



**HAL**  
open science

# Accelerating the distribution of financial products through classification and regression techniques.

Edouard Ribes

► **To cite this version:**

Edouard Ribes. Accelerating the distribution of financial products through classification and regression techniques.. 2023. <hal-04167233>

**HAL Id: hal-04167233**

**<https://hal.science/hal-04167233v1>**

Preprint submitted on 20 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

July 20, 2023

# Accelerating the distribution of financial products through classification and regression techniques.

*A case study in the wealth management industry.*

Edouard A. RIBES <sup>a,1</sup>

<sup>a</sup>*Researcher at Mines Paristech & C.T.O at Manymore, Paris, France*

---

<sup>1</sup>Corresponding Author: Edouard Augustin Ribes, Mines Paristech, Paris, France; E-mail: edouard-augustin.ribes@mines-paristech.fr

July 20, 2023

**Abstract.**

**Background context.** To date, financial products mostly consists in instruments used by households to prepare for retirement (consumption smoothing across the various stages of and one's life-cycle) and/or to transfer wealth across generations. However, the current financial system is not completely efficient in doing so as the level of participation of households remains low. This calls for potentially new tools and techniques aimed at accelerating the distribution of the said products.

**Specific knowledge gap the work aims to fill .** The current studies available on Fintechs & the distribution of financial products mostly revolve around notions of robo-advisors. Use cases usually orbit around automatically structuring & managing households portfolios under risk (incl. diversification) and performance constraints (e.g. what is the mix of bond and equity to use within a retirement plan to optimize the performance measured as the Sharpe ratio?). However there is no available study on the choices made by households on the products themselves, the associated drivers and the level of investments, a gap this study aims to bridge.

**Methods used in the study.** This study leverages a private data-set from a French Fintech describing at a macro level the structure of 1.5K+ households and their wealth. Post feature selection and data cleansing, the information is fed to standard classification algorithms (e.g. random forest, support vector machine...) and regression techniques (linear regression, multiple kernel estimators...) to predict whether or not households are likely to subscribe to a life insurance or a retirement plan or a real estate program over the forthcoming year and to forecast the associated level of investment .

**Key findings.** The current data-set shows that macro levels indicators on French household structure and wealth can be used to predict their subscription behavior over the next 12 months towards core investments products with a high level of performance ( $A.U.C > 90\%$ ). However, macro level information appears insufficient to predict the level of investments on those products ( $R^2 < 30 - 40\%$ ).

**Implications.** The study shows that standard classification techniques could be used by financial advisors to accelerate client discovery on their existing portfolio. This could thereby improve the distribution of financial products and result in productivity gains for those professionals.

**Keywords.** Wealth Management, Brokerage, Machine learning, Classification, Fintech, Technological Change.

July 20, 2023

## 1. Introduction:

Personal banking pertains to the way households manage their finances and leverage the associated eco-system to meet their needs (Campbell, 2006; Tilmes and Schaubach, 2006; Maude, 2010; Collardi, 2012). Those needs have traditionally been articulated around four pillars: payment, loans, investment & and insurances . This has given birth to an industry structured around two core activities. On one hand, advisors and brokers distribute/sell financial solutions and products addressing those needs. On the other, asset managers, banks and insurers process the associated orders and produce the associated services.

Personal banking has, however, been the subject of recent critics as the associated services are not only considered expensive but also have not fundamentally changed over the past decades. For instance, investments products (e.g. private retirement plans based on equity) which yield, on average, a 5% return, have been the subject to a 2-3% fee every year (Philippon, 2016; Bazot, 2018). As a result, for every \$ invested, only 2 or 3c are earned by a household, which barely covers the inflation (currently running at a 1-2% rate in most mature countries).

Those longstanding critics have thus given birth over the past decade to the Fintech movement (Philippon, 2022), an entrepreneurial stream aimed at leveraging automation technology to make personal finance more efficient. And of course, this entrepreneurial movement has had its academic counterpart (see (Knewton and Rosenbaum, 2020) for a definition). Fintechs nowadays target both households and financial professionals. When it comes to households, the questions Fintechs are trying to address are mainly one of speed. Recent evidences, for instance, show that automation technologies have proven useful to reduce by about 30% the time it takes to get loans and insurances. When it comes to professionals, the question then becomes one of productivity and Fintechs aim there at reviewing the value chain of a profession (Grossman and Rossi-Hansberg, 2008) and automating as many activities as possible so that professionals can serve more households (Todd and Seay, 2020; Chatterjee and Grable, 2022). Note that this should, in principle, also lead to a reduction in the price of the financial products/services.

If examples of Fintechs applications are numerous in the literature when it comes to loans and insurances (Vives, 2017; Navaretti et al., 2018; EBE, 2017), they are quite sparse when it comes to investments products. Most examples indeed focus on the topic of robo- advisors, which, if efficient, are not adopted in a wide fashion as their usage suffers not only from black-box considerations (Polansky et al., 2019) & lack of trust (Mutual, 2017; Cull, 2022) (for instance, studies show that only 29% (resp. 53%) of babyboomers (Gen X/Y) trust robo-advisor) but also is not so frequent (most financial portfolios associated to investments are indeed only re-balanced one every year).

The value chain associated to the distribution of financial investments products is however full of automation opportunities. In essence, it boils down to four steps (Maude, 2010). Contacts (i.e. households) are first garnered and qualified (e.g. level of revenue, financial objectives etc...). Based on this qualification, pre-sales activities then occur to match clients' needs with one or two products based on both financial & fiscal considerations. If interested, households then subscribe to the product. Finally, post subscription, households embark on a cycle of yearly reviews with their financial advisor. This last step, namely, the client review is, as of today, non efficient. An advisor has indeed about

July 20, 2023

100 clients (Foerster et al., 2017) in its rooster<sup>2</sup> and needs to spend about several hours for each clients. This time is indeed spent to contact them, understand if their household structure has changed and to assess if new products could be useful. This thus requires at least 100h of the total of 1750 working hours per year available per advisor<sup>3</sup>. The reason for this non efficiency is that advisors main role is to distribute financial products and that less than 10% of the existing clients actually subscribe to a new one during those reviews. Other activities (i.e. re-balancing of positions in existing financial products, information updates), which occur during those reviews, can indeed easily be (if not already are) automated (e.g. via robo-advisors, digital client information files (referred to later as K.Y.Cs)). An option to gain in productivity is therefore to accurately target this 10% section of the portfolio which needs new investments products. This is where automation technology, notably through the use of artificial intelligence (I.A.) can become useful, something that this paper will illustrate.

This illustration will be two-fold. First, this article will show that standard information available in K.Y.Cs can be used to assess which household is likely to subscribe to a financial product over the next year. This will be done by using an anonymised and randomized proprietary data-set from *Manymore*, a french Fintech firm specialised in productivity tools for financial advisors. Second, this paper will show that the same information (i.e. households' K.Y.C) can be also used to accurately predict the amount they are likely to invest. Using the previously detailed figures, it can be estimated that equipping advisors with this type of tools could boost their productivity by at least 5% as they would be not only able to target who to contact in their clients' rooster, but also be able to prepare pre-sale pitch to accelerate the distribution process. Note that the contribution of this article are likely to be broader than this simple use case. This will be further discussed in section (2) in light of the current challenges highlighted by specialists in the field of personal finance.

Now, in terms of structure, this paper will first begin with a small literature review highlighting, with more details, what's currently known about the Fintech movement and what is understood of the field of personal banking. A methodology section will then follow to explain the type of data that was used, the classification techniques which were employed to assess which household is likely to get an investments product and the regression methods which were used to predict the amount invested by households. Results will then be reviewed in details and a discussion will be articulated around the limitations of this study as well as areas of future research. A crisp conclusion will then wrap the article.

---

<sup>2</sup>Each client has on average 2 financial products leveraging a mixture of equity and bonds organised on 4 to 5 supports.

<sup>3</sup>This is based on the assumption of a 40h working week, 50 weeks per year and a 80% of time availability for clients related matters.

July 20, 2023

## 2. Theoretical background:

According to the recent review of (Grable and Chatterjee, 2022), personal banking is articulated around four domains: loans, insurances, payments means and investment tools. The methods described in this article are aimed at facilitating the distribution of this later category of product. The contribution of this paper will therefore be illustrated by first depicting what is known of households' investments behavior, then by proposing a view of the current challenges acknowledged by academic and practitioners associated to this field and finally by highlighting the specific contributions of the proposed methodology.

### 2.1. Households & investments:

Investments products are structured to help individuals & households accumulate capital according to smooth their consumption throughout their life-cycle. The objective is to achieve at a stable revenue under the form of a permanent income (Friedman, 1957). If loans are used to transfer future wealth to the present, investments products are used to transfer current wealth to the future. Investments are thus mainly used as a vehicle to accumulate wealth during the active period of life and to smooth consumption during one's retirement (for instance through pension plans/funds) and throughout generations. Empirical studies such as the one of (Foerster et al., 2017) have indeed shown that investments and the associated advice was indeed primarily a topic for households whose head is in his/her 50s and who seeks to prepare for retirement . Those households generally invest about 100k\$ on 1 to 2 products (e.g. a life insurance and a retirement plan) and investments are spread across 5 to 6 supports (e.g. a mixture of equity and bonds). Beyond retirement, households also leverage investment products to transfer wealth across generations. Transmission is generally facilitated through specific products such as life insurances or through products structured to facilitate bequests. Bequest are indeed a tool frequently used by households to tie together generations through a specific incentive schemes (the "strategic motive" described by (Bernheim et al., 1986)) and/or to serve altruistic objectives of more experienced individuals (Wilhelm, 1996).

Investments are generally categorized depending in their nature: physical investments, which notably encompasses households' primary residence & financial ones (e.g. pension plans etc...) (Cocco, 2005). On this front, empirical studies (Bogan, 2008; Christelis et al., 2013) indicate that 30% to 50% of households own financial investments products (with notable differences across countries) and that about 70% to 80% own physical investments products, notably through their home. General decumulation patterns associated to investments are simple: physical assets can generate a form of annuity through a reverse mortgage or be sold so that the associated capital is converted into financial investments. Financial investments can then either generate an annuity (although a large portion of households remain reluctant to leverage this option (Peijnenburg et al., 2016)) or be subject to annual withdrawals which are generally of about 4% of the total accumulated capital (Bengen, 1994) (with variations and adjustments based on individuals' risk appetite (Pfau, 2010; Murguía and Pfau, 2021)). As for bequest, they are generally structured either as a lump sum payment or as the transfer of the full or partial bare ownership of an asset (Masson and Pestieau, 1997). Bequests are yet heavily regulated, notably to limit an excessive polarization of wealth across households (De Nardi and Yang, 2016; De Nardi and Fella, 2017).

July 20, 2023

## 2.2. Investments challenges:

When it comes to investments-related instruments, a number of dynamics are currently at play. First, local regulations are being reinforced (Armour et al., 2016) to protect households from hazardous providers and avoid situations similar to the subprime financial crisis in 2009. If this generates an increment in labor for professionals distributing investments products and dampen their productivity, this also generates opportunities in terms of technological replacement since regulatory activities are heavily scripted (see (Grossman and Rossi-Hansberg, 2008) for the seminal theory on labor displacement and replacement). Those opportunities are currently addressed by RegTech actors, service providers specialized in digital technologies automating regulatory tasks (Anagnostopoulos, 2018; Buckley et al., 2020). Note that this regulatory pressure is increasing the specialisation of actors partaking to the distribution of investment products.

Second, the way households garner wealth (notably through labor related income) is shifting. Global labor displacement (off/near-shoring) and replacement (through automation technologies) are indeed driving a polarization of households income structure (David et al., 2006; Harrigan et al., 2016). As a result, households savings capability is, on average, decreasing (Wisman, 2009). The polarisation of the economy at play in mature countries indeed entails an increased proportion of low income jobs with no savings capability, a shrinking "middle class" who could save to invest and a reinforcement of the economic means of the wealthiest (who could already save about (if not more than) 20% of their income (Huggett and Ventura, 2000))). When it comes to investments products, those stylized fact simply translate is an increased challenge in terms of access. On one hand, wealthy households benefit from the change and invest more heavily, whilst the others actually exit financial markets. Note that the challenge nowadays appear different from before. In the nineties, the question of access was linked to informational constraint (Guiso and Jappelli, 2005) and access boomed with the development of the internet (see for instance (Bogan, 2008) for households measure on the evolution of the stock market participation). Several decades later, participation rates have not only plateaued but are on the decline (Merkoulova and Veld, 2022). Interestingly, the answer on the distributor side has been to attempt to lower investment fees (currently in the 2-3% range) which are deemed too "high" (Philippon, 2016; Foerster et al., 2017). This spurred massive investments towards robo-advisors (Fisch et al., 2019). However, the adoption of those tools has so far been limited and only a handful of technology fuelled initiatives have been recorded on other investment related topics (Ribes, 2022, 2023). Robo-advisors indeed appear to be of use for small transactions and investments and therefore limited to a certain fringe of households (Phoon and Koh, 2017). But the bulk of investments (measured in terms of asset under management - A.U.M) still come from wealthy households who prefer to leverage physical advisors. Their investment decisions indeed mix objective and subjective components which are best addressed through human interactions (Rossi and Utkus, 2020).

The third and last main investments challenge encountered by households revolves around portfolio choices and diversification, which is heavily intertwined with notions of financial literacy and behavioral finance. Empirical studies (such as the one of (GAUDECKER, 2015)) have indeed shown that households equipped with either basic financial notions or supported by a financial advisor achieve a reasonable level of diversification. However, considering that about 60% of households (Alhenawi et al., 2013) do

July 20, 2023

not possess enough financial knowledge<sup>4</sup> and that only 10 to 15% of households benefit from the services of a financial advisor (West, 2012), it would seem<sup>5</sup> that at least 30 to 40% of the households who can invest are making poor portfolio choices.<sup>6</sup> At an aggregated level, this translates into a landscape where investing households leverage on average only one equity stock, which is generally the stock of their employers (Kelly, 1995). But beyond notion of financial literacy, households are also known to have behavioral biases when it comes to investing. This is associated to personal traits such as risk taking, overconfidence and time preferences (Badarinza et al., 2016; Lyons and Kass-Hanna, 2022). This is something that the field of behavioral finance has consistently addressed at an individual level since the eighties (Hammond, 2015) and which is now mutating to consider how individual behavior are modeled by social norms as the field is shifting towards social finance (Hirshleifer, 2015).

### 2.3. Current contribution:

With respect to the macro-level challenges highlighted in the previous sub-section, the tools & methods described in this paper can contribute in two ways. Those tools are indeed aimed at democratizing the services provided by financial advisors. As such, they first contribute to the diversification of households portfolio as more advice leads to more diversification. Second, they act as productivity tools which can free advisors' time so that they can acquire new clients. At scale, this could potentially lead to an increased participation of households to the financial market. The proposed tools are therefore also a contribution to the access challenge highlighted above. To further illustrate the current contribution of the article, a short discussion will follow on the activities of financial advisors and on how the proposed methods could be integrated in their day to day.

Looking at both existing benchmarks (Foerster et al., 2017) and the professional experience gathered within *Manymore*<sup>7</sup>, a financial advisor serves between 70-80 households. Each household has 1 to 2 financial products (representing a total of about 110 products managed per advisor). From an advisor point of view, those products are divided in two blocks. About 80 products are considered as mature. They represent an investment of about 50-100k\$. Given that advisors are compensated retro-commission of 2 to 3% of the invested capital, they generate a recurring revenue of about 110-120k\$ per year. The activity on those products is mainly one associated to re-balancing: advisors indeed need to make sure that given the last market variations, the positions of their clients are diversified and meet households expectations in terms of risks and returns. Then about 20 to 30 products managed by advisors are products which have been subscribed during the

---

<sup>4</sup>This type of measure is however directional as there is no standard metrics/consensus when it comes to assessing financial literacy (Warmath, 2022). Besides reported figures mainly depicts the situation observable in the US and differences may exist across countries.

<sup>5</sup>Considering that financial literacy, the availability of savings and the use of advisors are independent.

<sup>6</sup>It would be interesting here to further invest in a comprehensive survey to assess the portion of households making poor investment decisions. Typically, households who invest generate a level of income and posses a higher level of education (incl. financial literacy).

<sup>7</sup>Manymore is currently one of the leading software providers for financial advisors in France. It equips more than 2400 advisors with a comprehensive list of solutions enabling regulatory activities as well financial advice and product distribution. Manymore's software suite provides tools ranging from C.R.M and digital K.Y.Cs used to onboard households and capture their data to digital subscription and re-balancing forms to activate the distribution of products.

July 20, 2023

year. Investments on this front are on average of 10k\$ and the pool of product generate a revenue for the advisor of about 5k\$ per year. Note that this is something that will be further described in the methodology section when the data-set available for the study gets depicted. In the meantime, this explains how the average productivity 120k€ of an advisor can be decomposed. Activity wise, advisors work about 210 days per year (of course, with variations depending in local regulations). Mature contracts generally require an annual re-balancing which takes at least a day of work, leaving about 150 days of work to distribute 20 to 30 new products (to either new or already known households). The activity advisors is thus heavily skewed toward the distribution of new products although the core of their revenue is associated the products which have been under management for a long time.

In light of this structure, the tools proposed in this article can be used in two ways. First, they can be positioned online and made available directly to the households. Given a certain level of promotional effort to make sure that households are made aware and leverage the service, this could be used to generate qualified leads and reroute households to financial advisors. This mechanism could serve to democratize financial advice and, as previously mentioned, contribute to more diversification at an aggregated level. Second, the tools could be offered directly to advisors so that they can scan their portfolio of existing clients. This would translate into an early identification of households who can benefit from financial products and the days of work saved here could be repurposed to increase either advisors' efforts towards households' education (therefore boosting the overall societal effort towards more financial literacy) or towards clients' portfolio expansion (which would benefit the overall eco-system).

### 3. Methodology:

There are two main goals to this study. The first objective is to test whether or not standard classification techniques can be used to predict accurately the subscription behavior of households over the next twelve months towards specific investment products (namely life insurance, private retirement plans and real estate assets). The second ambition is to assess the extent to which their level of investments during the first year can be forecasted with standard regression techniques. Note that this analysis is based on information which is normally available within financial distributors Customer Relationship Systems [C.R.M]. To better understand the current intent and its applicability, this section will focus on describing the kind of data-set (sub-section 3.1) that can be found in such information systems as well as the methods that were used (sub-section 3.2) to draw the results detailed in section (4) .

#### 3.1. Data:

The data used for this study stems from one of *Manymore's* database. This data set, which is privately held, is anonymized. The data holds anonymous information about 2439 households in France. Within this sample, 1622 (66%) households have detailed information in their "Know Your Client" records [K.Y.C] regarding their assets. Those households have been used as the baseline sample for the study. From a reference standpoint, households were represented by 488 distinct independant financial advisor [IFA].

July 20, 2023

Each IFA covers on average 12.7 (+20) households. Note that this is lower than what would be expected in the profession to date, as IFAs traditionally service about 80 to 100 households (Foerster et al., 2017). This can however be explained by the fact that the software associated to the selected database is primarily used by IFAs to provide in depth fiscal and financial advice. If this kind of advice comes with the capture of detailed information on households, which comes handy for this study, its scope is limited to a subset of IFAs' client rooster ( 20%).

Among the households considered in the study, 232 hold a private retirement plan (referred to as **RP** in the rest of the study) (14% of the sample) of which 59 were initiated in year prior to the study. 870 households hold a life insurance (referred to as **LI** in the rest of the study) (53% of the sample) of which 202 were subscribed in the year prior to the study. Finally 1435 households (88% of the sample) hold real estate assets (referred to as **ReA** in the rest of the study), 88 of which were bought in the year prior to the study. Investments represented, on average, respectively 280k\$ (+/- 869k\$) for LIs, 30.3 (+/- 61k\$) for RPs and 282k\$ (+/- 372k\$) for ReAs. This means, based on the current sample, that each professional has a 12% (resp. 3.6%) chance to distribute new LIs (resp. RPs) and has a 5.4% to support a new investment on ReAs in one year.

The data points recorded in *Manymore*'s system for the 1622 households (indexed by the letter  $n$  in the rest of this paper) are of two different natures. On one hand, structural information on the households gets captured. This includes the age of the main representative of the household ( $A_n$ ), the age of his/her partner ( $A_n^p$ ), marital status ( $M_n$ ), number of children ( $C_n$ ). Additional information is also recorded on existing bequests for fiscal reasons ( $\delta_n$ ). This later element was recorded as a dummy variable (i.e.  $\delta_n \in \{0, 1\}$ ). The median household was composed of a  $\mathbb{E}(A_n) = 53.4$  (+/-14/9) years old individual married to a  $\mathbb{E}(A_n^p) = 53.4$  (+/- 12.9) years old person with  $\mathbb{E}(C_n) = 1.6$  (+/-1.62) children. 181 households ( $\mathbb{E}(\delta_n) = 11\%$ ) have bequests in place.

On the other hand various data points are stored pertaining to households' wealth. This includes a detailed decomposition of the household's revenues, expenses, assets and loans.

Revenue was first analyzed according to three dimensions: the total revenue of the household ( $R_n$ ) [ $\mathbb{E}(R_n) = 63.3k/year$  (+/- 113K€)], the revenue generated by the main representative of the household ( $R_n^c$ ) (resp. his/her partner ( $R_n^p$ )) [ $\mathbb{E}(R_n^c) = 42.6k/year$  (+/- 96.6K€),  $\mathbb{E}(R_n^p) = 14.6k/year$  (+/- 32K€)]. The reason behind this split was that, intuitively, households' revenue levels (which is notoriously correlated to savings capability) as well as internal differences could play a role in their financial product subscription decisions. At a high level, the average profile shows that the households in the data-set were in the upper end of the French income distribution (top 10% according to the statistics recorded by the french state (I.N.S.E.E. <sup>8</sup>). Besides, this also shows that households presented a high level of discrepancy between partners in terms of revenue. This is not unusual as French records show that, on average the associated gap is of 22% <sup>9</sup>, but is perhaps exacerbated in the available sample as individuals are at the end of their career or already retired.

Revenue was also decomposed at the household level according to its source ( $R_n^s$  with  $s \in 1...15$ ). Sources were structured according to the referential in place in *Manymore*'s software. Details are displayed in table (6) and show that most of the revenue from

<sup>8</sup><https://www.insee.fr/fr/statistiques/5431993>

<sup>9</sup><https://www.insee.fr/fr/statistiques/6047789>

July 20, 2023

household was generated from standard wages and non commercial earnings, complemented by some pensions (probably associated to the fact that a non negligible portion of the data-set was composed of retired household), some level of income from stocks and shares and various allowances.

When it comes to expenses, assets and debts, information was captured both at an aggregated household level and according the expenses, assets, debts classification structure in place within *Manymore*'s software. In this sample, the average household exhibits a level of expenses of  $E_n = 15.4\text{k€}$  (+/-43k€) per year. On the overall sample about 43% of households were reported to have contracted a number of loan(s) of  $\mathbb{L}_x = 0.8$  (+/-1.5) associated to a charge of about 14.3k€ per year (+/- 166k€) [representing about 20 to 25% of households revenue]. Additionally, the average households owned  $\mathbb{A}_x = 7.1$  (+/-8.8) assets. The details of expenses and assets are displayed in tables (7), (8). Given that households had a very simple debt pattern (about 1 loan), loans were not classified according to specific local categories. Expenses, which represent, on average, about 20-25% of households revenue, are primarily composed of common expenses and income taxes. The average household had a total wealth in terms of assets worth 1.3M€ (+/-2.7M€), which was primarily allocated between a primary residence ( 300k€), some rental properties ( 200k€), life insurances contracts ( 130k€) and private equity ( 150-200k€).

### 3.2. Methods:

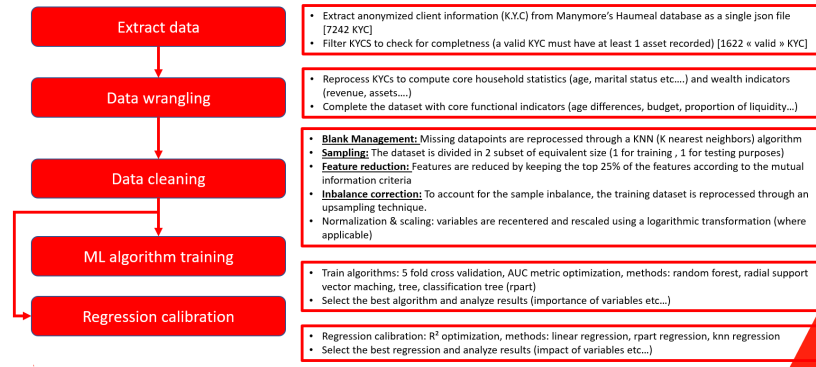
As seen in figure (1), three types of methodology were used in this article. First, standard feature selection techniques were leveraged to mine the data-set and extract valuable information which can be used to predict household's investment behavior (see subsection (3.2.1)). Second, standard classification algorithms were calibrated to the current data-set to understand if it could be feasible to predict whether or not a household would subscribe to a life insurance (LI) or a retirement plan (RP), or buy a real estate asset (ReA) over the next 12 months (see subsection (3.2.2)). Finally, regressions were used to evaluate if households' investments on these newly opened contracts or newly bought assets could be anticipated (see subsection (3.2.2)). Note that the aforementioned techniques were implemented using the Caret package (Kuhn, 2008) available on R.

#### 3.2.1. Feature selection techniques:

The amount of variables that can be extracted or generated (through combination) from the current data-set is important. Therefore feature selection is mandatory to avoid potential over-fitting problems and achieve the best prediction/classification performance possible (Jović et al., 2015). In light of the existing literature (Chandrashekar and Sahin, 2014), this gets nowadays performed through two main techniques: a correlation test (Guyon and Elisseeff, 2003) or a mutual information test (Battiti, 1994). In this case, both methods were tested and the mutual information yielded the best performance.

On top of feature selection, additional pre-processing tasks were completed notably to handle class imbalance on the classification problems at hand (i.e. is the households likely to subscribed to a LI,RP or to buy a ReA over the next 12 months?). This is indeed a common theme in machine learning (Chawla et al., 2004) which is known to lead sub-optimal performances when calibrating classification algorithms. Looking back at the

July 20, 2023



**Figure 1.** Methodology used to calibrate the proposed classification and regression methods.

statistics displayed in section (3.1), the current data-set and thus this study are no stranger to this phenomena. This is usually tackled through one of three core techniques (Provost, 2000). A first option consists in down-sampling the training data-set (i.e. re-sampling the majority class to make its frequency closer to the rarest class.). Another option consists in up-sampling the training data (i.e. replicating casing from the minority). The third option "ROSE" (Menardi and Torelli, 2014) consists in the generation of artificial data points. In order to find the most suitable alternative, all three methods were tested and the one yielding the best results was kept.

Note that additional alternatives exist (e.g. assigning weights to each class (Chen et al., 2004) or combining up- and down-sampling (Chawla et al., 2002)) but were not tested in this study.

### 3.2.2. Classification & regression methods:

As highlighted in the recent review of (Zhang et al., 2017), machine learning techniques dedicated to classification problems are numerous. Yet, two main ones appear to dominate the field, namely **random forests** (Breiman, 2001) [referred to as **RF** in this rest of this paper] and **support vector machines** [referred to as **SVM** in the rest of this paper] (Cortes and Vapnik, 1995). Given their prominence, those two algorithms were pre-selected for the purpose of this study. Note that in light of the non linearity of the classification problem (which was assessed through a Principal Component Analysis on the training data set), the chosen implementation of the SVM technique relied here on radial/exponential kernels. In addition to RF & SVM algorithms, an additional contender techniques was selected amongst the ones known to yield the best performance on real world problems (Zhang et al., 2017), namely conditional inference trees (Hothorn et al., 2015) [referred to as **CTREE** in the rest of this paper].

When it comes to regressions, three types of classical models (Ciaburro, 2018; Fernández-Delgado et al., 2019) were used to explore if the amount households are likely to invest in the next 12 months following the subscription of a LI or RP (resp. the amount invested in ReA) can easily be explained. First, standard linear regressions were used. Second, non linear techniques such as random forest regressions (Breiman, 2001) and CART regressions (Breiman, 2017) were also used. Additional techniques (for instance neural networks etc...) were also tested in this study, but their usage did not yield any

July 20, 2023

Model type	$R^2$
Linear regression	40.1%
CART	13.1%
Random Forest	32.1%

**Table 1.** Regression results - amount invested in life insurance by households over the past year

performance improvement compared to the one previously mentioned (see section (4)). Those attempts have therefore not been reported here.

#### 4. Results:

##### 4.1. Key learnings for life insurance (LI) products:

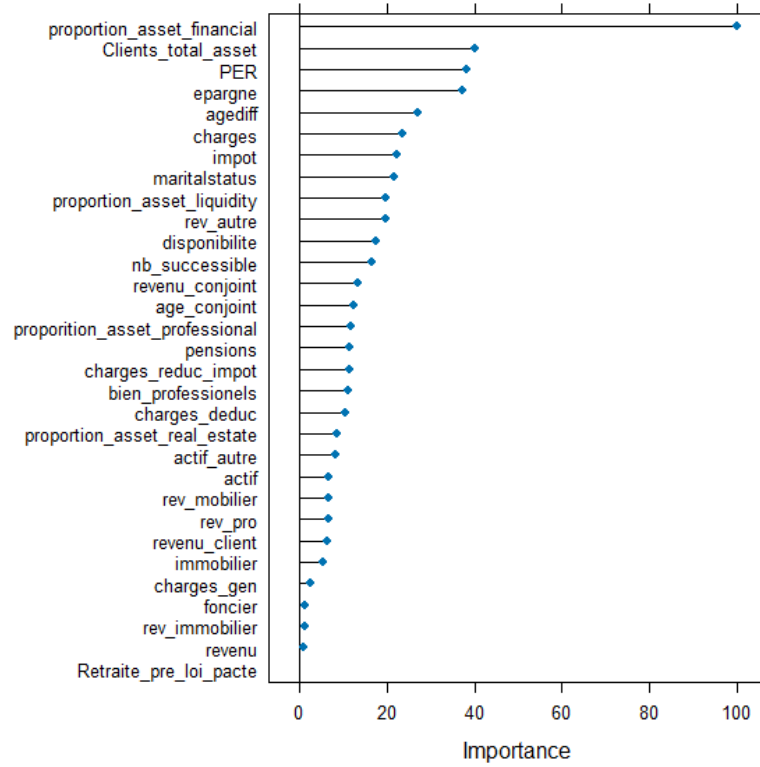
When it comes to predicting the amount invested by households in a newly subscribed life insurance product over the past year, the methods used in this article (linear regression, tree based model (CART) and random forest) did yield mixed results (see table 1). The available variables were indeed only able to explain 40% of the variance observed on the data-set with the best performing model here being the linear one. Looking at the input variables used to feed to the regression, there are 4 data points yielding a normalized importance above the 25% threshold (see figure ??). First, households' investments with respect to life insurance product appear to be primarily driven by the proportion of financial asset they hold. The more financial products they have, the more they are likely to invest in life insurance instruments. Second, wealthier clients ("client\_total\_asset") have higher savings levels ("epargne") and thus higher levels of financial investments. Third, the higher the amount invested in a retirement plan ("PER"), the higher the amount invested in life insurance products. Here the correlation is likely to be driven by the amount of savings the household is able to generate.

When it comes to predicting whether or not a household is likely to subscribe to a life insurance over the next 12 months, the proposed methods are efficient (A.U.C 80-85%). The selected algorithms yield relatively similar performances (see table (2)) on a balanced data-set of 320 households (class imbalance was handled through down sampling). As seen in figure (3), 7 variables have a normalized importance above 25% (when fed to the most performing algorithm). The pattern behind the algorithm can be summarized in the following fashion: the wealthier a household and the more it is inclined to contract financial products, the more likely they are to subscribe to a life insurance. Besides, the more liquidities they have, the higher the chance that they contract such a product. So here, statistical techniques simply replicate and industrialize professional knowledge available in the field.

##### 4.2. Key learnings for retirement plans (RP):

The current experiments do not yield statistically significant results when it comes to predicting the amount invested by households in a retirement plan in France over the past year. The methods used in this article (linear regression, tree based model (CART) and random forest) indeed did yield poor results with the current data-set ( $R^2 < 30\%$ ) (see table ??).

July 20, 2023



**Figure 2.** Importance of the variables used to predict the amount invested in life insurance by households over the past year (linear model).

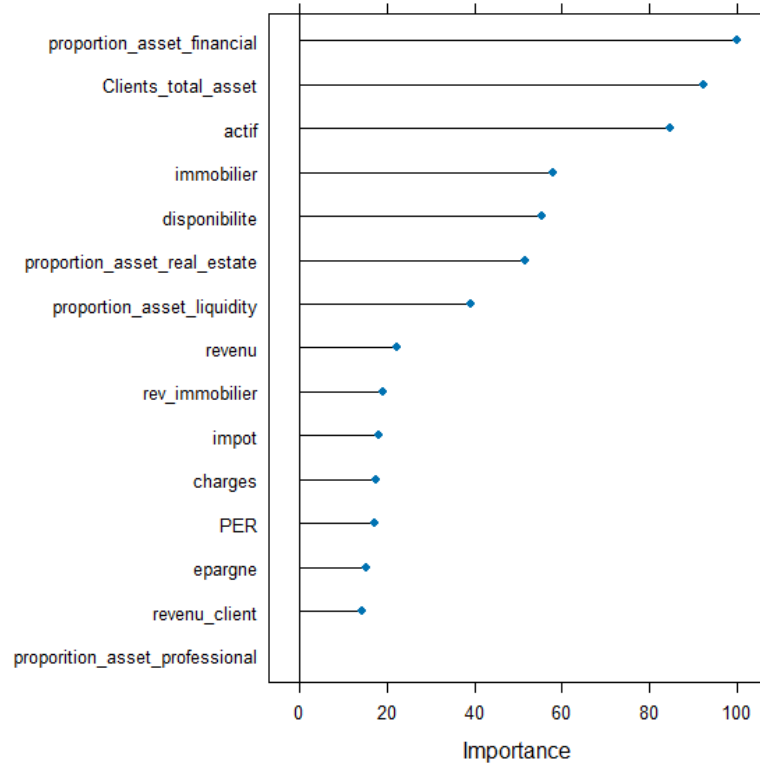
Model type	A.U.C
C Tree	81.9%
C Part	79.2%
SVM Radial	83.2%
Random Forest	83.4%

**Table 2.** Classification results (random forests) - Will a household subscribe to a life insurance over the next 12 months?

Model type	$R^2$
Linear regression	7.9%
CART	15.1%
Random Forest	26.0%

**Table 3.** Regression results - amount invested in private retirement plans by households over a 12 months period.

July 20, 2023



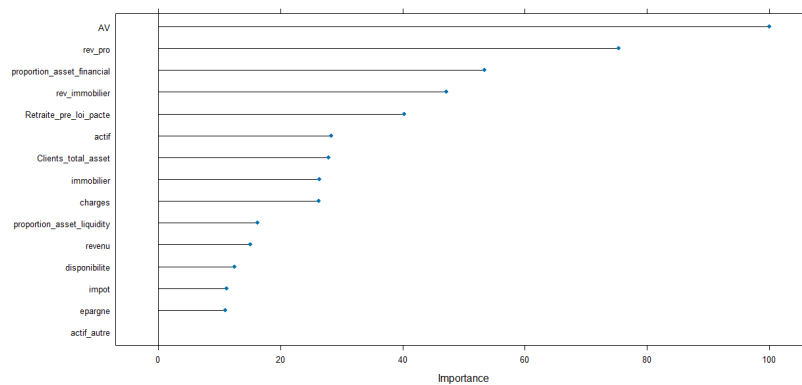
**Figure 3.** Importance of the variables used to predict whether or not a household is likely to subscribe to a life insurance over the next 12 months.

If predicting the amount invested by household to prepare privately for retirement appears challenging, the likelihood of contracting such a product over the next 12 months can easily be assessed (see table (4)). Standard machine learning methods indeed yield very good results on this type of problem with an A.U.C > 85% (post down-sampling to train the models on balanced data-set made of 96 observations). Analyzing the importance of the variables (see figure (4)) (i.e. random forest) shows that, for the best performing algorithm, the likelihood of opening a private retirement plan is highly linked to households' total revenue (from both professional activities or real estate properties) and existing financial market participation (amount invested in life insurance products, proportion of wealth invested in financial products etc...). Once again, statistical methods simply activate and industrialise professional knowledge materializing the fact that the higher households' revenue, the lower the replacement rate allocated by state sponsored public pensions scheme in France and the greater the incentive to capitalize and save through private mechanisms.

July 20, 2023

Model type	A.U.C
C Tree	86.6%
CART	82.4%
SVM Radial	83.6%
Random Forest	88.4%

**Table 4.** Classification results (random forests) - Will a household subscribe to a private insurance plan over the next 12 months?



**Figure 4.** Importance of the variables used to predict whether or not a household is likely to subscribe to a private retirement plan in France over the next 12 months.

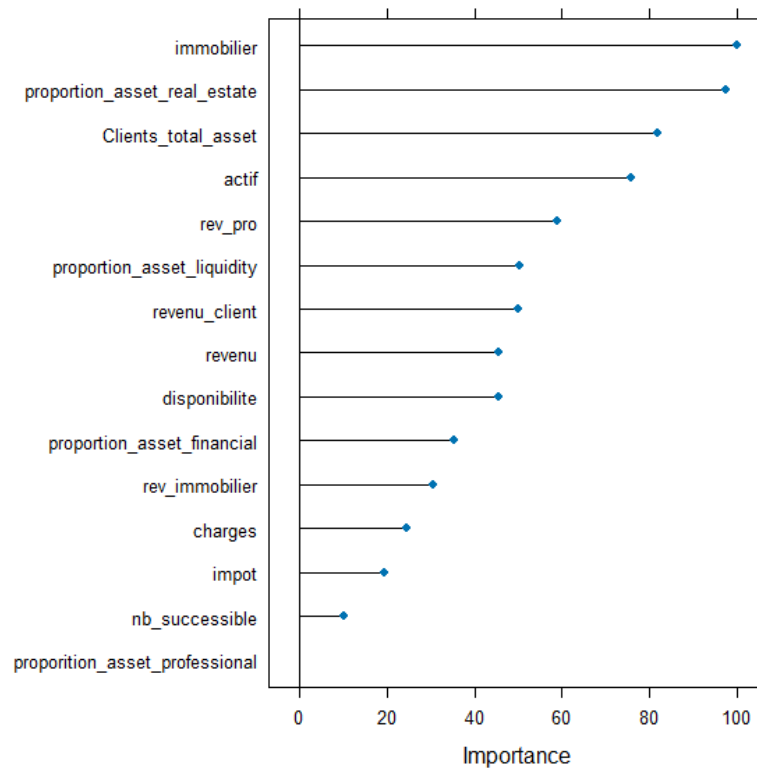
#### 4.3. Key learnings for real estate assets (ReA):

Investigating the profile of households buying real estate properties within the current data-set also yields mixed results. On one hand, it proves difficult to predict the amount invested by individuals. When calibrating linear regression or tree based models (e.g. random forest or ctree) to the available information, no statistical techniques was able to explain more than  $R^2 = 10\%$  of the inherent variance in the data. But, when it comes to detecting a statistical pattern to infer whether or not a household will buy a propriety over the next 12 months, results appears much more interesting as seen in table (5). Once the data-set has been normalized through down-sampling ( 100 data points), standard algorithms indeed show a "good" performance (expressed in terms of A.U.C) of about 80%. On this front, key variables used to yield a prediction revolve around the current worth of real estate assets owned by the household, the proportion of physical versus financial assets in households' portfolio as well as households revenue and savings levels (see figure (5)). This aligns with two stylized fact known within the financial advisor profession. First, real estate is considered as household's primary investment vehicle (as soon as they have some savings capability). Second, this type of investment vehicle is persistent. Rather than diversifying with financial products, households (which net worth is within the 500k€ to 1 - 2M€ as described within this data-set) indeed tend to accumulate wealth within physical assets over time.

July 20, 2023

Model type	A.U.C
C Tree	73.3%
CART	69.7%
SVM Radial	74.3%
Random Forest	78.9%

**Table 5.** Classification results - Will a household buy real estate over the next 12 months?



**Figure 5.** Importance of the variables used to predict whether or not a French household is likely to buy real estate over the next 12 months.

## 5. Discussion:

### 5.1. Limitations:

The current study presents a couple of limitations worth mentioning. First, the sample used to calibrate the proposed algorithms remains "small" and biased towards a certain social layer. It only covers 1622 households, which belong to the top 1% (income-wise) of the French economy. The households in the sample indeed showcase an average revenue of 113k\$ per year, while public sources <sup>10</sup> highlights that the median revenue of

<sup>10</sup><https://www.insee.fr/fr>

July 20, 2023

French households is close to 20 to 25k\$ per year. This comes with two consequences. On one hand, results displayed here are likely to be inappropriate when applied to the general population. The criteria driving the appetite for a financial product such as a retirement plan are very likely to differ between a household with 100+k\$ of revenue per year and a household with 20k\$ of annual revenue. On the other, there is natural question of portability of the results for the fringe of households in the top 1% of the economy. This fringe indeed contains about 300k households, of which the available sample covers about 0.5%. The data-set may therefore present some underlying biases, which could not be assessed prior to the study.

Second, the data used to fuel the study only encompasses French households. Despite being one of the largest economy in the world, France has several specific regulatory characteristics which may hinder the portability of the results and calibrated algorithms. France is known to have a very strong public system catering for retirement (in line with the European "Bismarkian" tradition) and benefits from one of the highest public replacement rate across O.E.C.D countries<sup>11</sup> when distributing pensions. This means that the elements driving French households to subscribe to private retirement plans are likely to be very different from the ones driving the decision of an American or a English household, even within the portion of the population with a very high level of income. There are however good chances that results would be structurally similar across Germany, Italy and Spain for similar population groups.

Third in France, retirement plans and life insurance products come with very specific tax incentives which are more than likely to affect the subscription behavior of households. In France, life insurances come with specific inheritance tax exemptions up to about 100k. This usually tend to dimension the major part of households investments, especially as 60% of the social fringe encompassed in this study tend to look for means to optimize its succession (redadd citation of prior work). A similar mechanism is also at play for retirement plans as deposits comes with a rebate on income taxes up to 20k per year. This again creates a strong incentive with respect to the size of the deposits, their frequency but only for households with a certain level of income. Those two specificities are inherently intertwined with the French tax system and when replicating the study in alternative geography, some attention should be paid to the associated context as it may impact the results (for instance the importance of the "tax" variables highlighted in table ??).

Finally, the application of standard algorithms on current data set does not present good predictive capabilities ( $R^2 < 30 - 40\%$ ) with respect to the level of investments performed by households on the core investments media (RE,LI,RP) available in France. Although more efforts could be invested in leveraging a broader array of techniques, the reason may very well be that the data-set in itself is insufficient. Deposits are indeed of two types: one off and recurring. One off deposits are usually triggered by specific events, for instance the perception of a lump sum because of an inheritance, and are often subject to constraints (for instance some distributors require an initial upfront deposit to open a contract). It is highly likely that those one off contextual elements are the key elements determining the level of investments over the first year of a contract. However, exploring this hypothesis would require a more granular data-set encompassing the history of operations (deposits, withdrawal, re-investments, portfolio re-balancing etc..) on contracts.

---

<sup>11</sup><https://www.oecd.org/>

July 20, 2023

## 5.2. Future research opportunities:

There are four streams of activities that could be undertaken to build upon the current research. First, additional research in the context of France could be useful. To address the topic of portability to other social layers, it could be interesting to expose the proposed algorithms to the public via a software platform used at scale, and thereby analyse a bigger volume of information. It could indeed be interesting to see if collecting data for 15 to 30k households with a "high earner" profile would cement the current findings or not. This could be easily done by deploying and re-calibrating the algorithms in a financial advice network with more 2000 counsellors (e.g. a large french bank).

Second, it would be interesting, still in the French context, to see how the proposed methods perform for households with a different income profile. The topic of retirement planning is indeed vibrant in O.E.C.D countries with Bismarkian retirement systems since replacement rates offered by public pensions are getting lower due to population ageing (Disney, 2000; Fenge and Werding, 2004). This in turns triggers a number of questions around the usage of private retirement plans and their democratization.

Third, additional research in the context of France would be necessary to understand the levels of investment displayed by households. As stressed at the end of the previous subsection (5.1), this would most likely require another type of data-set. This could be done in the context of France by leveraging anonymous information from private financial data aggregators such as *Harvest* or *Manymore* or by accessing anonymous information stemming from the back office a large financial product distribution network (for instance a large bank). This effort could however prove complex as it would require to blend two types of information. On one hand, it would require granular data on financial contracts, their operations and their ownership structure. On the other hand, it would need to have detailed information about the wealth of the contracts' owners (e.g. their level of revenue etc...). Data collection and reconciliation would therefore present a significant challenge.

In light of the considerations developed in this article, a last natural area for future research lies in the extension of the study to other countries. If the usual key markets (USA, UK, Germany, Spain, Italy, Japan) jump to mind, there a few things to consider. The level of participation of households to financial markets is high in the US in light of the prominence of private retirement systems and life insurance do not benefit from the same aura as in Europe. Differences are therefore expected on this front. When it comes to European countries, regulation differs between countries. The UK has notably implemented the latest Mifid regulations (change in terms of the compensation mediums associated to the distribution of financial products) and as a result has completely altered its local structure in terms of financial advice and product distribution. This is therefore likely to have an effect on the variables driving households investments behavior and decisions. As per the other European countries (Germany, Spain & Italy), local retirement systems, demographic pressure (notably for japan) and structural differences when it comes to the financial institutions are also more than likely to generate some heterogeneity in the proposed results.

## 6. Conclusion:

This article shows that simple information depicting households' wealth can be used to infer their subscription behavior towards standard financial products (i.e. life insurance

*July 20, 2023*

and retirement plans) as well as their purchasing intention towards real estate assets. This can be done using standard machine learning techniques such as random forest with a high level of performance (A.U.C >90%) for French households.

However the data-set used for this study proves insufficient to predict the exact level of investments of those households based on their macro level characteristics ( $R^2 < 30 - 40\%$ ). The considerations developed in this study suggests that accessing a more granular data-set (for instance, the past behavior of households in terms of deposits and/or withdrawals on already owned financial products) could be of use.

July 20, 2023

## 7. Appendix

Revenue sources	Average household revenue (k€/year)	Std deviation (k€/year)
Wages ( $R_n^1$ )	34.1	74.1
Non commercial earnings ( $R_n^2$ )	8.9	42.2
Commercial earnings ( $R_n^3$ )	1.7	13
Agricultural wages ( $R_n^4$ )	0.5	18.3
Pensions ( $R_n^5$ )	5.12	18.3
Non taxable pensions ( $R_n^6$ )	<0.1	0.9
Alimony ( $R_n^7$ )	<0.1	1
Life annuity ( $R_n^8$ )	<0.1	0.8
Income from stock & shares ( $R_n^9$ )	1.8	20.5
Movable income and distributions ( $R_n^{10}$ )	0.9	14
Other movable income ( $R_n^{11}$ )	<0.1	-
Tax free movable income ( $R_n^{12}$ )	<0.1	0.3
Gain from the sale of securities ( $R_n^{13}$ )	0.7	22.1
Non taxable capital gains ( $R_n^{14}$ )	<0.1	0.2
Gains from life insurance contracts ( $R_n^{15}$ )	<0.1	2.3
Revenue from real estate	1.1	7.5
Ordinary property income	-	-
Family allowances	0.5	20.2
Other revenues	0.9	15.1

**Table 6.** Households revenue decomposition in the *Haumeal* sample

July 20, 2023

Expense sources	Average household expenses (k€/year)	Std deviation (k€/year)
Common expenses	6.4	23.3
Education expenses	0.2	1.8
Exceptional expenses	0.1	1.5
Alimony	0.3	1.9
Hospitality expenses	<0.1	0.2
other deductible expenses	0.4	5.7
Employment of an employee at home	0.4	1.7
Childcare costs	<0.1	0.6
Elderly care expenses	<0.1	<0.1
Donations to organizations	<0.1	<0.1
Donation to charities	<0.1	0.4
Union dues	<0.1	<0.1
Compensatory allowance	<0.1	<0.1
Other tax deductible expenses	0.3	5
Direct investments in SME	<0.1	0.4
Investment in innovation funds	<0.1	0.1
Income tax	5.1	21.5
Property wealth tax	0.5	3.3
Residential tax	0.2	1.2
Property tax	0.3	1.3
Other taxes	0.2	2.3

**Table 7.** Households expenses decomposition in the *Haumeal* sample

## References

- (2017). Discussion paper on the eba's approach to financial technology (fintech). *European Banking Authority, EBA*.
- Alhenawi, Y. et al. (2013). Financial literacy of us households: Knowledge vs. long-term financial planning. *Financial services review 22*.
- Anagnostopoulos, I. (2018). Fintech and regtech: Impact on regulators and banks. *Journal of Economics and Business 100*, 7–25.
- Armour, J., D. Awrey, P. L. Davies, L. Enriques, J. N. Gordon, C. P. Mayer, and J. Payne (2016). *Principles of financial regulation*. Oxford University Press.
- Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016). International comparative household finance.
- Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on neural networks 5*(4), 537–550.
- Bazot, G. (2018). Financial consumption and the cost of finance: Measuring financial efficiency in europe (1950–2007). *Journal of the European Economic Association 16*(1), 123–160.
- Bengen, W. P. (1994). Determining withdrawal rates using historical data. *Journal of Financial planning 7*(4), 171–180.
- Bernheim, B. D., A. Shleifer, and L. H. Summers (1986). The strategic bequest motive. *Journal of labor Economics 4*(3, Part 2), S151–S182.

July 20, 2023

Asset nature	Average worth (M€)	Std deviation (M€)
Main residence	0.31	0.56
Secondary residence	0.09	0.75
Rental property (rented bare)	0.18	0.62
Furnished rented property (French LMP status)	0.02	0.15
Share of real estate civil society	0.07	0.63
Share of real estate investment firm	0.01	0.04
Long-term rented rural properties	0.01	0.18
Land (not leased)	0.01	0.07
Current accounts	0.05	0.19
Passbook accounts	0.01	0.09
French "A" accounts	0.01	0.04
Others accounts	0.01	0.07
Term accounts	0.02	0.73
Savings housing plan (French "PEL")	0.01	0.02
Capitalization bonds	0.02	0.33
Business savings plan (French "PEE")	0.01	0.03
Securities account	0.02	0.14
European & french shares	0.01	0.24
French PEA investment accounts	0.01	0.07
Private equity (French "Pacte Dutreil")	0.01	0.24
Life insurance	0.13	0.49
Retirement account (French PERP) 53	0.01	0.03
Private equity	0.17	0.95
Private equity (french "entreprise individuelle")	0.01	0.16
Goodwill	0.02	0.43
Other professional assets	0.01	0.12
Other rental properties (outside of the French LMP)	0.02	0.27
Other assets	0.01	0.06

**Table 8.** Households assets decomposition in the *Haumeal* sample

- Bogan, V. (2008). Stock market participation and the internet. *Journal of Financial and Quantitative Analysis* 43(1), 191–211.
- Breiman, L. (2001). Random forests. *Machine learning* 45, 5–32.
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Buckley, R. P., D. W. Arner, D. A. Zetsche, and R. H. Weber (2020). The road to regtech: the (astonishing) example of the european union. *Journal of Banking Regulation* 21(1), 26–36.
- Campbell, J. Y. (2006). Household finance. *The journal of finance* 61(4), 1553–1604.
- Chandrashekar, G. and F. Sahin (2014). A survey on feature selection methods. *Computers & Electrical Engineering* 40(1), 16–28.
- Chatterjee, S. and J. E. Grable (2022). 34 the future of personal finance: An educational and research agenda. *De Gruyter Handbook of Personal Finance*, 599.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16, 321–357.

July 20, 2023

- Chawla, N. V., N. Japkowicz, and A. Kotcz (2004). Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter* 6(1), 1–6.
- Chen, C., A. Liaw, L. Breiman, et al. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley* 110(1-12), 24.
- Christelis, D., D. Georgarakos, and M. Haliassos (2013). Differences in portfolios across countries: Economic environment versus household characteristics. *Review of Economics and Statistics* 95(1), 220–236.
- Ciaburro, G. (2018). *Regression Analysis with R: Design and develop statistical nodes to identify unique relationships within data at scale*. Packt Publishing Ltd.
- Cocco, J. F. (2005). Portfolio choice in the presence of housing. *The Review of Financial Studies* 18(2), 535–567.
- Collardi, B. F. (2012). *Private banking: Building a culture of excellence*. John Wiley & Sons.
- Cortes, C. and V. Vapnik (1995). Support-vector networks. *Machine learning* 20(3), 273–297.
- Cull, M. (2022). 30 the growing role of fintech and robo-advisors. *De Gruyter Handbook of Personal Finance*, 529.
- David, H., L. F. Katz, and M. S. Kearney (2006). The polarization of the us labor market. *American economic review* 96(2), 189–194.
- De Nardi, M. and G. Fella (2017). Saving and wealth inequality. *Review of Economic Dynamics* 26, 280–300.
- De Nardi, M. and F. Yang (2016). Wealth inequality, family background, and estate taxation. *Journal of Monetary Economics* 77, 130–145.
- Disney, R. (2000). Declining public pensions in an era of demographic ageing: Will private provision fill the gap? *European Economic Review* 44(4-6), 957–973.
- Fenge, R. and M. Werding (2004). Ageing and the tax implied in public pension schemes: simulations for selected oecd countries. *Fiscal Studies* 25(2), 159–200.
- Fernández-Delgado, M., M. S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, and M. Febrero-Bande (2019). An extensive experimental survey of regression methods. *Neural Networks* 111, 11–34.
- Fisch, J. E., M. Labouré, and J. A. Turner (2019). The emergence of the robo-advisor. *The Disruptive Impact of FinTech on Retirement Systems* 13.
- Foerster, S., J. T. Linnainmaa, B. T. Melzer, and A. Previtro (2017). Retail financial advice: does one size fit all? *The Journal of Finance* 72(4), 1441–1482.
- Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pp. 20–37. Princeton University Press.
- GAUDECKER, H.-M. V. (2015). How does household portfolio diversification vary with financial literacy and financial advice? *The Journal of Finance* 70(2), 489–507.
- Grable, J. E. and S. Chatterjee (2022). *De Gruyter Handbook of Personal Finance*. Walter de Gruyter GmbH & Co KG.
- Grossman, G. M. and E. Rossi-Hansberg (2008). Trading tasks: A simple theory of offshoring. *American Economic Review* 98(5), 1978–97.
- Guiso, L. and T. Jappelli (2005). Awareness and stock market participation. *Review of Finance* 9(4), 537–567.
- Guyon, I. and A. Elisseeff (2003). An introduction to variable and feature selection. *Journal of machine learning research* 3(Mar), 1157–1182.

July 20, 2023

- Hammond, R. C. (2015). Behavioral finance: Its history and its future.
- Harrigan, J., A. Reshef, and F. Toubal (2016). The march of the techies: Technology, trade, and job polarization in france, 1994-2007. Technical report, National Bureau of Economic Research.
- Hirshleifer, D. (2015). Behavioral finance. *Annual Review of Financial Economics* 7, 133–159.
- Hothorn, T., K. Hornik, and A. Zeileis (2015). ctree: Conditional inference trees. *The comprehensive R archive network* 8.
- Huggett, M. and G. Ventura (2000). Understanding why high income households save more than low income households. *Journal of Monetary Economics* 45(2), 361–397.
- Jović, A., K. Brkić, and N. Bogunović (2015). A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 1200–1205. Ieee.
- Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of us households. *Journal of Economic Behavior & Organization* 27(1), 87–96.
- Knewton, H. S. and Z. A. Rosenbaum (2020). Toward understanding fintech and its industry. *Managerial Finance* 46(8), 1043–1060.
- Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of statistical software* 28, 1–26.
- Lyons, A. C. and J. Kass-Hanna (2022). 25 behavioral economics and financial decision making. *De Gruyter Handbook of Personal Finance*, 433.
- Masson, A. and P. Pestieau (1997). Bequests motives and models of inheritance: a survey of the literature. *Is inheritance legitimate?*, 54–88.
- Maude, D. (2010). *Global private banking and wealth management: the new realities*, Volume 610. John Wiley & Sons.
- Menardi, G. and N. Torelli (2014). Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery* 28(1), 92–122.
- Merkoulova, Y. and C. Veld (2022). Why do individuals not participate in the stock market? *International Review of Financial Analysis* 83, 102292.
- Murguia, A. and W. D. Pfau (2021). A model approach to selecting a personalized retirement income strategy. Available at SSRN 3788232.
- Mutual, N. (2017). Planning & progress study 2016.
- Navaretti, G. B., G. Calzolari, J. M. Mansilla-Fernandez, and A. F. Pozzolo (2018). Fintech and banking. friends or foes? *Friends or Foes*.
- Peijnenburg, K., T. Nijman, and B. J. Werker (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control* 70, 18–35.
- Pfau, W. D. (2010). An international perspective on safe withdrawal rates from retirement savings: The demise of the 4 percent rule? *Journal of Financial Planning, Forthcoming*.
- Philippon, T. (2016). The fintech opportunity. Technical report, National Bureau of Economic Research.
- Philippon, T. (2022). Harnessing the promise of fintech. *Shifting Paradigms: Growth, Finance, Jobs, and Inequality in the Digital Economy*, 95.
- Phoon, K. and F. Koh (2017). Robo-advisors and wealth management. *The Journal of Alternative Investments* 20(3), 79–94.

July 20, 2023

- Polansky, S., P. Chandler, and G. R. Mottola (2019). The big spenddown: Digital investment. *The Disruptive Impact of FinTech on Retirement Systems*, 129.
- Provost, F. (2000). Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets*, Volume 68, pp. 1–3. AAAI Press.
- Ribes, E. (2022). Using classification techniques to accelerate client discovery: a case study for wealth management services.
- Ribes, E. A. (2023). Transforming personal finance thanks to artificial intelligence: myth or reality? *Financial Economics Letters* 2(1), 11–21.
- Rossi, A. G. and S. P. Utkus (2020). The needs and wants in financial advice: Human versus robo-advising. Available at SSRN 3759041.
- Tilmes, R. and P. Schaubach (2006). Private banking und private wealth management—definitionen und abgrenzungen aus wissenschaftlicher sicht. *Private Banking und Wealth Management, Bankakademie-Verlag GmbH, Frankfurt am Main*.
- Todd, T. M. and M. C. Seay (2020). Financial attributes, financial behaviors, financial-advisor-use beliefs, and investing characteristics associated with having used a robo-advisor. *Financial Planning Review* 3(3), e1104.
- Vives, X. (2017). The impact of fintech on banking. *European Economy* (2), 97–105.
- Warmath, D. (2022). 27 measuring and applying financial literacy. *De Gruyter Handbook of Personal Finance*, 473.
- West, J. (2012). Financial advisor participation rates and low net worth investors. *Journal of Financial Services Marketing* 17(1), 50–66.
- Wilhelm, M. O. (1996). Bequest behavior and the effect of heirs' earnings: Testing the altruistic model of bequests. *The American economic review*, 874–892.
- Wisman, J. D. (2009). Household saving, class identity, and conspicuous consumption. *Journal of Economic Issues* 43(1), 89–114.
- Zhang, C., C. Liu, X. Zhang, and G. Almpandis (2017). An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications* 82, 128–150.