



**HAL**  
open science

## The need for data products in personal finance.

Edouard Ribes

► **To cite this version:**

Edouard Ribes. The need for data products in personal finance.: A perspective on financial instruments distribution chains with an application to the French investment eco-system.. 2023. hal-04015599

**HAL Id: hal-04015599**

**<https://hal.science/hal-04015599>**

Preprint submitted on 16 Mar 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.

# The need for data products in personal finance.

A perspective on financial instruments distribution chains with an application to the French investment eco-system.

Edouard Augustin Ribes<sup>1</sup>

<sup>1</sup>CERNA, Mines Paristech\*

March 16, 2023

**Background context:** Current societal challenges around healthcare, education and retirement require households to increasingly leverage personal finance instruments. To meet this trends the lending, insurance and investment industries need to become more efficient and affordable.

**Specific knowledge gap the work aims to fill:** To date, the distribution chains of financial instruments remains costly and inefficient. To transform, the associated industries need to further leverage digital media to accelerate products distribution and maintenance. Some of the benefits of digitalization have already been captured & depicted in the recent literature sitting at the frontier between personal finance and financial technologies. However the scope of those studies has, so far, been limited to the distribution of those instruments & there has been little discussion about the opportunities associated to the maintenance of financial contracts, notably through the structuration of data products & warehouses. This is a gap this article aims to address.

**Methods used in the study:** This paper leverage standard economic modeling techniques as well as option and game theory to describe the impact of digital media and notably data products on the financial instruments' brokerage system. It also leverages order of magnitude found in the literature to perform a high level calibration of those models to one of the Big 5 European financial market, namely the French investment industry.

**Key findings:** The proposed models show 3 stylized facts about data

---

\*Contact: edouard.augustin.ribes@gmail.fr

products when applied to the French investment industry. First, such a market can only support two data product suppliers. Second, it comes with a large asymmetry in prices (prices differ by a factor 2 or 4 between actors) and clients profiles between the two data suppliers. Third, the market is not completely efficient as its equilibrium results in about 60% of the market not being equipped with data products.

**Implications:** Data products can yield a 10 to 20% productivity increase for independent financial advisors and brokers distributing financial instruments. Those gains will likely be passed in some form to households, thereby increasing the overall efficiency of the financial system and supporting households financial professionalization.

**Keywords.** Personal finance, households economics, wealth, technological change, financial services.

**JEL.** D31,L84,O33.

# 1 Introduction:

**Context:** Today's mature countries are facing three profound societal challenges. First, education systems are ongoing a deep transformation as the value human capital is unwrapped. Second, the healthcare world is also in the midst of a (re)volution as technological advances mixed with demographic shifts are generating tensions. Finally retirement systems are also ongoing changes as population ageing requires adjustments to neoclassical Bismarkian systems (Fanti and Gori, 2012; Fanti et al., 2015; Ribes, 2022b).

Blatant differences aside, there is one thing that all of those of phenomena have in common: they all require individuals to increasingly take charge of themselves across the entirety of their life-cycle (see (Ando and Modigliani, 1963) for the underlying theory) as states'/public involvement is reduced. This, in turn, calls for a better access to and knowledge across all pillars of personal finance (i.e. payments, loans, insurances, investments products) (Grable and Chatterjee, 2022). But all of those pillars yet face numerous challenges.

Let us first consider the lending industry. It is divided in two segments (Zinman, 2015). On one hand, we can find mid to long terms loans including mortgages (50% of the mid to long term loans) as well as car (30%) and student loans (19%). On the other, we can find short terms loans (which have a penetration of 60% in the US & the UK and with various levels of usage across European countries) which include credit cards and consumption credits. Challenges to date here are that households generally over pay for debt (across mature countries, 50% of the loans have a 7%+ interest rate) and that the credit scoring & attribution process associated to loans distribution remains lengthy & costly. One can thus easily imagine the implication of those costs and miss-matches when considering the need for financing unexpected healthcare expenses, education bills or even more simply, basic housing etc...

When it comes to the access to investments instruments, most mature market still face heavy inter-mediation costs (2 to 3%) (Philippon, 2016; Bazot, 2018) which strongly affects the value they yield (the average return is of 5% whilst inflation has remained, on average, as high as 1 to 2% over the past decades). Given that those instruments are used across the entirety of households life-cycle (Jappelli and Modigliani, 1998)(notably to prepare for retirement), there is a real need to increase the efficiency of the system to make it useful. When considering an average length of service of 40 years in the labor market, an increase (or decrease) of 1% on the return of capital has very important implications for individual households... Besides, the investment industry comes nowadays with additional challenges in terms of access as the associated instruments are mainly leveraged by households with a high level of income (and therefore savings capability) (Huggett and Ventura, 2000).

Insurances, on their end, are used to cover four core risks incurred by households (Grable and Kruger, 2022): risks pertaining to one's health (diseases, the need for long term care and potential disabilities), unemployment risks, death (via life insurances) and risks pertaining to one's properties (mostly home & vehicles). Even though the industry has reached a form of maturity as the associated markets appear saturated (Grable and Chatterjee, 2022)], it still faces challenges of two natures. First, competition drives in-

insurance firms to look for opportunities to decrease their costs (mostly linked to claims processing, regulatory reporting & fraud detection) which progressively translates in lower fees for households (King et al., 2021). Second, technological development comes with opportunities to better assess the risks incurred by households as more data gets generated and collected in today's world (thereby reducing the traditional problem of information asymmetry underlying the insurance industry) (Boobier, 2016; Cull, 2022). All of those improvements should ultimately lead to households having a more appropriate value for money when they contact such products. Those savings could then be used for other purposes such as retirement savings.

The only segment of personal finance which seems to experience a different spin to date lies within the payment side of the industry. The current ecosystem is indeed made of digital mediums (debit & credit cards) whose costs (with fees worth about 1-2% of each transaction (Humphrey et al., 2003)) are mostly incurred by retailers and of physical mediums (cash & checks) whose costs are supported by households. To date, the most efficient media are electronic and they already have a high level of penetration: about 80% in Europe (60% in the US). The core challenges (Pieters, 2021) now are about ensuring that physical payments forms are still available where appropriate (notably in rural areas), while digital payments systems are becoming less vulnerable so that households data is fully protected and remains private. So here, the next wave of transformation is not so much about lowering costs to help households finance their retirement as well as their health and human capital, but rather about safety and control.

**Problem:** In short, the need for more personal finance underpins the need for a continuous transformation of the existing eco-system in order to increase its efficiency and affordability. This evolution is already occurring as changes are seen on the distribution chains associated to loans, insurances investments instruments (i.e. the classical brokerage and financial advice activities) (see (Vives, 2017; Navaretti et al., 2018; Suryono et al., 2020) for recent reviews or (EBE, 2017) for institutional reports). Those changes are of two nature.

First, "regulatory" activities (subscription, contractual reviews & amendments etc...) are increasingly digital, thereby increasing the productivity of brokers and advisors. This comes with a number of business opportunities for digital entrants (e.g. "Fintechs"), which were described in a prior work (Ribes, 2021). Second, there is an increasing need for a centralization and standardisation of information related to loans, insurances and investments vehicles (i.e. the structuration of specialized 'data products'). Brokers and advisors indeed work with a large catalog of instruments and manage an increasing complex and decentralized portfolio <sup>1</sup>. For them to operate efficiently, there is therefore a need for the definition of data standards and for the set up of aggregation capabilities, which have not been, to my knowledge, the subject of any discussion so far (Takeda and Ito, 2021). This is a gap that this article aims to bridge.

---

<sup>1</sup>Note that similar products could also be used to help financial instruments providers (e.g. an insurance company) shape better instruments (e.g. data collection on households to help assess their risk when it comes to insurance design). However, this is not something which will be discussed here

Note that the development of data products should naturally pave the way for the introduction of advanced analytical tools (leveraging for instance artificial intelligence techniques) enabling a better match between households and the instruments available within the financial sphere (see (Ribes, 2022a) for a quick discussion).

**Structure:** To understand the need for such data products, this paper presents a bundle of models describing the value of those products for brokers and advisors who operate either in an independent fashion (section 2) or within large networks (section 3). Each section discusses the core productivity levers at play for those segments & the value proposal of a data product suppliers. It also models the equilibrium which exists in a local market. Note that the proposed models are generic enough to enable an assessment of the value of a new entrant or existing actor providing data products independently of its specialisation (loans, insurances, investments) [provided that the productivity of a distributor has been modeled].

Each section also features a series of examples, which consists in applying the proposed models to various use cases. If those use cases have been, for personal convenience, centered around typical situations observed in the investment pillar of the personal finance industry, they could be easily extended to cover other activities.

## 2 Modeling the value of "retail" data products:

### 2.1 The rationale behind "retail" data products:

#### 2.1.1 The productivity of an independent financial advisor/broker:

The main purpose of independent financial advisors [I.F.A] and brokers nowadays is to distribute financial instruments (loans, insurance, investments) to households and then to ensure their maintenance. Instruments are generally provided to a portfolio of  $N$  clients. Now, the way productivity is articulated for independent brokers & financial advisors varies across each pillar of the personal finance industry and across countries (because of variations in local regulations). The general pattern consists in brokers & financial advisors being compensated either through commissions paid by the financial instrument provider which instrument got distributed &/or through one time fees incurred by the household the instrument was distributed to (see for example (Guillem, 2015; Ragin and Halek, 2016) for a discussion on the insurance industry). Now, for the sake of simplicity, we will only consider markets where the dominant compensation scheme consists in commissions. Besides, in line with the original aim of the paper (i.e. to illustrate what is happening on the investments side of the personal finance industry), this article will focus on a small model explaining the commission scheme sustaining the distribution of investment instruments. Note that a similar exercise could easily be performed for either loans or insurance instruments.

In the investment side of personal finance, what matters is the capital  $C(t_n)$  stored by each households  $n \in \{1, \dots, N\}$ . The accumulated capital depends in the length of time  $t_n$  during which they have been performing investments activities. Households indeed

generally invest a lump sum  $I$  at every point in time and then benefit from a return  $\mu \approx 5\%$  for their investments (Bazot, 2018). This behavior can be described as follows:

$$C(t_n + 1) = C(t_n) \cdot (1 + \mu) + I \leftrightarrow C(t_n) = ((1 + \mu)^{t_n} - 1) \cdot \frac{I}{\mu} \quad (1)$$

Households invest during a time horizon of  $\theta \approx 15$  years as they usually as they reach 50 years of age and capitalize until retirement which occurs when they turn about 65 years old. They then start to decumulate usually through withdrawal  $W$  until their death, which occurs around 80 years of age. The decumulation and accumulation periods are roughly similar in duration. Assuming that households tends to withdraw amounts in a fashion similar to an annuity, they end up withdrawing  $W = \frac{I \cdot ((1 + \mu)^\theta - 1)}{1 - (1 + \mu)^{-\theta}}$ <sup>2</sup>. For the sake of simplicity, let us assume that an advisor distribute products to  $d \approx 5$  new households per year and that they all have the same generic capitalization mechanism. Given that the advisor generates revenue through retro commission worth  $\phi \approx 2 - 3\%$  percent (Philippon, 2016) of the total amount of asset they manage, their activities generates a revenue of:

$$R(t) = \frac{\phi \cdot d \cdot I}{\mu} \left( \underbrace{\sum_{k=0}^{\min(\theta, t)} ((1 + \mu)^k - 1)}_{\text{Revenue from capitalizing households}} + \overbrace{\frac{((1 + \mu)^\theta - 1) \cdot \sum_{k=0}^{\min(\theta, t - \theta)} (1 - (1 + \mu)^{k - \theta})}{1 - (1 + \mu)^{-\theta}}}_{\text{Revenue from decumulating households}} \right) \quad (2)$$

Since an advisor/ broker works about  $\Psi = 1680$  hours per year (i.e. 210 working days of 8 hours)<sup>3</sup>, this means that their hourly productivity  $\rho$  increases over time (as per standard economic theory - see (Anzanello and Fogliatto, 2011) for a review) in the following fashion:

$$\rho(t) = \frac{\phi \cdot d \cdot I}{\mu \cdot \Psi} \left( \frac{((1 + \mu)^{\min(\theta, t) + 1} - 1)}{\mu} - \min(\theta, t) + \frac{(1 + \mu)^\theta - 1}{1 - (1 + \mu)^{-\theta}} \cdot (\min(\theta, \max(t - \theta; 0)) - (1 + \mu)^{-\theta} \cdot \frac{(1 + \mu)^{\min(\theta, \max(t - \theta; 0))} - 1}{\mu}) \right) \quad (3)$$

This also implies that the volume of households  $N(t)$  supported by a financial advisor is a function of his/her experience (i.e.  $N(t) = d \cdot \min(t, 2 \cdot \theta)$ ). Given that advisors & brokers generally support a portfolio of  $\approx 100$  households in the investment pillar of the personal finance industry (Foerster et al., 2017; Linnainmaa et al., 2021) to which  $\approx 2$  financial capitalization instruments are distributed (Linnainmaa et al., 2021), it can be assumed that a broker distribute  $d \approx \frac{100 \cdot 2}{2 \cdot \theta} = 6 - 7$  new contracts per year.

<sup>2</sup>When households decumulate by withdrawing fixed amounts on their savings, their capital follows  $\forall t \geq \theta; C(t) = (1 + \mu) \cdot C(t - 1) - W$ . When they calibrate their withdrawals to consume all their capital over a fixed time horizon  $\theta$  (i.e.  $C(2 \cdot \theta) = 0$ ), it comes that  $W = \frac{\mu \cdot C(\theta)}{1 - (1 + \mu)^{-\theta}}$ . The fact that  $C(\theta) = ((1 + \mu)^\theta - 1) \cdot \frac{I}{\mu}$  according to equation (1) then yields the suggested decumulation program  $W$ .

<sup>3</sup>This of course varies depending in local labor regulations. The proposed estimate is based on what can be found in the top 5 countries across the European Union.

### 2.1.2 "Retail" data products as productivity enhancers:

To date, most of the available time of independent financial advisors and brokers is spent on maintenance activities (Maude, 2010). This entails reviewing their stock of  $N(t)$  contracts at least  $M$  times a year. To do so, they must consolidate information about each contract, which are provided by  $L$  different counterparts (insurers, banks, asset managers...). Scanning through those counterparts initially entails a search time  $\tau \in [1/60, \dots, 1/6]$  hours per contract as the information is not centralized.

Here, the primary goals of data products is to completely streamline those search activities. This means that, in essence, "retail" data products are reporting ones. For data products covering  $l \in \{1 \dots L\}$  providers, this translates into a productivity boost of  $\rho(t) \cdot \tau \cdot M \cdot N(t) \cdot \frac{l}{L}$  as advisors are able to refocus their time on client acquisition & relationship management activities.

Data products are made by centralizing information from  $l \in \{1 \dots L\}$  sources into one repository with a unified model. Given the current state of digital technologies, consolidation costs can be considered linear (i.e. production costs are worth  $c \cdot l$  per unit of time for  $l$  sources). Besides given the distributed nature of digital products, as long as there is one supplier, the demand from all the advisors who can benefit from a productivity boost for a price tag  $p$  per financial instrument for which data gets collected and  $l$  sources can be met.

## 2.2 Market structure for "retail" data products:

### 2.2.1 A monopoly on data products:

Let us first assume that data products are only supplied by a unique actor with a data catalog of  $l$  sources and that the distribution of advisors/brokers according to their experience level is depicted by a function  $f(t)$ . In this case, to maximize the benefits of data products for financial advisor & brokers, the data product unit price tag  $p^*$  is dependent in the size of the supplier's catalog and, given that suppliers are profit maximizing agents, price is given as a solution of the following constrained problem :

$$\begin{cases} \max_{(p^*)} (p^* \cdot \sum_{t \geq t^*} (N(t) f(t))) & (\text{Supplier profit maximization behavior}) \\ p^* \cdot \sum_{t \geq t^*} (N(t) f(t)) \geq c \cdot l & (\text{Supplier profitability constraint}) \\ t^* = \min_{t \in \{1 \dots T\}} (\rho(t) \cdot \tau \cdot M \cdot l \geq p^* \cdot L) & (\text{Advisor productivity gain constraint}) \end{cases} \quad (4)$$

This type of problem can easily be solved in a "greedy algorithm fashion" via a grid search on the price  $p^*$  variable. The productivity constraint indeed leads to a price which is in the  $[0; \rho(T) \cdot \tau \cdot M \cdot \frac{l}{L}]$  range. Therefore setting up a grid with a granularity of a few cents enables a simple search for the monopoly set up.

**Example.** Let us consider a market for investments products similar to the French one. In France, there are about 680 financial products providers, 60% of which provide



products to households (i.e.  $L \approx 400$  with different data models <sup>4</sup>. Assume that an independent financial advisor spends  $t \approx 1$  min(s) searching for information on a financial product, where households invest, on average,  $I \approx 12\text{k€}$  per year <sup>5</sup>. The market is structured such that  $F \approx 5000$  independent financial advisors are distributed evenly according to their experience and retire after  $t_{max} = 35$  years (i.e.  $F(t) = F \cdot \min(\frac{t}{t_{max}}; 1)$ ). From a technology standpoint, the costs of adding and maintaining an entry to a data catalog is assumed to represent 4 days of work for a data engineer, which nowadays represents a cost <sup>6</sup> of  $c \approx 2\text{k€}/\text{year}$  in a mature market.

In this set up, assume that there is only one data provider with a data catalog of length  $l = 250$  sources. Solving problem (4) entails that the provider will equip all the advisors with an experience  $t > t^* = 21$  years for a yearly fee  $p^* = 5.3\text{€}/\text{year}/\text{contract}$ . The firm will thus have a retail market share of 40% (i.e. 60% of the advisors will not be equipped), generate an overall yearly revenue of 1.5M€ with a margin of 60%.

Looking at the advisor's side of the market, the orders of magnitude used in this example lead to a productivity curve of a sigmoid nature. Their first 10 to 15 years of activity is marked by an exponential progression towards a productivity of about 100k€/year, followed by a slower growth to a plateau of about 400k€/year. This plateau is reached after 20 years of activity. In this set up, data products are mostly targeted towards advisors who have reached this level of maturity and their adoption translates into a productivity boosts for advisors of 2 to 3%.

Additionally, a sensitivity analysis was run based on the size of the catalog (i.e.  $l$ ). Results displayed in figures (1a) and (1b) shows that for a data supplier to exist its catalog must reach the minimal critical size of 15 sources <sup>7</sup>. Second it shows that prices  $p^*$  linearly increase with the size of the supplier's catalog and range between 1 and 7€/contract/year. Third, it shows that the maximum amount of revenue a firm can generate with such an offering is of about 2M€ per year on a market such as the French one.

### 2.2.2 Multiple data suppliers:

Let us assume now that "retail" data products are supplied by  $J$  actors indexed by  $j \in \{1; \dots; J\}$ . Actors are sorted by the size of their data catalog (i.e.  $\forall j_1 < j_2; l_{j_1} < l_{j_2}$ ). Advisors can now choose between different data products: an independent financial advisor with  $t$  years of experience is therefore comparing the value proposal of each data supplier  $j$  (expressed as  $\rho(t) \cdot \tau \cdot M \cdot l_j - p_j^* \cdot L$ ) and chooses the supplier  $j^*$  who maximises its gains.

In this case, the market is set as a  $J$  player game, where each data supplier can choose a level of price  $p_j$  and where advisors of experience  $t$  select the supplier which maximises

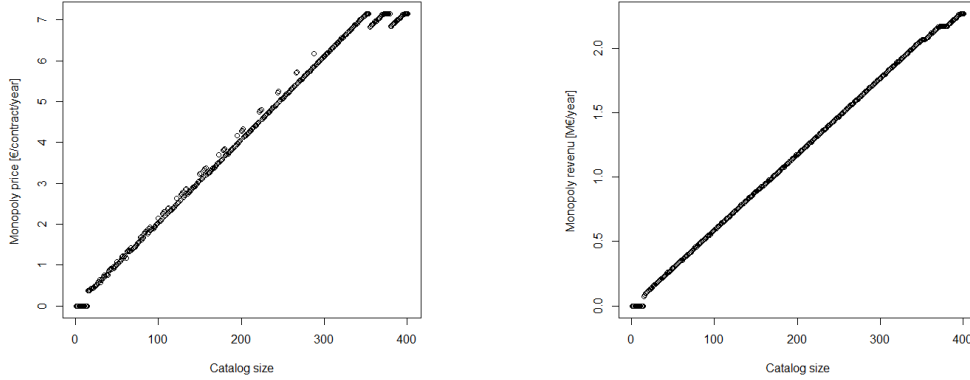
---

<sup>4</sup>the latest report of the [AFG](#)

<sup>5</sup>This represents a saving rate of 10 to 15% for a household with an income of 100k€ per year

<sup>6</sup>This represents the costs of having 8 incidents per data feed per year (or about 1 every month without holiday). Each incident requires the intervention of a resource costing about 500€ per day for about 0.5 day.

<sup>7</sup>This interestingly recoups the long standing idea that for firm to be viable they must reach a minimum efficient size ([Simon and Bonini, 1958](#))



(a) Monopoly products prices with catalog size.

(b) Monopoly yearly revenue

Figure 1: Data suppliers economic results in a monopoly set up.

their productivity gains:

$$\max_{j \in \{1 \dots J\}} (\rho(t) \cdot \tau \cdot M \cdot \frac{l_j}{L} - p_j) \quad (5)$$

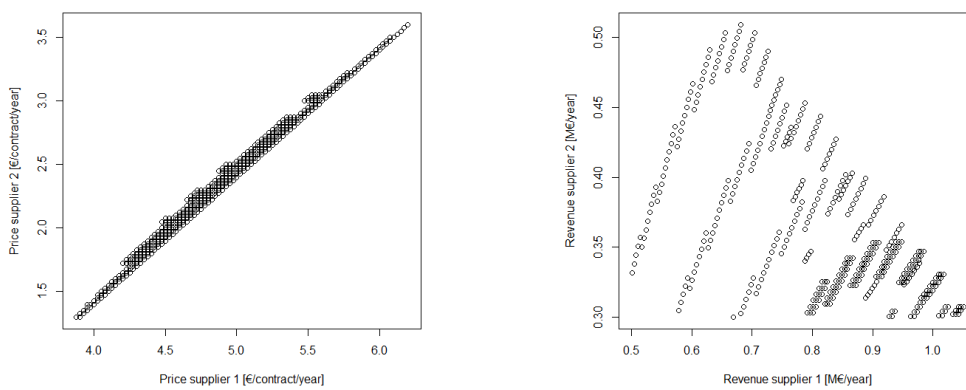
When the number of actor is relatively small <sup>8</sup>, this kind of game can be discussed by exploring & discussing the possible choices made by data suppliers through a greedy search approach. To do so, the state of possible prices  $p_j$  for each actor  $j$  is depicted by a vector  $\{0, \dots, \rho(T) \cdot \tau \cdot M \cdot \frac{l_j}{L}\}$ , where possible prices differ by a step  $\Delta p = 0.01$  c€. Then, for each possible price combination  $(p_1, \dots, p_J)$ , the advisors allocation  $A : t \in \{0 \dots T\} \rightarrow j \in \{1 \dots J\}$  is drafted as a solution of (5) given the data suppliers price choices. Note that this allocation can account for different suppliers having the same value proposition for an advisor/broker. In this case, the pool of advisors/brokers of experience  $t$  is divided equally amongst the suppliers having the same proposal. Financial results are then derived for data suppliers in terms of revenue and profitability. Should the position be non profitable for an actor  $j$ , its revenue and profits are normalized to 0.

**Example.** Let us reconsider the example of the French investment market presented in section (2.2.1). But, this time, let us assume that there is a second competitor presenting a data catalog of size  $l_2 = 150$  sources. In this case, two options appear. First, the main data supplier can assume that its competitor will attempt to maximize its revenue and given this knowledge will try to find a price which maximizes its own revenue in a non cooperative fashion. Given that the main data supplier has a larger catalog and therefore a competitive advantage, this leads to a price  $p_1^* = 2.3$ €/contract/year a yearly revenue of  $\approx 0.7$ M€ with a margin of 27%. In this set up, the competition is such that the second actor can not exist and 60% of all advisors are equipped with a data product ( $t^*1 = 14$  years). For the main data supplier, entering a perfect competition results in a heavy

<sup>8</sup>This is generally the case in the "Fintech" market where there is only a few actors.

loss compared to the monopoly scenario described in section (2.2.1), although this comes with a better coverage of data products on the financial distributor side.

The second scenario is similar to a cooperation. In this context, each data supplier attempts to find a position where it can exist. The associated results are displayed in figures (2a) and (2b). In this case, if the suppliers cooperate to extract the maximum of revenue (and therefore profit since costs are linear in the size of their catalog) from the market, the main data supplier (resp. its competitor) will choose a price tag of  $p_1^* = 5.5\text{€}/\text{contract}/\text{year}$  (resp.  $p_2^* = 3\text{€}/\text{contract}/\text{year}$ ), yield a yearly revenue of 1M€ (resp. 0.3M€) with a margin of 52% (resp. 1.5%). The main competitor equips 35% of the advisors ( $t_1^* = 23$  years), while the second equips an extra 5% ( $t_2^* = 21$  years). Coverage is therefore relatively similar to the monopoly scenario with a minor fringe of the financial brokers and advisors benefiting from the introduction of the second actor.



(a) Data suppliers products prices.

(b) Data suppliers yearly revenue

Figure 2: Data suppliers options in a cooperative market set up.

Note that in this specific example, additional tests were run to assess whether or not adding a third data supplier with a catalog  $l_3 \leq 30$  sources would result in a different cooperative market set up. The associated analysis did show that there was no place for such a third entrant.

### 3 The value of data products dedicated to large distribution networks:

#### 3.1 Data products value proposition for large distribution networks:

If financial advisors and brokers can distribute financial products independently by setting up their own practice, a number of them are yet integrated as part of large distribution networks. In this case, the productivity of an advisor/broker  $\bar{\rho}$  is no longer a function of his/her own experience, but becomes normalized across professionals and

depends in the size of the network  $s$  to which they belong (i.e. the number of financial advisors/brokers in the network).

The market segment made by large distribution networks is also very different from the "retail" environment. First, those networks have specific ties with  $\bar{L}$  selected financial providers. As a result, the competition among the  $J$  data products' suppliers is no longer based on the size of their catalog since they have to adjust their offerings to meet the network needs. Second, the size of those networks calls for bespoke support and infrastructure which are both completely correlated to their size  $s$ . As a result, the total cost incurred by data suppliers are made of three parts:

- a production cost  $c.L$  associated to the supply of a custom catalog composed of  $L$  sources (the unit price of adding a data source is similar to what is observed on the "retail" market and is purely technology-dependent).
- an infrastructure cost  $c_I^j.s.N.K$ , which is dependent in the number of financial contracts which must be centralized ( $s.N.K$ ), given that storage and computing costs per contract is comes at a price of  $c_I^j$  (which can be specific to the technology choices made by the  $j$ -th data supplier).
- a support cost  $c_S^j.L.U$ , which is dependent on the number of sources which must be centralized ( $L$ ), the average number of incidents  $U$  stemming from a data feed associated to a financial products' provider and the cost  $c_S^j$  of support associated to the diagnostic and resolution of such an incident (which is also specific to the technology & organisational choices made by the  $j$ -th data supplier).

Finally, the way large distribution networks choose their data supplier also differs from the one seen in the "retail" segment as those networks consider that their choice creates a binding relationship with the supplier. Their evaluation therefore tends to be structured as a tender where the value proposal  $V_j(s)$  of the  $j$ -th data supplier for a network of size  $s$  includes a notion of robustness, which can be represented by the data product supplier default/survival probability  $\theta_j$ . Given that the  $j$ -th data suppliers comes with a price tag  $p_j^*(s)$  in the tender, its value proposition is summarized as:

$$\begin{cases} V_j = (1 - \theta_j).(\bar{\rho}(s).s.\tau.N.K.M - p_j^*(s)) \\ p_j^*(s) \geq (c + c_S^j.U).L + (c_I^j).s.N.K \end{cases} \quad (\text{Supplier's profitability constraint}) \quad (6)$$

## 3.2 Market structure:

### 3.2.1 Large networks facing a monopoly on data supply:

The underlying mechanism of a tender is to maximize the benefits of the data product for a potential client of size  $s$ , which translates into:

$$\max_{p_j^*} (V_j(s)) = \max_{p_j^*} ((1 - \theta_j).(\bar{\rho}(s).s.\tau.N.K.M - p_j^*(s))) \quad (7)$$

**Lemma 1:** *When there is only one data supplier on the market (i.e.  $J = 1$ ), the tender structure implies that data products will be priced "at cost". The associated price tag  $p_{monopoly}^*$  will thus have the following structure:*

$$p_{monopoly}^*(s) = (c + c_S^1 \cdot U) \cdot L + (c_I^1) \cdot s \cdot N \cdot K \quad (8)$$

The lemma (1) is interesting as it shows that the price of data products for those large distribution networks can be very dynamic depending in their growth. As a results prices articulated a tender appear best expressed when indexed to the size of the potential client.

As a quick remark, the situation described here is a bit hypothetical. In practice, in a monopoly situation, the tender process will result in a negotiation as the data supplier will elaborate a pricing scheme following a "cost plus model". The data supplier will leverage the "at costs" price  $p_{monopoly}^*(s)$  to which a margin will be added and the said margin will be the subject of a negotiation.

**Example.** Let us consider the case described in section (2.2.1) (i.e.  $N \approx 100$ ,  $K \approx 2$ ,  $M \approx 2$ ,  $c \approx 4\text{k€}$  per year) for a tender with  $\bar{L} = 10$  data sources. Let us assume that the unit costs associated to the infrastructure of data is mainly driven by storage costs and is worth  $c_I \approx 0.03\text{€}/\text{contract}/\text{year}$ . Respectively, let us assume that the probability of a data feed generates about  $U \approx 15$  defects per year and that solving the defect requires 2 to 3 hour of work for a data engineer (i.e.  $c_S \approx 0.3\text{k€}/\text{defect}$ ). The resulting prices offered by a data supplier in a tender process are displayed in figure (3). The results

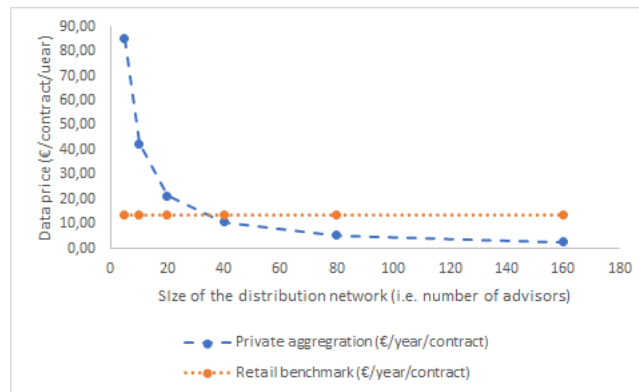


Figure 3: Tender prices offered by a data supplier in a monopoly set up.

display two interesting features. First, for a distributor who leverages  $L \approx 10$  sources, it becomes interesting to enter in a tender process to get a bespoke service as soon as its size becomes greater than  $s \approx 40$  advisors. For financial investments products distributors, this means that tenders becomes of interest as soon as the distributors generated more than 5M€ of revenue per year, which corresponds to an A.U.M (asset under management) of about 250M€. Second, it shows that prices converge slowly to be aligned at a data element level (e.g. a retirement plan distributed to a household) with infrastructure costs  $c_I$  to reach a few c€ per year.

### 3.2.2 Multiple data suppliers:

Let us now assume that multiple data suppliers (i.e.  $J > 1$ ) compete in a tender and attempt to put forward the best value proposal for a potential client of size  $s$ . Infrastructure costs are relatively normalized nowadays as data suppliers leverage cloud providers to host their solution and those cloud providers have a relatively similar cost footprint  $c_I$  (i.e.  $c_I^1 \approx c_I^2 \approx \dots \approx c_I^J \approx c_I$ ). As a result, the competition between data suppliers mostly revolve around the support model (incl. the associated technology) they dedicate to ensure the availability of their data feeds <sup>9</sup>.

**Lemma 2:** *Without prior knowledge or belief with respect to the competition and/or its technology, data suppliers prices are indexed on their production costs (i.e.  $p_j^*(s) = (c + c_S^j.U).L + c_I.s.N.K$ ). The tender selection process then end up selecting the supplier  $j^*$  with the best long term production profile:*

$$\max_{j \in \{1 \dots J\}} ((1 - \theta_j).(\bar{\rho}(s).M.\tau - c_I).s.N.K - (c + c_S^j.U).L) \quad (9)$$

Note that should the support model of suppliers be similar, the tender process will end up favoring the incumbent with the highest survival probability. Given that age and size are a known first proxy to assess firm survival (Evans, 1987), this should lead the client of size  $s$  to choose the oldest and /or the largest competitor.

Additionally, the tender process may not necessarily involve all of the suppliers. Tenders indeed start by an identification phase to pre-select some suppliers based on references, market share and/or prior work. Depending on how this phase gets played, the tender may result in a sub optimal choice. This will however not be further discussed in this paper.

**Example.** Let us consider a market similar to (2.2.2) where the two suppliers production structures are aligned with (3.2.1). The core difference lies in the age of the suppliers. The first one is the oldest and largest and has been present on the market for about 10+ years, while the second has only existed for a few years. Given that the first three to five years of existence of a firms are marked by a progressive increase and convergence in the firm survival probability (see for instance the OECD database)) towards a market standard, the largest actor will naturally be favored in the process.

## 4 Discussion & conclusion:

**Conclusion:** This paper models the value of data products sold to financial intermediaries (i.e. financial advisors and brokers) operating across three of the main pillars of

---

<sup>9</sup>Note that an additional element which could also come into play is the service offerings which are peripheral to the data product at stake in the tender (financial distribution network indeed tend to adopt a one stop shop approach in a tender ), but this is not going to be a topic of discussion in this article.

personal finance (insurances, loans & investment). It also provides examples & applications of the proposed theory to the investment arms of this eco-system. It notably highlights the differences in the underlying mechanisms driving value in the "retail" market segment (composed of independent brokers & advisors) and the market dedicated to large financial inter-mediation networks.

The methods described in this article are mostly applied to the French investment products distribution market (one of the largest of its kind in Europe). The models show 3 stylized facts when applied to the French "retail" environment. First, such a market can only support two data product suppliers. Second, it comes with a large asymmetry in prices (prices differ by a factor 2 or 4 between actors) and clients profiles between the two data suppliers. Third, the market is not completely efficient as the best option for data suppliers results in about 60% of the market not being equipped with a data products. Application to the French "large distribution network" market segment, then highlights that the value of data products is network dependent and that its price follows the production costs of and that prices display a decreasing structure. Estimates also shows that "large networks" and their associated tenders start to sprout when the size of the network reaches a threshold of about  $s \approx 40$  professionals.

**Opening remarks:** The modeling exercise described in this article could be improved in several ways. First, the matching mechanism used to depict the "retail" market structure could be enhanced to account for heterogeneity between financial products distributors as well as heterogeneity between households behaviors. This could be done, for instance, through the introduction of a degree in variability in individual professionals productivity. Second, the model could be improved to account for a limited usage of data catalogs. Professionals could indeed leverage only a random number of sources rather than the full portfolio, which could, in turn, have an impact on data products value proposals. Third, the scope of the applications could be enriched to account for flat fees as well as extended to other pillars of the industry (i.e. loans & insurance) to create a 360° view of the personal finance industry. Finally, when looking at the tender mechanism, it could be interesting to further decompose the identification phase used to prepare the said tender. This phase appears indeed often critical to ensure success for a data supplier, especially in niche environments where competition is limited and production technologies are normalized.

## References

- (2017). Discussion paper on the eba’s approach to financial technology (fintech). *European Banking Authority, EBA*.
- Ando, A. and F. Modigliani (1963). The” life cycle” hypothesis of saving: Aggregate implications and tests. *The American economic review* 53(1), 55–84.
- Anzanello, M. J. and F. S. Fogliatto (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics* 41(5), 573–583.
- 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.
- Boobier, T. (2016). *Analytics for insurance: The real business of Big Data*. John Wiley & Sons.
- Cull, M. (2022). 30 the growing role of fintech and robo-advisors. In *De Gruyter Handbook of Personal Finance*, pp. 529–550. De Gruyter.
- Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The journal of industrial economics*, 567–581.
- Fanti, L. et al. (2015). Growth, payg pension systems crisis and mandatory age of retirement. *Economics Bulletin* 35(2), 1160–1167.
- Fanti, L. and L. Gori (2012). Fertility and payg pensions in the overlapping generations model. *Journal of Population Economics* 25(3), 955–961.
- 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.
- Grable, J. E. and S. Chatterjee (2022). *De Gruyter Handbook of Personal Finance*. Walter de Gruyter GmbH & Co KG.
- Grable, J. E. and M. Kruger (2022). 19 the role of insurance as a household financial management tool. *De Gruyter Handbook of Personal Finance* 2, 307.
- Guillem, M. Á. L. (2015). Shedding the light on insurance brokers’ remuneration: The role of the fees. *Opción: Revista de Ciencias Humanas y Sociales* (2), 654–676.
- 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.
- Humphrey, D., M. Willeson, T. Lindblom, and G. Bergendahl (2003). What does it cost to make a payment? *Review of Network Economics* 2(2).



- Jappelli, T. and F. Modigliani (1998). The age-saving profile and the life-cycle hypothesis. *Long-run Growth and Short-run Stabilization: Essays in Memory of Albert Ando*.
- King, M. R., P. D. Timms, and T. H. Rubin (2021). Use of big data in insurance. *The Palgrave Handbook of Technological Finance*, 669–700.
- Linnainmaa, J. T., B. T. Melzer, and A. Previtro (2021). The misguided beliefs of financial advisors. *The Journal of Finance* 76(2), 587–621.
- Maude, D. (2010). *Global private banking and wealth management: the new realities*. John Wiley & Sons.
- Navaretti, G. B., G. Calzolari, J. M. Mansilla-Fernandez, and A. F. Pozzolo (2018). Fintech and banking. friends or foes? *Friends or Foes*.
- Philippon, T. (2016). The fintech opportunity. Technical report, National Bureau of Economic Research.
- Pieters, G. (2021). Digital currencies and central banks. *The Palgrave Handbook of Technological Finance*, 139–160.
- Ragin, M. A. and M. Halek (2016). Market expectations following catastrophes: An examination of insurance broker returns. *Journal of risk and insurance* 83(4), 849–876.
- Ribes, E. (2021). What is the impact of introducing productivity tools for wealth management professionals? a case study for the french market.
- Ribes, E. (2022a). Transforming personal finance thanks to artificial intelligence: myth or reality?
- Ribes, E. (2022b). What are the financial implications of an ageing population for european citizens?
- Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *The American economic review* 48(4), 607–617.
- Suryono, R. R., I. Budi, and B. Purwandari (2020). Challenges and trends of financial technology (fintech): a systematic literature review. *Information* 11(12), 590.
- Takeda, A. and Y. Ito (2021). A review of fintech research. *International Journal of Technology Management* 86(1), 67–88.
- Vives, X. (2017). The impact of fintech on banking. *European Economy* (2), 97–105.
- Zinman, J. (2015). Household debt: Facts, puzzles, theories, and policies. *economics* 7(1), 251–276.