between the predictors and mortgage type. The **pseudo R-squared** value of 0.0620 suggests that while the predictors included in the model have an effect, there may be other unobserved factors not taken into account that also influence the families to choose an ARM.

- **DEBI12A** (remaining mortgage debt): This is the only predictor with a **statistically significant coefficient** (4.72e-06) *at the 5% level*. It indicates that as the remaining mortgage debt increases, so does the probability of the mortgage being an ARM, possibly giving us the chance to propose a strategy in the borrower's debt management and payment, opting for the initially lower payments of an ARM.
- **RATA_AR** (annual instalment): The **coefficient** is negative, but **not statistically significant** (p = 0.678), suggesting that the annual instalment amount may not be a strong predictor of mortgage type, but still gives the same results in terms of logic from what we've seen so far.
- **Y** (income): The positive coefficient assigned is **not statistically significant**, but the **positive coefficient** gives us the idea that Higher Income Families could have higher risk attitude and choose an ARM over the FRM.

4.4 EMPIRICAL RESULTS: FOCUS ON HOUSING PROPERTY CHARACTERISTICS

The current model analyzes the effects of property value (VALABIT), annual income (Y) and remaining mortgage debt (DEBI12A) on the type of mortgage chosen.

```
27 . probit TYPE VALABIT ANNPOSS Y DEBI12A

   Iteration 0:  Log likelihood = -97.073553
   Iteration 1:  Log likelihood = -89.647428
   Iteration 2:  Log likelihood = -89.503303
   Iteration 3:  Log likelihood = -89.503091
   Iteration 4:  Log likelihood = -89.503091

   Probit regression                                 Number of obs =    144
                                                     LR chi2(4)    =  15.14
                                                     Prob > chi2   = 0.0044
   Log likelihood = -89.503091                       Pseudo R2     = 0.0780

------------------------------------------------------------------------------
        TYPE | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     VALABIT |   2.33e-06   1.57e-06     1.48   0.139    -7.56e-07    5.41e-06
     ANNPOSS |   .0388515   .0471971     0.82   0.410     -.053653    .1313561
           Y |   6.75e-07   5.48e-06     0.12   0.902    -.0000101    .0000114
     DEBI12A |   2.52e-06   2.20e-06     1.14   0.253    -1.80e-06    6.83e-06
       _cons |  -78.94758   95.06894    -0.83   0.406    -265.2793    107.3841
------------------------------------------------------------------------------

28 . mfx

   Marginal effects after probit
         y  = Pr(TYPE) (predict)
            =  .61272011
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 VALABIT |   8.90e-07      .00000     1.49   0.137  -2.8e-07  2.1e-06    243472
 ANNPOSS |   .0148766      .01808     0.82   0.411  -.020554  .050307    2015.44
       Y |   2.59e-07      .00000     0.12   0.902  -3.9e-06  4.4e-06    47735.7
 DEBI12A |   9.63e-07      .00000     1.14   0.254  -6.9e-07  2.6e-06    132329
------------------------------------------------------------------------------
```

Table 12 - Probit Regression Model with Housing Property Variables

The **log likelihood** of -89.503091 indicates that the model fits the data, and the **LR chi-squared statistic** of 15.14, with a p-value of 0.0044, confirms the significance of the overall model. The variables taken into account gives different points of view to watch.

- **VALABIT** (property value): The **coefficient** is *positive* (2.33e-06) but **not statistically significant** (p = 0.139), suggesting a tentative positive association between property value and the choice of an ARM. The higher the value, the greater the likelihood of choosing an ARM, possibly due to the greater financing flexibility of owners of higher valued properties;

- **ANNUAL INCOME** (Y): The income variable has a **positive coefficient** (6.75e-07), but it is **not statistically significant** (p = 0.902). A possible conclusion to that, with all the cautions needed, is that higher incomes leads to higher risk tolerance for families;

- **DEBI12A** (residual mortgage debt): The **coefficient** here is also *positive* (2.52e-06) but **lacks statistical significance** (p = 0.253), suggesting that this suggests that an increase in mortgage outstanding slightly increases the probability of choosing an ARM;

- **ANNOPOSS** (years in which you owned the property): The **coefficient** is *positive*, but **not statistically significant**. This implies that, going on the years, families could have found more suitables for their mortgages to choose ARM over FRM.

4.5 COMMENTS ABOUT THE EMPYRICAL RESULTS: PROBIT REGRESSIONS ISSUES

In the probit regression described above, most of the variables represented were not statistically significant. Given the experimental nature of this analysis, it would be interesting to examine possible causes that could have influenced the overall consistency of the results we have.

Therefore, we will try to formulate a hypothesis as to what could be the consistent causes of this result.

First, we can certainly address the issue of the **sample size** we used: With only 144 observations, the sample size may be too small to detect all of the effects except for the strongest ones. In general, larger sample sizes provide more power to detect statistically significant relationships. It should be noted that we struggled to find consistent information regarding the choice of mortgage (see Annex B).

Secondly, another issue can be represented by the **selection of variables**: the variables included may not be those with the strongest effect on the dependent variable. The effects that are found to be significant may be altered by including more relevant variables or excluding less relevant ones. We may have excluded some important information from the SHIW because we thought that it could be left out,

The problem can be intrinsic into the **model specifications**: the Probit model assumes a specific relationship between the independent variables and the probability of the outcome. If this relationship is incorrectly specified, it may lead to insignificant results for variables that do have an effect, or to significant results for variables that do not have an effect.

The possible causes just presented are hypotheses based on possible errors resulting only from my analysis. We should not forget, however, that we could have at least two other possible exogenous causes, such as random variation and the quality of the data.

In **random variation**, some of the variation in the dependent variable may simply be random and have no systematic relationship with any of the independent variables under measurement.

Equally important to the outcome of the analysis may have been the **quality of the data**: measurement errors in the data can weaken associations and make true impacts harder to detect. This includes errors with how data are reported, recorded or coded. Given the way the Bank of Italy collects information, we cannot rule out that the people involved have made mistakes.

4.6 COMMENTS ABOUT THE EMPYRICAL RESULTS: COMPARISON WITH THE LITERATURE GUIDELINES

The analysis showed that *traditional economic variables*, such as annual income and property value, *are influentia*l. However, they are not the sole determinants of mortgage preferences. This suggests a more nuanced decision landscape, with *social*, *economic* and *cultural factors* playing an **important role**.

It is important to note, however, that the *studies carried out on this subject are extremely few in numbers* and are also contextualized within periods characterized by massive economic crises in recent years (just think of the Internet bubble in the US) and the crisis at the end of the first decade of the 2000s. They are, however, of the **utmost importance** because they *represent models* to be referred to in order to *make progress in research*. Let us start with the results of our work that can be assimilated with the literature.

First of all, it is certainly significant that the literature tends to suggest that in *periods of relative economic stability*, or even when *interest rates tend to fall slightly*, households prefer to take out an **adjustable-rate mortgage**. The motivation may lie in the fact that interest rates are expected to remain at favourable levels or to fall further, and that one can therefore take advantage of this dynamic to obtain further favourable debt payments. This dynamic is *explained* by the **positive coefficient** of the statistically significant variable **ANNO**: from 2012 to 2019, as the Italian domestic economy stabilises, we have a positive propensity in the overall probit.

Another point of contact is the concept of "*financial Education*" introduced at the beginning of the research. In the literature there is a **positive relationship** between progress in *studies* and the choice of an **adjustable rate mortgage**. The same dynamic is found in the empirical analysis: the **STUDIO** variable, although not statistically significant, increases the propensity to choose an ARM rather than an FRM as it increases.

The literature (13) mentions that households with a relatively *high mortgage payment and debt* to repay are *more likely* to choose a **fixed-rate mortgage** than an

adjustable-rate mortgage. This information is *confirmed* by the fact that the *statistically significant* variable used in the survey, **RATA_AR** (the instalment paid in the year of the survey), has a *negative marginal coefficient*: as the numbers associated with it increase, the probability of choosing an interest-only mortgage tends to decrease.

We do not find any specific situations concerning the tendencies of households in their choice of mortgage as the value of the *house acquired varies*: in our analysis, the marginal coefficient of the variable **VALABIT** (var. stat. significant) is positive in the general Probit. This result should lead us to conclude that a positive change in the value of the dwelling should correspond to a higher propensity to take out an adjustable rate mortgage. *This inference is somewhat at odds* with the explanation provided by *Cocco and Campbell* (11), who instead tend to associate a rapid re-mortgaging with the choice of an ARM mortgage (note: not necessarily the property value associated with the willingness not to change dwelling!)

We do not have any data in the literature related to *purely financial issues* (risk aversion, reaction to unexpected fluctuations and hypothetical changes in income) to compare with the empirical results obtained.

In general, therefore, although the empirical analysis shows some areas for improvement, such as the general quality of the data available and the consequent absence of certain variables that we had assumed to be statistically significant, most of the work carried out *can be associated* with what has been found in the papers on the problem of the "**optimal choice of mortgage**".

5 IMPROVEMENTS: AN OVERVIEW

The research conducted so far has highlighted various aspects that influence mortgage choice, but there is potential to deepen and broaden this research for a more nuanced understanding. The integrity of our conclusions is inextricably linked to the quality and comprehensiveness of our data. Future efforts should focus on improving data collection methods to ensure accuracy and avoid misinterpretations that may mask the true dynamics at play. In addition, expanding the range of variables considered - beyond typical financial metrics to include fluctuating factors - could provide a more complete understanding of the drivers of mortgage decisions.

In addition, broadening the pool of participants in our studies will facilitate the identification of significant patterns and trends, overcoming the current limitations of limited data. Examining how mortgage preferences change over time or vary across different demographic groups can provide deeper insights, requiring longitudinal studies and comparisons across different groups. The use of advanced technologies, such as big data analytics, could uncover overlooked patterns and simplify the handling of large datasets for more comprehensive analysis.

It's also appropriate to explore aspects of behavioural finance, looking at how psychological traits and biases influence mortgage decisions. This research should include risk aversion and other key behavioural factors. Expanding research to assess the economic and family utility of housing finance, especially against the background of recent global challenges affecting economic stability, may also provide insight into the underlying causes of mortgage delinquency and renegotiation. This broadened perspective could be used to inform policy and regulatory decisions, thereby enhancing the welfare of consumers and the resilience of the financial system.

6 CONCLUSION

In my dissertation, I explored the factors that shape *Italian families' mortgage choices*, focusing on a stable *economic period from 2012 to 2019*. Using detailed data from the Bank of Italy, I examined how social, economic and cultural elements interact in the context of mortgage choice, offering new perspectives on the expanding field of household financial behaviour.

The empirical analysis examines the factors that influence mortgage preferences. It shows that while traditional *economic indicators* such as annual income and property value play a role, *they aren't the only factors at play*. A mix of social, economic and cultural elements also play a significant role in the decision-making process.

The findings are consistent with the existing literature in several respects. For example, during periods of economic stability or falling interest rates, households tend to favour adjustable-rate mortgages (ARMs), anticipating the benefit of potentially lower future payments. This preference is related to the notion of "financial literacy", with higher levels of education correlating with the choice of an ARM, although this was not statistically significant in the study.

In addition, the analysis supports the idea that households with high mortgage payments are more likely to opt for fixed-rate mortgages (FRMs) to avoid the risk of rising interest rates. Contrary to some literature, the study found a positive association between house value and the likelihood of choosing an ARM, suggesting that as house values rise, so does the propensity to choose an ARM, contrary to some existing theories linking rapid remortgaging or a desire not to move with ARM choice.

7 BIBLIOGRAPHY

(1) Guiso, Luigi, and Paolo Sodini. "Household finance: An emerging field." Handbook of the Economics of Finance. Vol. 2. Elsevier, 2013. 1397-1532.

(2) https://www.bancaditalia.it/pubblicazioni/indagini-alfabetizzazione/2023-indagini-alfabetizzazione/statistiche_AFA_20072023.pdf

(3) https://www.bancaditalia.it/pubblicazioni/ricchezza-settori-istituzionali/2022-ricchezza-settori-istituzionali/index.html#:~:text=Il%20totale%20delle%20passività%20delle,basse%20se%20rapportata%20alla%20popolazione.

(4) https://www.moneyfarm.com/it/glossario-finanza/quantitative-easing/

(5) https://temi.camera.it/leg17/temi/riforma_previdenziale_eta_pensionabile.html

(6) https://mutuionline.24oreborsaonline.ilsole24ore.com/guide-mutui/irs.asp

(7) https://mutuionline.24oreborsaonline.ilsole24ore.com/guide-mutui/euribor.asp

(8) https://www.euribor-rates.eu/it/grafici-euribor/

(9) https://quifinanza.it/mercato-immobiliare/mutui-e-prestiti/mutuo-casa-famiglie-italiane/742804/

(10) https://www.borsaitaliana.it/notizie/sotto-la-lente/mutuifissovariabile.htm#:~:text=Se%20fisso%2C%20il%20tasso%20di,dell%27andamento%20del%20tasso%20EURIBOR.

(11) Campbell, John Y., and Joao F. Cocco. "Household risk management and optimal mortgage choice." *The Quarterly Journal of Economics* 118.4 (2003): 1449-1494.

(12) Coulibaly, B., & Li, G. (2009). Choice of mortgage contracts: Evidence from the survey of consumer finances. *Real Estate Economics*, *37*(4), 659-673.

(13) Johnson, Kathleen W., and Geng Li. "Are Adjustable-Rate Mortgage Borrowers Borrowing Constrained?." Real Estate Economics 42.2 (2014): 457-471.

(14) https://paolapozzolo.it/matrice-correlazione/

(15) https://infostat.bancaditalia.it/inquiry/home?spyglass/taxo:CUBESET=/PUBBL_00/PUBBL_00_02_01_04/PUBBL_00_02_01_04_02&ITEMSELEZ=MIR0400:true&OPEN=false/&ep:LC=IT&COMM=BANKITALIA&ENV=LIVE&CTX=DIFF&IDX=2&/view:CUBEIDS=MIR0400/&timestamp=1646397336224.

(16) https://murraylax.org/rtutorials/linearprob.html

(17) https://sites.google.com/site/econometricsacademy/masters-econometrics/probit-and-logit-models

(18) https://it.wikipedia.org/wiki/Modello_probit

(19) Kreft, Ita GG, and Jan De Leeuw. Introducing multilevel modeling. Sage, 1998.

APPENDIX A

DATABASE ANALYSIS ON SOCIAL AND DEMOGRAPHIC BASE.

I canceled duplicates of the NQUEST variable, effectively taking into consideration only the head of household (highest income earner). Thus, the study variables study is aimed at the head of the household. As male-headed households, we are slightly above 50 %. To facilitate the understanding of a possible table, I have considered three income brackets (**IRPEF income brackets**).

Fasce di REDDITO		uso scaglioni IRPEF, da internet
1	<28000	
2	tra 28000 e 50000	
3	>50000	

Table 13 - IRPEF Income Brackets

Let us summarize the variables considered in this first analysis:

ETA5	Age groups	Age groups
QUALP7	Main employment, work status	Main employment,
ANNO	Year	Survey period (year).
AREA3	Geographical area (3)	Division of Italy into 3
SESSO	Sex	Sex
STUDIO	Educational qualification	Educational

Table 14 – List of Variables

Each variable has *several characteristics*:

1 = North; 2 = Center; 3 = South and Islands;
1 = male; 2 = female
1 = none; 2 = elementary school; 3 = middle school; 4 = high school; 5 = bachelor's degree; 6 = post-graduate qualification

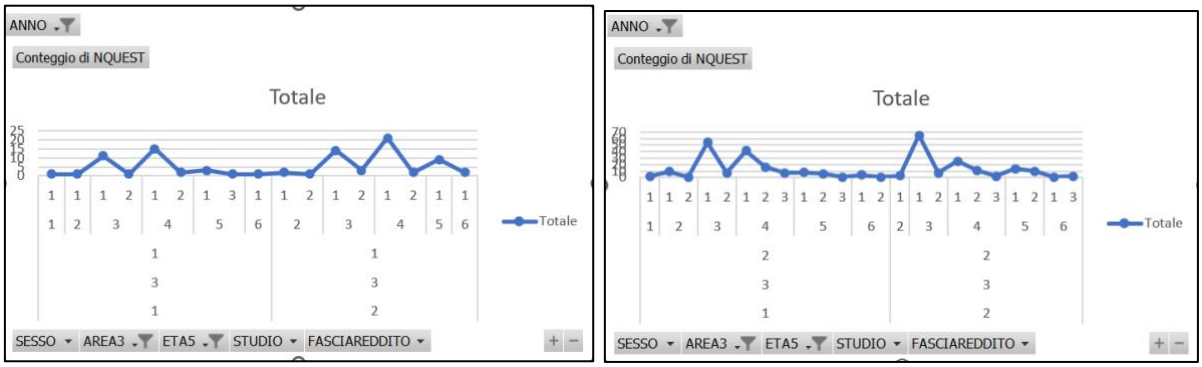
1 = up to 30 years; 2 = from 31 to 40 years; 3 = from 41 to 50 years; 4 = from 51 to 65 years; 5 = more than 65 years
QUAL7N
1 = blue-collar worker or similar; 2 = office worker or school teacher; 3 = manager; self-employed:
4 = member of the arts or professions; 5 = sole proprietor;
6 = free lance; 7= not employed (unpaid family member included).

Table 15 - Variable Explanation

When analysing AREA3 #3, ETA5 #1 and the year 2012, the first filters should be inserted. Additionally, the variable FASCIAREDDITO, which was created, should be inserted into the pivot table. If the output nvalues of the NQUEST are entered, the resulting scenario is as follows.

AREA INCOME BRACKET: How does it vary by ETA5?

(Instructions on how to read the graph: The bottom two bars represent gender, the two bars with the number 3 represent area3 = 3, and the two bars with the number 1 represent ETA5. The bars numbered 1-6 above represent educational qualifications, while the bars numbered 1-3 at the top represent income brackets).



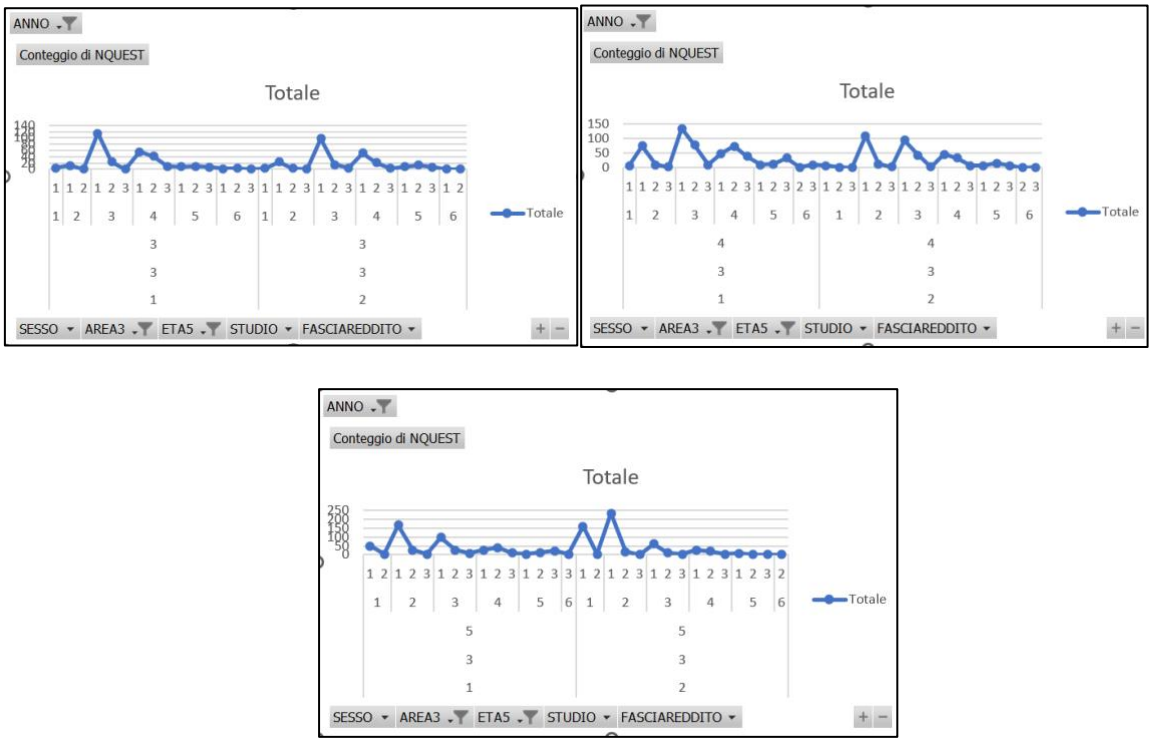


Figure 9 - pivot graphs with varying ETA5

Regarding the period between 2012 and 2014, it is observed that, among single males in AREA3 with INCOME BAND=1, the last age group has no cases of STUDIO=6. Additionally, the highest number of cases occurs at ETA5=2, with educational qualification n=3 (middle school). This trend is in line with expectations. It would be worthwhile to investigate the job qualifications associated with the earned degree and its impact on low income. It is observed that the low-income dynamic has less impact on the first two age groups, as shown in the bottom figure.

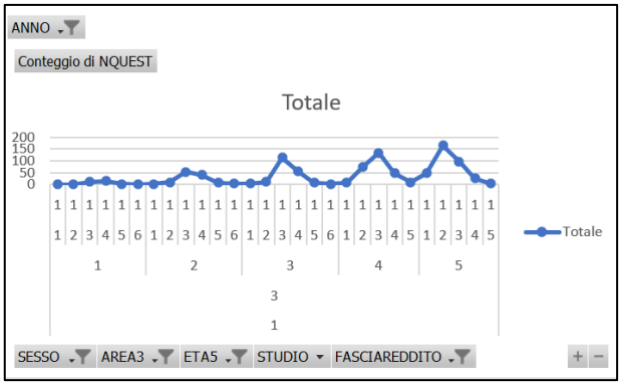


Figure 10 - pivot graph with SESSO=1,ETA5=2, AREA3=3

.

Looking at the data from 2012-2014, we can see that for the single female participants in AREA3 with FASCIAREDDITO=1, the overall numbers are lower than those of the male participants, except for the ETA5=5 bracket. However, we do see a similarity with the male participants in the last age bracket, where the largest number of minimum income earners have eighth grade as their STUDENT (see figure below).

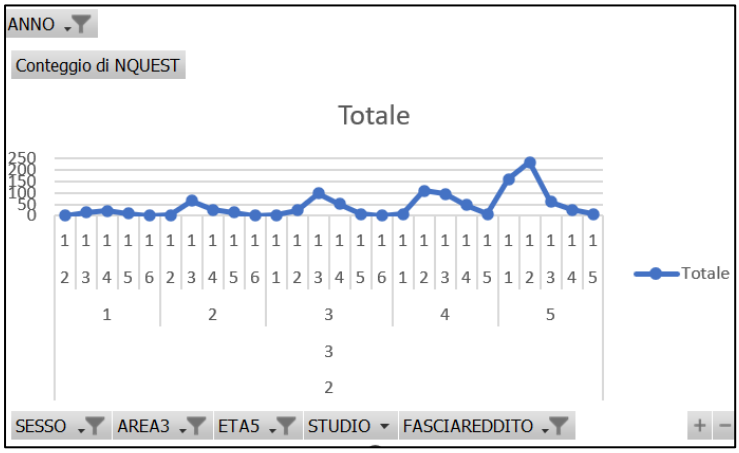


Figure 11 - Pivot graph with SESSO=2, AREA3=3, FASCIAREDDITO=1

In other areas, however, what is happening?

When going up the peninsula, we considered the male sex with FASCIAREDDITO=1 and ETA5=1,2,3. The age group most affected by low income is No. 3, with a peak among those with an eighth-grade education (the recorded peak is 32). Overall, the peak of low incomes is also associated with eighth grade, which generally has low numbers. It is expected that there will be an increase in the last two age groups, as shown in the second image.

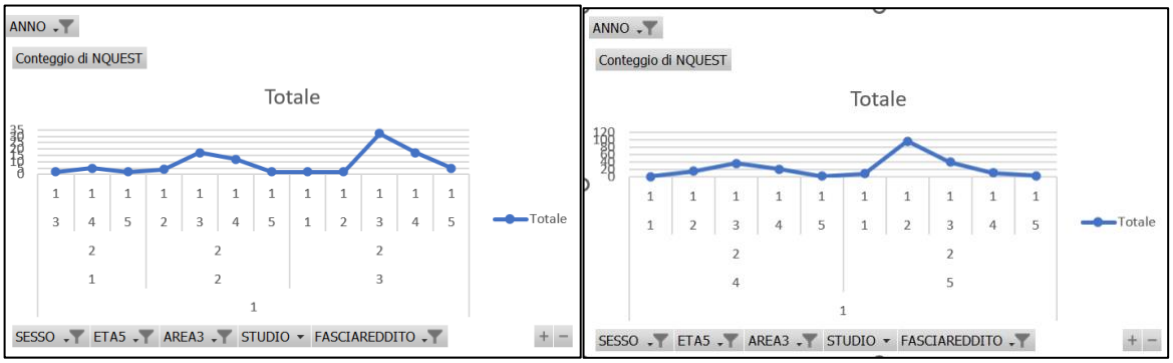


Figure 12 - Pivot graph with SESSO=1, AREA3=2, FASCIAREDDITO=1

It is evident that the number of income-earners in the lowest bracket (<28,000) increases significantly in most of the case histories, with the highest peak observed in STUDY 2 among the last age group (96 case histories with attained elementary school qualifications).

Regarding women in AREA5 with lower income, what is the current situation?

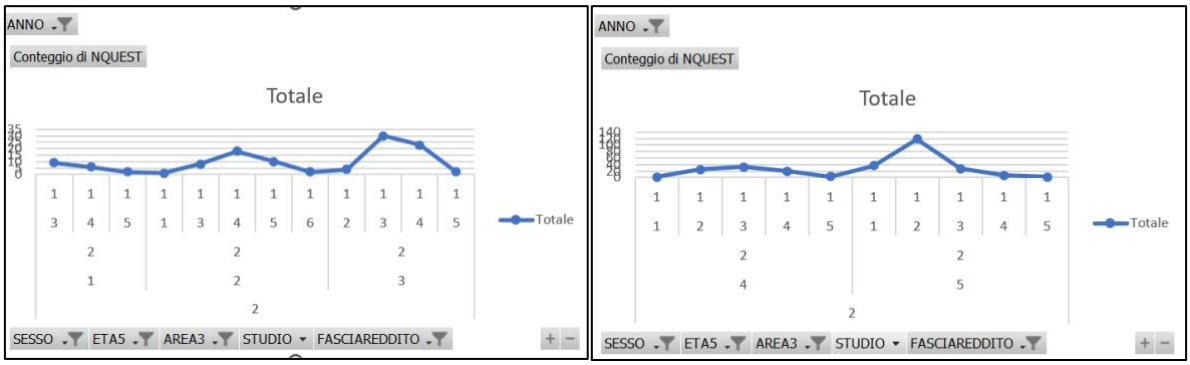


Figure 13 - Pivot graph with SESSO=2, AREA3=2, FASCIAREDDITO=1

The highest point is reached at ETA5=5, with a value of approximately 118, even higher than the same situation for males, when only an elementary school education is obtained. Generally, the numerical values appear to be slightly higher. A new trend is observed in this area compared to the previously studied one: between the ages of 31 and 40, low-income women's peak is associated with obtaining a high school degree.

In AREA3=1, a slightly different situation is expected. The first query will be for SESSO=1, and FASCIAREDDITO=1.

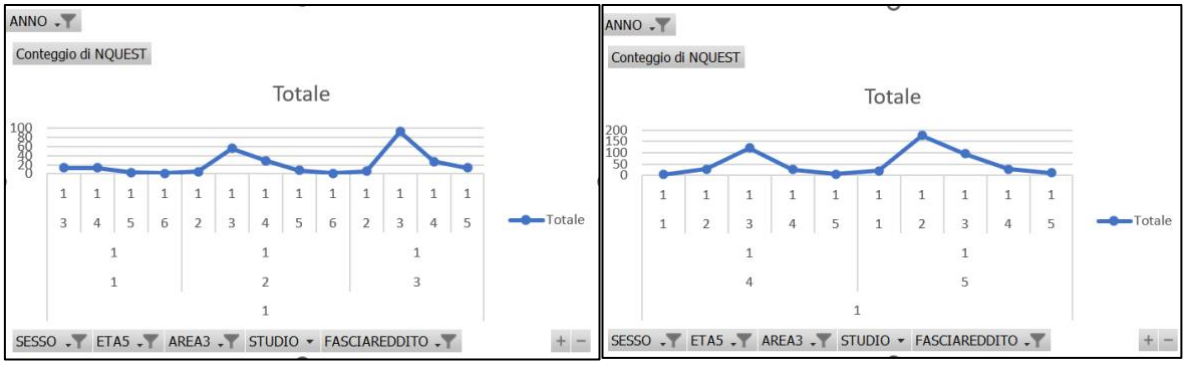


Figure 14 - Pivot graph with SESSO=1, AREA3=1, FASCIAREDDITO=1

In terms of trends, similar case histories are found compared to the previous two areas, with peaks of citizens who have low educational attainment corresponding to those who have completed only up to eighth grade. The absolute maximums, however, are found in the last age group, where individuals have only completed elementary school as their highest qualification.

I anticipate a slightly different scenario here. We will begin by querying AREA5=1, SESSO=1, FASCIAREDDITO=1 and observing the outcome.

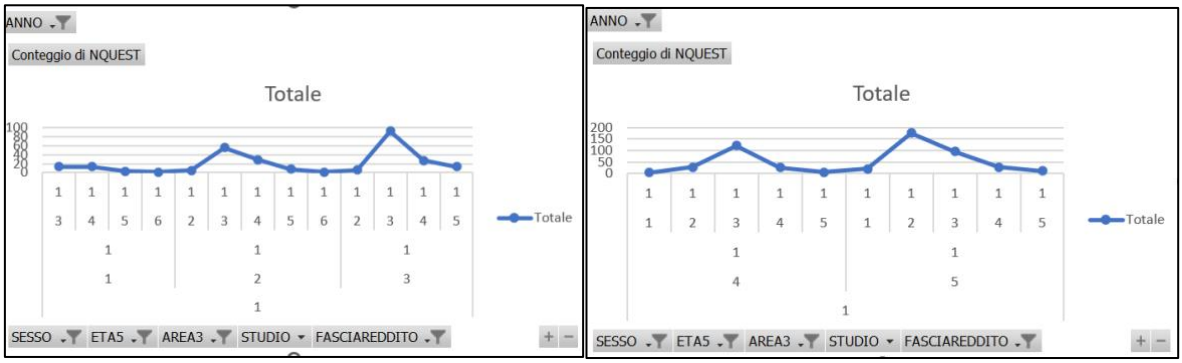


Figure 15 - Pivot graph with SESSO=1, AREA3=1, FASCIAREDDITO=1

We will first query AREA3=1, SEX=2, FASCIAREDDITO=1.

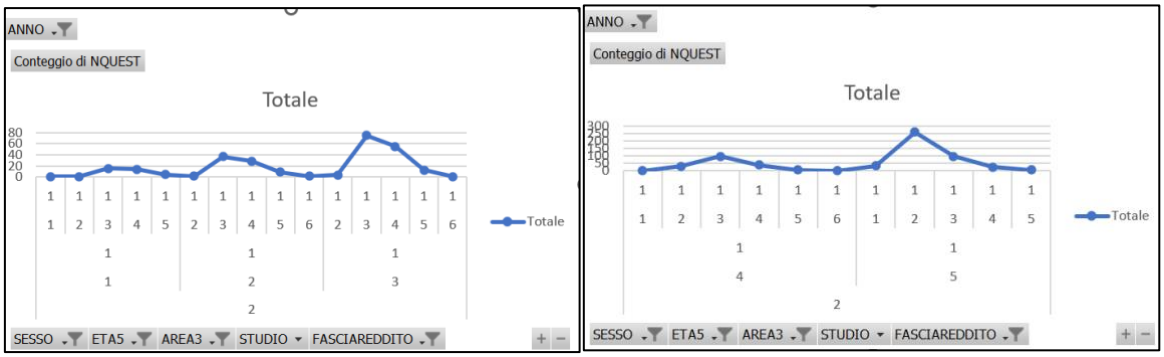


Figure 16 - Pivot graph with SESSO=2, AREA3=1, FASCIAREDDITO=1

In this initial section, I have introduced the variable SAVINGS S. To examine the relationship between income bracket and savings, I will not consider area or age group.

- In the first income bracket, the numerics of positive savings are in a 2:1 ratio to positive savings;
- The second bracket shows an increase in the number of votes in favour of TRUE, resulting in approximately 88.39%.
- In the final income bracket, there is an additional increase of 95.5 percent.

Leaving aside the smaller poll numbers when dealing with the third income bracket, it is clear that income and savings are directly proportional in this case.

Conteggio di NQUEST		
FASCIAREDDITO	SAVINGS positivo?	Totale
1	FALSO	3408
	VERO	6081
1 Totale		9489
2	FALSO	641
	VERO	4882
2 Totale		5523
3	FALSO	166
	VERO	3542
3 Totale		3708
(vuoto)	(vuoto)	
(vuoto) Totale		
Totale complessivo		18720

Figure 17 - Pivot Results of Savings Positivo for each FASCIAREDDITO

Regarding the occupations of our respondents, we aim to investigate whether there is a correlation between educational qualifications, job titles, and income brackets.

To achieve this, we will analyze the first variable in the survey: educational qualifications.

Conteggio di NQUEST			
STUDIO	QUALP7N	FASCIAREDDITO	Totale
1	1	1	20
	1 Totale		20
	6	1	5
		2	1
	6 Totale		6
	7	1	589
		2	41
		3	4
	7 Totale		634
1 Totale			660
2	1	1	154
		2	32
		3	7
	1 Totale		193
	2	1	1
		2	1
	2 Totale		2
	3	1	1
	3 Totale		1
	5	1	2
		2	1
		3	1
	5 Totale		4
	6	1	27
		2	40
		3	20
	6 Totale		87
	7	1	2754
		2	681
		3	105
	7 Totale		3540
2 Totale			3827

Table 16 - Pivot table with STUDIO =1,2; QUALP7N and FASCIAREDDITO

In the STUDIO, where bracket = 1 (corresponding to no degree), it is not surprising that most people are unemployed. Out of 589 people in age bracket 5, 551 are unemployed. The 589 unemployed people, therefore, reside in the lowest income bracket.

In the elementary degree (STUDIO=2) band, the trend should be similar to what was observed in the previous case, with higher numerical values.

The two employment figures are almost identical in the working class and unemployment, as intended. The income groups most represented are the minimum and lower levels for both qualifications. analyzed).

Conteggio di NQUEST			
STUDIO	QUALP7N	FASCIAREDDITO	Totale

In my opinion, the scenario analysed with respondents who possess a STUDIO=3 (eighth grade) may be the final case history similar to the first two:

- Numbers have significantly increased compared to previous cases.
- The two most common case histories are unemployment (57.2%) and blue-collar work (27%). In both cases, the most frequent income bracket is the lowest, with a net increase in Band 2 for both qualifications (Band 2 being approximately half of Band 1 income).

⊟3	⊟1	1	1031
		2	534
		3	94
	1 Totale		**1659**
	⊟2	1	158
		2	177
		3	55
	2 Totale		**390**
	⊟3	1	10
		2	8
		3	7
	3 Totale		**25**
	⊟4	3	3
	4 Totale		**3**
	⊟5	1	7
		2	10
		3	9
	5 Totale		**26**
	⊟6	1	233
		2	173
		3	130
	6 Totale		**536**
	⊟7	1	2233
		2	1041
		3	251
	7 Totale		**3525**
3 Totale			**6164**

Table 17 - Pivot table with STUDIO =3; QUALP7N and FASCIAREDDITO

- New phenomena are emerging as there has been an increase in the number of respondents in the freelance sector. It is important to note that this study may not be complete, which could result in partial skills and incomes for approximately 43.7% of respondents earning below EUR 28,000.

Conteggio di NQUEST			
STUDIO	QUALP7N	FASCIAREDDITO	Totale
4	1	1	337
		2	202
		3	28
	1 Totale		567
	2	1	402
		2	593
		3	299
	2 Totale		1294
	3	1	21
		2	78
		3	92
	3 Totale		191
	4	1	2
		2	14
		3	43
	4 Totale		59
	5	1	41
		2	64
		3	99
	5 Totale		204
	6	1	159
		2	187
		3	208
	6 Totale		554
	7	1	762
		2	788
		3	470
	7 Totale		2020
4 Totale			4889

When considering individuals with a high school degree, the statistics show a shift: while the largest group remains the unemployed (41.31%), the second largest group is now QUALP7N=2, which refers to office workers or school teachers, and generally office clerks. Consequently, the dominant income bracket changes to the second position.

When examining the most represented group, the unemployed, it is important to investigate their age. It is necessary to determine whether they are actively seeking employment or retired. To investigate the nature of unemployment, the variable ETA5 will be used. The variable etapens will not be considered due to the potential for bias arising from the nature of employment, such as workers with their own pIVA. Therefore, ETA will be used as an approximation method. As #5 represents those aged 65 and older, we can use this as a guide.

The figure representing the input of the variable ETA5 is shown below:

Conteggio di NQUEST				
STUDIO	QUALP7N	FASCIAREDDITO	ETA5	Totale
	7	1	1	58
			2	72
			3	118
			4	214
			5	300
		1 Totale		762
		2	1	10
			2	19
			3	40
			4	238
			5	481
		2 Totale		788
		3	1	7
			2	1
			3	10
			4	163
			5	289
		3 Totale		470
	7 Totale			2020
4 Totale				4889
Totale complessivo				4889

Figure 18 - Pivot graph with STUDIO=4; QUALP7N; ETA5; FASCIAREDDITO

The purpose of this analysis is to examine the 'unemployed' component (QUALP7N=7). It is noted that income bracket No. 2 is the most represented, which is consistent with the macroanalysis associated with STUDY=4.

In terms of age, income bracket No. 5 is the most represented, with 53% of the sample (300 for income=1, 481 for income=2, and 289 for income=3).

It can be assumed that most of these numbers are associated with the retiree bracket, but the remaining 47% cannot be.

Generally, the numbers of unemployed people in the 31-50 age groups are considerably lower than in the later age groups. Specifically, INCOME BAND=1 has more significant numbers in this regard.

What to say about STUDIO=5?

Conteggio di NQUEST			
STUDIO	QUALP7N	FASCIAREDDITO	Totale
5	1	1	43
		2	22
		3	6
	1 Totale		71
	2	1	177
		2	290
		3	239
	2 Totale		706
	3	1	23
		2	73
		3	120
	3 Totale		216
	4	1	4
		2	31
		3	181
	4 Totale		216
	5	1	49
		2	87
		3	418
	5 Totale		554
	6	1	47
		2	43
		3	112
	6 Totale		202
	7	1	158
		2	249
		3	473
	7 Totale		880
5 Totale			2845
Totale complessivo			2845

Unemployment is currently the most prevalent issue, followed by the second most common occupation, which is the same as in previous training. New occupations, such as sole proprietorship, are emerging. Further details on unemployment will be provided later.

The analysis highlights the increasing prominence of occupations such as self-employment and artisan work. It is important to note that the income brackets for all professions have risen. Bracket #1, which was in the minority in all professions in previous studies, has seen a significant drop in percentage. Income band No. 2 is locally higher for civil servants and teachers, while private professions have a strong component of income band

STUDIO	QUALP7N	FASCIAREDDITO	ETA5	Totale
5	7	1	1	23
			2	18
			3	14
			4	32
			5	71
		1 Totale		158
		2	1	7
			2	11
			3	10
			4	49
			5	172
		2 Totale		249
		3	1	3
			2	4
			3	8
			4	93
			5	365
		3 Totale		473
	7 Totale			880
5 Totale				880
Totale complessivo				880

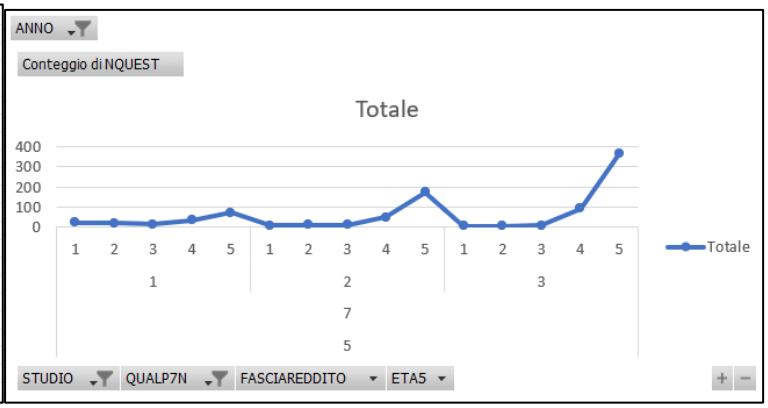


Figure 19 - Pivot graph with STUDIO=5; ETA5; QUALP7N and FASCIAREDDITO

As previously mentioned, respondents with a bachelor's degree predominantly fall into the highest income bracket (No. 3). This trend is also observed among the unemployed category. It is noteworthy that 88.86% of the pool of respondents are over 50 years old, with the majority of this percentage being represented by those over 65 (70% of the total). The majority of the unemployed in this pool are of retirement age.

WITH A STUDY LEVEL OF 6 (a master's degree and specializations)

Conteggio di NQUEST			
STUDIO	QUALP7N	FASCIAREDDITO	Totale
6	1	1	1
		2	1
	1 Totale		2
	2	1	11
		2	26
		3	24
	2 Totale		61
	3	1	4
		2	6
		3	24
	3 Totale		34
	4	2	5
		3	56
	4 Totale		61
	5	1	9
		2	10
		3	96
	5 Totale		115
	6	1	3
		2	4
		3	5
	6 Totale		12
	7	1	11
		2	10
		3	29
	7 Totale		50
6 Totale			335
Totale complessivo			335

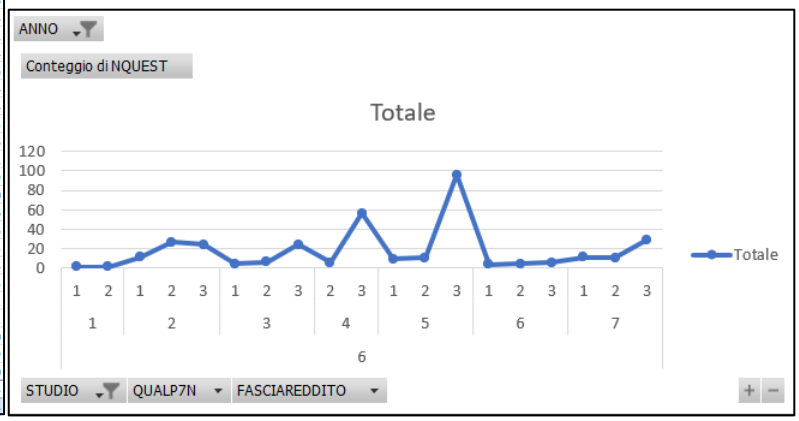
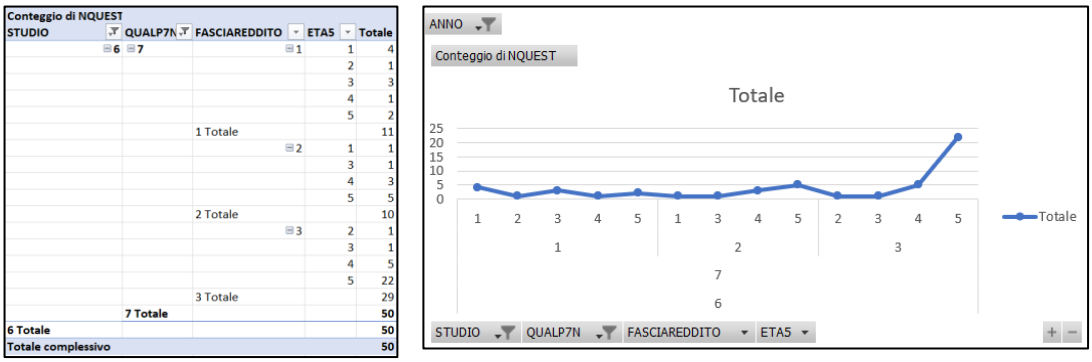


Figure 20 -Pivot graph with STUDIO=6; ETA5; QUALP7N and FASCIAREDDITO

In general, the numerical values are much lower (as previously analysed). FASCIAREDDITO n=3 accounts for approximately 69.85% of the total, while FASCIAREDDITO = 1 accounts for only 13.13%. The remaining 27% can be attributed to FASCIAREDDITO = 2.

The most common occupations among those with the aforementioned qualifications are those requiring 5, 4, and 2 years of experience. The number of unemployed respondents is a new addition to the analysis and represents the fourth category. The number of unemployed respondents is a new addition to the analysis and represents the fourth category. It is expected that a significant portion of this group is made up of individuals over the age of 65.



Conteggio di NQUEST				
STUDIO	QUALP7N	FASCIAREDDITO	ETA5	Totale
6	7	1	1	4
			2	1
			3	3
			4	1
			5	2
		1 Totale		11
		2	1	1
			3	1
			4	3
			5	5
		2 Totale		10
		3	2	1
			3	1
			4	5
			5	22
		3 Totale		29
	7 Totale			50
6 Totale				50
Totale complessivo				50

Figure 21 - Pivot graph with STUDIO=6; QUALP7N=7; FASCIAREDDITO and ETA5

Of the total number of unemployed individuals, 58% are associated with retirement age. This percentage increases to 76% when including individuals between the ages of 51 and 65. It is important to note that there are few outliers within this subcategory.

APPENDIX B

INTRODUCTION AND OVERVIEW ON DEBT FINANCING VARIABLES

The debt and credit items associated with the surveyed households will now be studied in detail. A comparison with income will be made to determine who requested these necessary resources.

At the outset, it is not advisable to focus solely on expenditures for renovation, real estate purchases, and related mortgages. Since there is no direct data available to understand the mortgage choices made by Italian households between 2012 and 2019, it is interesting to broaden the study to include hypothetical expenditures on durable and non-durable goods.

The FAM archive contains items that can assist with this.

ANTIC	Advances for real estate
MASTRIP	Extraordinary maintenance expenses
CONSAL	Food expenses
CONDIV	Expenses other non-durable goods
CONNDA	Non-durable goods expenses
CDUR1	Expenses for purchasing means of transportation
CDUR2	Expenditure on other durable goods
CDUR	Expenditure on durable goods
CONSA	Expenditure on durable and nondurable goods
CREALI	Expenditure on the purchase of real assets

VREALI	Amount of real goods sales

Table 18 - 'FAMI' dataset variables

On the same file, there are debt entries, such as:

DEB12A	Debt for purchase/restructuring
DEB12B	Debts for the purchase of other real assets.
DEB12C	Debts for the purchase of means of transportation
DEB12D	Debts for the purchase of other durable goods
DEB12E	Debts for the purchase of non-durable goods
DEB12F	Debts for other reasons
DEB12G	Informal Debts
PFCARTE	Credit card debts
RATA_AR	Rate for purchase/renovation debts
RATA_CONS	Debit interest rates for purchasing durable goods

To begin with, it is important to understand the basis for the choice of a fixed interest rate on a mortgage. This is determined by socio-economic variables, but what is the starting point for assuming the rate? The answer lies in the search for a home.

As a starting point, it is necessary to consider the debt related to individual properties, including consumer debt and its interest, which is typically more expensive than real estate. It is also important to consider household consumption and whether it can be traced back to the debt.

To find information about mortgages taken out by participating households, we need to access the 'FAMI' archive in our database. This archive contains the following variables:

DEB12A	Debiti per acquisto/ristrutturazione	Ammontare dei debiti verso banche o società finanziarie esistenti alla fine dell'anno e contratti per l'acquisto o la ristrutturazione di beni immobili.
DEB12B	Debiti per acquisto altri beni reali	Ammontare dei debiti verso banche o società finanziarie esistenti alla fine dell'anno e contratti per l'acquisto di altri beni reali (preziosi, oro..). Nei dati dell'Indagine sul 1987 comprende i debiti contratti anche con soggetti diversi da banche e società finanziarie.
RATA_AR	Rate debiti per acquisto/ristrutturazione	Ammontare totale rate pagate nel corso dell'anno per l'acquisto o la ristrutturazione di beni immobili.
RATA_CONS	Rate debiti per acquisto beni e servizi di consumo	Ammontare totale rate pagate nel corso dell'anno per altri debiti per esigenze familiari.

Table 19 - Debts & Payment's' variables

It should be noted that the debts to financial institutions associated with the mortgage are equivalent to the total amount existing at the end of the survey year. It is possible that many households have little remaining debt to financial institutions.

Another parameter that could be relevant for our investigation is the debt-to-real-assets ratio, which we will define as the ratio of consumer credit granted by financial institutions to participating households and the corresponding installments paid in the reference year. Additionally, the variables RATA_AR and RATA_CONS indicate the total amount of installments paid during the year to repay the two debts incurred. By understanding the average interest rates for consumer credit and mortgages, we can analyse the preference of Italian households towards the type of rate used.

First, we need to examine the number of households that took out a mortgage to buy or renovate their property.

When we filter the identification numbers of households from 2012 to 2020 in the "Debts" section, we notice the presence of inconsistencies that could mislead our analysis: there are in fact certain identification codes (NQUEST) that belong to households interviewed at the beginning of the 1980s (rising).

We recall, in fact, that the definition of DEB12A is the amount of debt existing at the end of the year attributable to the mortgage: it is obvious that these households interviewed 30 years ago still had a residual debt to pay off. We have therefore identified the first household interviewed in the questionnaire since 2012 with the number 871001, and consequently excluded from the analysis the identifiers that are smaller than the above number in all the years taken into account.

Before the proposed skimming, out of a total of 29966 identifiers, about 10.38% had debts to banks for the purchase/restructuring of their property, and of these 99.64% had paid an instalment to settle the debt contracted with financial institutions. The remaining 0.35% can be linked to two situations: a delay in the payment of the instalment (which generally does not have to end within the 180-day delay) or even, in the worst case, the termination of the contract by the banking institutions in the event of repeated delays in the payment of the agreed instalments. Another hypothesis that should not be dismissed is the usual presence of unintentional errors in the questionnaire.

In summary, the figures just described:

Conteggio di NQUEST SI O NO DEB12A	SI O NO? RATA__AR	Totale
FALSO	non presente	99,98%
	SI	0,02%
FALSO Totale		**89,62%**
VERO	NO	0,35%
	SI	99,65%
VERO Totale		**10,38%**
Totale complessivo		**100,00%**

Table 20 - Pivot table on the correlation between DEBI12A and RATA_AR

A new scheme to be carried out is to check for possible duplication of the very households that, on the contrary, fit perfectly into our reference period. In the same way that we have witnessed the repetition of past identification keys in the years of our

interest, this could also be the case for the households surveyed. In this case, however, the possible duplicates could be useful to us in analysing the payment status of the debt itself.

A new skim to be carried out is to check for possible duplication of the very households that, on the contrary, fit perfectly into our reference period. In the same way that we have witnessed the repetition of past identification keys in the years of our interest, this could also be the case for the households surveyed. In this case, however, the possible duplicates could be useful to us in analysing the payment status of the debt itself. After a considerable amount of skimming, we were able to match the numbers of the unique NQUEST (household number) code of the two different datasets: the first, which contains the characteristics of the households, the COMP dataset (revised at the beginning of our analysis), with FAMI, which, as already mentioned, contains information on mortgages and household debt to financial institutions. We can therefore first summarise the actual number of households that owed money to banks between 2012 and 2020:

		Valori	
Mutuo?	Rata_AR?	Conteggio di NQUEST	Conteggio di NQUEST2
⊟no	no	12451	99,98%
	si	2	0,02%
no Totale		**12453**	**88,26%**
⊟si	no	7	0,42%
	si	1649	99,58%
si Totale		**1656**	**11,74%**
Totale complessivo		**14109**	**100,00%**

Table 21 - Pivot table, figures on how many households opened a Mortgage

We note that, after removing the superfluous codes, the percentage of households applying for a mortgage rises by around 1.4 percentage points and stabilises at 11.74%. On the other hand, the anomalies mentioned above persist: 0.5% of households with debts to banks have not paid any instalments.

A very useful piece of information can be the information on the real estate of the households surveyed. After the usual skimming of the numerical identification codes, we have selected some variables in the 'IMMP' database relating to :

RESID: Main residence of household (1= main residence)
USOIMM: Use of the property (by setting RESID=1, automatically set as main residence)
VALABIT: Property value;
POSS3: Method of acquisition of the property, which can vary as follows: 1: purchased; 2: received as a gift; 3: purpose-built.
ANNPOSS: year of ownership of real estate;
ANNCOSTR: building construction year.

Table 22 - 'IMMP'useful variables

By entering information about the ownership and method of purchase of the property in the pivot, we can check that the information found in the previous pivot remains consistent. Below is the pivot showing the two scenarios:

casa di proprietà	Mutuo?	Rata_	POSS3	Valori Conteggio di NQUEST	Conteggio di NQUEST2
no casa	**no**	no	no casa	3450	100,00%
		no Totale		3450	100,00%
	no Totale			**3450**	**99,80%**
	si	si	no casa	7	100,00%
		si Totale		7	100,00%
	si Totale			**7**	**0,20%**
no casa Totale				**3457**	**24,50%**
proprietà	**no**	no	acquistato	4600	51,11%
			costruito apposit.	909	10,10%
			ricevuto in dono	3492	38,80%
		no Totale		9001	99,98%
		si	acquistato	2	100,00%
		si Totale		2	0,02%
	no Totale			**9003**	**84,52%**
	si	no	acquistato	5	71,43%
			costruito apposit.	1	14,29%
			ricevuto in dono	1	14,29%
		no Totale		7	0,42%
		si	acquistato	1232	75,03%
			costruito apposit.	60	3,65%
			ricevuto in dono	350	21,32%
		si Totale		1642	99,58%
	si Totale			**1649**	**15,48%**
proprietà Totale				**10652**	**75,50%**
Totale complessivo				**14109**	**100,00%**

Table 23 - Pivot table with DEBI12A, RATA_AR, POSS3

Households are firstly divided into two macro-areas, those who own their home and those who do not. The first group represents 75.5 per cent of the total, with 24.5 per cent of the interviewed households not owning a dwelling and probably living in rented accommodation.

The first group is therefore excluded from our analysis: it is in this group, however, that the 7 households without property, without an open mortgage and with an instalment debt over the years are to be found. This case may be related to an error in the collection of the questionnaire responses.

For the remaining 75.5% of households with a home, it should be noted that the majority of them, around 84.52%, do not have an outstanding mortgage with financial institutions, as opposed to the remaining 15.48%.

To better understand the nature of the mortgage, I thought I would highlight the method of acquisition of the property: in the pivot, you can see that there are three different categories in the column "POSS3": purchase, received as a gift and purpose built.

If we look only at the share of households with their own home, the share of properties received as a gift is remarkable (around 38.8%), if we exclude the households that applied for a mortgage; the share of owner-occupied dwellings is much lower.

On the other hand, the percentage drops considerably (we are at 21.1%) when households seek financial liabilities from banks. In the scenario we are going to analyse, the largest share of ownership (75.03%) will come from the purchase of the building, against a small share of construction (3.65%).

Therefore, 75.03% of the households that applied for a mortgage during the period we analysed needed help to buy/renovate an existing property.

In order to divide the type of mortgage into two macro-categories (fixed or variable), new assumptions have to be made.

- First, in the absence of information on the duration of the mortgage contract between the financial institutions and the households, it will be necessary to assume an average duration of 20 years for the amortisation plan. These data will be extremely useful, as it will be crucial to obtain the figures for the nominal amount paid over the years.

- As we do not know the total amount of loans granted by banks to households, we assume that it is equal to 80% of the mortgage value of the house;
- This assumption is used for all types of ownership of the property: whether it was built for own use, whether it was received as a gift (home improvement loan) or whether it was purchased. The hypothesis is extended to all types of ownership in order to remain as consistent as possible;

The next step will therefore be to study the behaviour of the households over the years of the survey: in the case that they have only been surveyed once, we will try to study the evolution and the decrease of the debt to the bank, starting from the year of ownership of the property up to the year of the survey.

However, for households interviewed more than once over the years, it will be interesting to study the trend between one measurement and the next. This could lead to more accurate measurements and analyses, or even to discrepancies between theoretical and actual results.

In order to study this phenomenon, it is necessary to divide the households analysed into sub-groups: The first and most significant of these will be households that have 'bought' (renovated, etc.) a property from 2012 onwards (ending in 2020).

In this case, knowing the average values of the current interest rates year by year, we will be able to outline the two possible scenarios, i.e. the total hypothetical values of the open mortgages in the case of fixed and/or variable interest rates.

By comparing the empirical data in the dataset and the results of our hypotheses, we should be able to draw some initial conclusions.

ANALYSIS OF THE VARIATION IN DEBT12A FOR DOUBLE HOUSEHOLDS

The first step was to identify those households whose unique code was repeated more than once within the analysis period.

Excluding all households interviewed for the first time in a period prior to 2012, there were 10 codes in the system that repeated across years:

Some of these were in 2 consecutive year bands, others repeated the interview 4 times in 2012, 2014, 2016 and 2020 respectively.

Examining the trend and variation of DEB12A and the resulting RATA_AR might reveal a trend attributable to the two mortgage types; or whether the same numbers present lead to random sequences and thus cannot be associated with any selectable mortgage type.

In fact, the 10 households highlighted did not show any patterns associated with an existing repayment plan:

- 50% of them had zero debt, although they had purchased the property in previous years and then reported a mortgage other than 0 in the subsequent and final interviews;

NQUE	ANNO	DEB12A	ANNOPO
871191	2012	0	2004
871191	2014	0	
871191	2016	**150000**	
871666	2012	**1000**	1997
871666	2014	**0**	
871697	2012	70000	1975
871697	2014	**0**	
871697	2016	**0**	
871697	2020	**0**	
872125	2012	0	1972
872125	2014	19000	
872125	2016	0	
872137	2012	0	1969
872137	2014	25000	

Table 24 - Variation of DEBI12A during the year for households

These are not hypothetical scenarios, or at least difficult to study theoretically, given the time lag between the purchase of the property and the taking out of a mortgage, the small sums involved in some cases, and the high level of repayment and cancellation of debt in a single two-year period (see families 871697 and 872125) in relation to their income bracket (FASCIAREDDITO).

Even if we assume that the numbers shown are not prohibitive for these families, in 4 out of 5 cases, excluding code 871125, the variations in DEB12A do not lead to anything: the null residual does not allow us to sketch any past time series. The situation does not improve for the remaining five households covered by this first scenario: as we will see in the next table, the deltaDEB12A is in fact difficult to read, between increases in the amount owed to banks and irregular and unworkable decreases from one two-year period to the next

NQUE	ANNO	DEB12A	ANNOPO	DELTADEB1
871385	2012	0	2007	
871385	2014	**30000**		30000
871385	2016	**19000**		11000
872012	2012	180000	2009	
872012	2014	170000		10000
872077	2012	55000	2006	
872077	2014	55000		
872077	2016	25000		30000
872082	2012	70000	2005	
872082	2014	75000		-5000
872082	2016	10000		65000
872190	2012	0	2001	
872190	2014	125000		
872190	2016	100000		25000

Table 25 - Delta(DEBI12A) of households

Even for the remaining five households, as expected, it is difficult to see a pattern that can be attributed to a mortgage repayment plan.

Taking NQUEST 872077 as an example, the debt contracted with the banks seems to remain the same over the two years 2012-2014 (it is difficult to imagine any form of renegotiation of the mortgage over such a long period) and then falls dramatically in 2016 by around EUR 30,000.

Or again, the household identified by code 872082 has a contracted residual debt of EUR 70,000, a figure that increases over the next two years with a delta of EUR -5,000, and then collapses to only EUR 10,000 in the 2016 biennium (a repayment delta of around EUR 65,000 in just two years, hardly credible).

In conclusion, the study of mortgage changes for 'double' households did not yield any significant results. The trends seem confusing and of little use for our analysis. We will now proceed with the study of all those households (about 287) that appear once in the SHIW database.

It will be necessary to study a method to validate the actual data, corresponding to the outstanding debt (DEB12A) or the instalment paid in the survey year (RATA_AR), without being able to take advantage of the variations between the different survey years.

The first attempt in this sense focused on trying to identify the real variable DEBI12A proposed by the database in one of the two theoretical variables I created, FIXRATEMORTG and VARRATEMORTG.

The first step was to calculate, as explained above, 80% of the mortgage value of each property owned by the households applying for a mortgage, which at the same time corresponded to the role of main residence (=1 in TYPEIMM).

The calculated value will, in fact, hypothetically correspond to the amount that financial institutions will be willing to grant as mortgage financing.

We have some missing data and some unknowns in our database: since we do not know the start date of the mortgage, we will use the two variables ANNO and ANNOPOSS, the year of the survey and the year of the start of ownership of the property respectively, to calculate what we are looking for.

Assuming that ANNOPOSS is the year of mortgage inception, the difference between (ANNO-ANNOPOSS) will give the duration of ownership of the property and, in parallel, the progressive duration of the mortgage itself.

It is now time to dust off the annual interest rates we obtained at the beginning of our analysis: we will need them to determine the values of fixed or variable rate mortgages for each year of the study.

The first idea is very simple: we artificially construct, for each household, the total hypothetical value to be repaid to the financial institutions for granting the mortgage (nominal part + interest part) according to two scenarios:

- Fixed-rate mortgage with amortization plan over 20 years;
- Variable-rate mortgage with amortization plan over 20 years.

As these values are purely theoretical, a margin of error of ± 5% should be included in the identity to avoid possible errors due to approximations and/or assumptions made.

For respondents with mortgages that have already been partially repaid, since we have no guidelines on the type of mortgage chosen, we work backwards and linearly add the instalment paid in the current year to the number of years of mortgage already paid. This is a fairly obvious stretch, but it is the only way to sketch a total debt from a residual value.

We could not use a linear function to sketch the total because we did not have a time series that would allow us to do so.

In summary, therefore, the variables in this specific data analysis are for the most part already known and defined, such as:

VARIABLE	DESCRIPTION	CODING
ANNO	*Survey reference period (year).*	
NQUEST	*Questionnaire number – household ID*	
TIPOIMM	*Age class*	
DEBI12A	*Amount owed to banks or finance companies at the end of the year for the acquisition or improvement of immovable property.*	
ANNOPOSS	*Year in which ownership started*	
DURATAPOSS	*#years owning the property*	
FIXRATEMORTGAGE	*Theor. total repayment for a fixe – rate mortgage*	
VARRATEMORTGAGE	*Theor. total repayment for a vari – rate mortgage*	
CONVALIDA	*(see below)*	

Table 26 - Table with all the built-Fictitious Variables

Let us consider the variable CONVALIDA for a moment:

It contains a nested logic function which checks whether the value associated with the current variable DEBI12A is in one of the two value ranges FIXRATEMORTGAGE before and VARRATEMORTGAGE after. If it is not in one of them, it returns the string 'verify'.

```
=SE(E(I2<=T2;I2>=U2);$Q$1;SE(E(I2<=V2;I2>=W2);$R$1;"verific"))
```

Figure 22 - CONVALIDA nested logic function.

However, the results shown by the CONVALIDA variable are extremely negative. In fact, out of a sample of 277 households, there are 272 households whose mortgage does not fall into either of the two specially created batches. However, the remaining five codes are mostly assigned to FIXRATEMORTGAGE and VARRATEMORTGAGE:

Conteggio di NQUEST	
CONVALIDA	Totale
FIXRATEMORTG	4
VARRATEMORTG	1
verific	272
Totale complessivo	277

Table 27 - Pivot table, division as TYPEOFMORTGAGE

This cannot be the right way forward: the factors that have led to this very high percentage of unidentified types can certainly be traced back to one or more of the following hypotheses:

- It is not correct to exclude a priori other debt repayment schedules: households may have chosen schedules spread over several years or even fewer years; considering only 20 years may exclude several unique codes;
- The linear approximation used to obtain the DEBI12A total (sub-values) is conceptually incorrect and has led to inaccuracies in the calculation, since the

instalment paid in the current year is not necessarily fixed for the whole duration of the mortgage (as it would be in the case of a fixed-rate mortgage).

It is therefore necessary to follow another path, one that is more precise and less based on loose assumptions, one that allows us to reach the point where we can classify mortgages according to their type.

Therefore, if we abandon the hypothesis of using the DEBI12A variable, we are left with only one variable that can be linked to the payment of mortgage instalments, namely RATA_AR.

This measure is certainly more concrete than the one used previously: DEBI12A refers to a sub-variable, whereas RATA_AR refers to the individual payment during the year of analysis; in this case, we could use the different interest rates found in the different years and compare the hypothetical instalments with the actual ones.