# Models of Wealth and Inequality Using Fiscal Microdata: Distribution in Spain from 2015 to 2020

Ignacio González García [†] and Alfonso Mateos Caballero *,[†]

Artificial Intelligence Department, E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid, 28660 Madrid, Spain; igmigonzalezgarcia@gmail.com
* Correspondence: amateos@fi.upm.es
† These authors contributed equally to this work.

**Abstract:** In this research, we used Spanish wealth distribution microdata for the period 2015–2020 to provide a general framework for comparing different models and explaining different empirical datasets related to wealth distribution. We present a methodology to output the current value of assets and participations held by the population in order to calculate their real and current distribution. We propose a new methodology for mixture analysis, whereby we identify and analyze subpopulations and then go on to study their influence on wealth distribution. We use concepts of symmetry to identify two internal processes that are characteristic of the wealth accumulation process for the subpopulations of entrepreneurs and non-entrepreneurs. Finally, we propose a method to adjust these results to other empirical data in other countries and periods, providing a methodology for comparing results output with differing data granularity.

**Keywords:** wealth; distribution; Pareto; Gibrat; econophysics; Benford; assets

## 1. Introduction

The discipline of econophysics originated during the 1970s and has evolved significantly over a period of just 50 years [1–4]. From the very first, it was effectively applied to economic problems, such as the puzzle of excess volatility, where other consolidated theories such as dynamic stochastic general equilibrium had proved insufficient ([5], p. 2), and strange regularities as observed in the distribution of the number of workers of companies [6,7] or the study of wealth and income.

Economists' traditional interest in wealth distribution resurged in the 1970s. Some studied the question from the normative point of view [8,9], and many other pragmatists considered its effects on growth [10]. With the turn of the century, the interest in inequality [11,12] extended the focus beyond the academic sphere, renewing econophysicists' interest in the problem [13,14].

As is well-known, Pareto observed that the distribution of the number of individuals with income above a specified level of wealth presented a strange regularity. In 1953, Champernowe [15] proposed a model according to which a taxpayer's annual income was a function of the preceding year's income and a stochastic factor. Gibrat [16] used data on company size as a variable to study the proportional effects leading to lognormal distributions. This was the first attempt to explain this specific skewed pattern of distribution in stochastic terms. He showed that the cumulative effect of a succession of independent random shocks would generate a variable whose log would be uniformly distributed regardless of the distribution that governed the shocks. The idea is sometimes explained by saying that company growth is independent of company size.

These ideas inspired countless empirical studies [17–20]. Some tried to confirm Pareto's law with a better fit [21–24], and others attempted disproof by counterexample [25]. It has been consistently found in $ln - ln$ plots that, when drawing the survival curve,

the complement to the cumulative curve, 90–95% of incomes (above a threshold) can be fitted to a lognormal distribution, whereas the 5–10% of the richest fit a Pareto distribution.

The cause of the duality is an enigma. According to [26], the upper tail of the income distribution has long been regarded as a source of fascination for economists (for a recent review, see [27]).

Scholars have focused on three aspects:

1.  Distributions. Many researchers have tried to achieve a comprehensive unique and universal fit of the observed density or distribution functions with: biparametric (lognormal, gamma, Weibull) distributions, triparametric (generalized gamma, Singh-Maddala or Dagum) distributions [28], and pentaparametric (generalized beta) distributions [29]. Economists have preferred lognormal distributions [30], while statisticians [31] and econophysicists [32] have shown the utility of gamma distributions. Gibbs' distributions have been used to fit cumulative data.
2.  Models. They have been investigated by scholars who believe that it is plausible for a unique simple physical law or a stochastic model to explain the observed regularity. For a complete review of the models, see [18].
3.  Subpopulations. Some studies, interested in the design of effective policies, have attempted to adjust the density function as a mixture of component subpopulations of wage earners, pensioners, etc. [33].

Many questions remain open today:

1.  What are the underlying causes of the distribution?
2.  How is it possible that such skewed distributions are observed in competitive capitalist markets that should be efficient?
3.  How many different groups are there within the society? [34,35].
4.  What are the processes governing the constitution of existing subgroups, and their influence on wealth and income distributions?

In this research we have had the opportunity to process microdata, which provide a very detailed picture of wealth in Spain over the period 2015–2020. The aim was to:

1.  Determine which of the income and wealth variables best determines the form of a universal law of wealth distribution provided there is access to microdata.
2.  Identify the subgroups existing in a society whose mix determines the wealth distribution and find the distribution of each subgroup in the case of Spain.
3.  Model the wealth generation processes of the subgroups.
4.  Quantify the deviations and errors that occur in estimating income and wealth caused by the involuntary use of variables with partial or biased information.
5.  Explain the emergence of the shape of the upper tail of the wealth distribution.

This paper has four sections. In Section 2, we describe the methodology with special attention to the method used to obtain the indirect wealth using microdata of tax declaration. In Section 3, we apply the methodology to the data of the wealth distribution in Spain during the period 2015–2020. We present the results of the method used to identify the mixture of subpopulations, to explain their effects in the observable distribution using the concept of symmetry. In Section 4, we resume the results obtained.

## 2. Methodology

The study of wealth and inequality has traditionally come up against a systematic scarcity of detailed data, because: (i) statistical data collection is designed according to the objectives of each survey, and it is not easy to adapt the collected data items for other purposes; (ii) tax data are subject to special protection and are not easily accessible for academia; (iii) wealth is hidden; (iv) the value of wealth components is volatile; (v) some wealth is associated with legal persons, like companies, and the task of associating such assets with their holders is complex, and, last but not least, (vi) considerable computational power is required to process all the data on all forms of wealth.

The methodology used in this research is divided into four phases: (1) data collection; (2) variable selection; (3) decision on data binarization; and (4) analysis.

1.  *Data collection.* We used detailed data from the declarations presented to the Spanish State Tax Agency (AEAT): taxpayer census, information submitted by public bodies (e.g., cadastre), tax (personal income tax, corporate tax, inheritance tax) self-assessments, as well as informative declarations submitted by insurance companies and financial institutions (the authors are solely responsible for their conclusions, which have not been reviewed by AEAT).

    While some authors [26] have already used income tax microdata, this study increases the level of detail. We selected the variables of interest from many tax declarations, including, for example, the gross tax base from Income Tax Form 100; data on partners and % participation in other companies from Corporate Tax Form 200, and data on assets and liabilities from other forms, such as F 720, F 174, F 196, F 714 and F 189. All the forms and variables used are described on the AEAT web page (AEAT, 2020).

    The number of taxpayer declarations in an average year (2018) include 19.98 million for income tax, 1.59 million for corporate tax and 207,225 for inheritance tax. A total of 94.7 million current accounts, and loans worth around 139.5 million euros are processed annually, using several hundreds of millions of data items sourced from the Spanish cadastre and from the informative declarations provided by financial institutions. These data are included in the AEAT annual reports [36].

    We have studied three variables:

    (a)  *Income.* Gross tax base, declared on Form 100.
    (b)  *Assets.* This is the sum of direct assets (value of property, financial products, deposits, insurance policies, shares in listed companies, etc.) and indirect (or corporate) assets, that is, the value of the % participation in the assets of non-listed companies.
    (c)  *Net worth* ($NW$). Assets minus passives liabilities (loans and accounts payable) is the sum of two components $DNW$ (direct net worth) and $INW$ (indirect net worth).

        For each inhabitant, indexed by $i$, $i \in \{1, \cdots, N\}$, ($N \approx 59{,}73$ M),

        $$NW_i = \sum_{j \in DA_i} DNW_{ij} + \sum_{k \in C_i} INW_{ik}, \tag{1}$$

        where $DA_i$ is the set of direct assets belong to inhabitant $i$ and $C_i$ is the set of companies in which inhabitant $i$ has a share.

    The calculation of the components of indirect wealth is complex because taxpayers may or may not to have participations in non-listed companies, which may or may not hold shares in other companies and so on. It is calculated according to two different processes explained using the toy example in Figure 1 with three natural persons and six companies {N, A, B, C, D, E}.

    Following the example of Natural Person 1, we calculate the direct wealth of taxpayers from the declarations presented by the taxpayer and by others (e.g., financial institutions). Indirect wealth is calculated by accumulating the market value of participations in non-listed companies {A, B, C, D} that are downstream in the flow of relations between participations. In this example, Natural Person 2 is a shareholder of A only. The relationship between company A, Natural Person 1 and Natural Person 2 is declared by A on page 2 of Form 200. The relationship between A and {B, C} is declared by A on page 24 of Form 200 and includes the % participation and its value. This information can be crossed with the information provided by B and C.
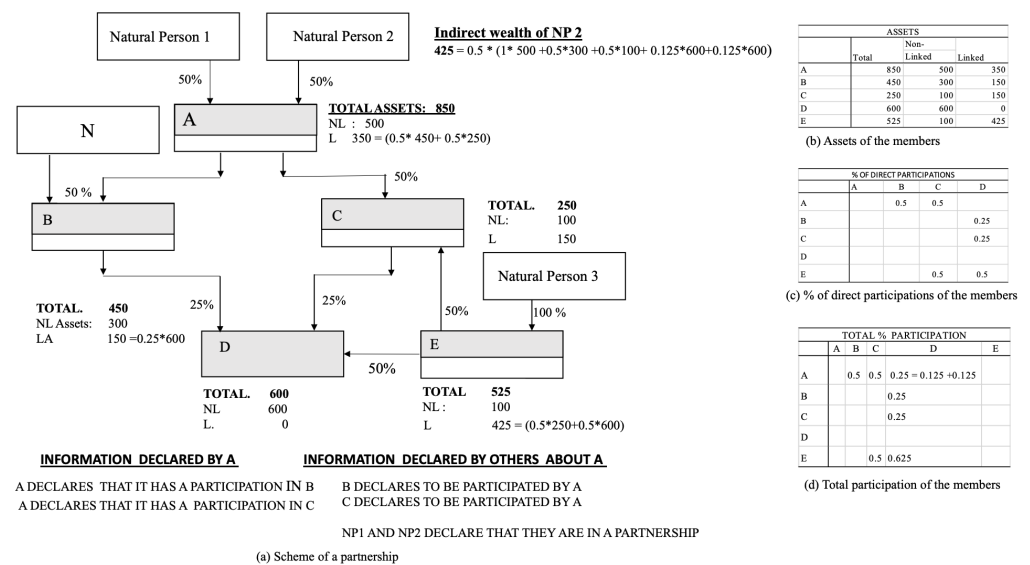
**Figure 1.** Example of calculation of indirect wealth.

The process is divided into two phases:

(a) *Calculate the % participation of each taxpayer in each company.* This is a network-based calculation using the information provided by companies on page 2 and 24 of Form 200. For example, the % participation of Natural Person 1 in D is output by adding two components along paths A-B-D and A-C-D with a total of 0.25, as illustrated in Figure 1d.

(b) *Calculate the target variable (assets or net worth).* We used balance-sheet data declared on Form 200 to output, for each network node, two set of values: (a) linked assets and liabilities; (b) non-linked assets and liabilities (that is, the balance-sheet items that appear due to company participation in other companies) as illustrated in Figure 1 for the example. Now,

$$INW_i = \sum_{c \in C_i} \sum_{p \in P_{ic}} \prod_{l \in p} \%PAR_l NLNW_l, \tag{2}$$

where $C_i$ is the set of companies in which inhabitant $i$ has a share, $P_{ic}$ is the set of all the paths from $i$ to $c$ and $l$ is a company along the path $p$.

The indirect net worth of a taxpayer's participation in a company is the sum, for the companies, indexed by $c$, in which it has a share, and for all the paths, indexed by $p$, that connect the taxpayer with a company, of the product of % participations $PAR$ multiplied by the non-linked net worth ($NLNW$) of these participations in the $L$ companies along a path. The non-linked assets (referred to in Spanish as "no vinculados") are calculated using data from Forms 200 and 220, including the balance sheet items. A close approximation to the linked value can be calculated from the balance sheet by deducting the value declared under items 00118, 00153, 00160 referring to assets and items 00223, 00238, 00243 referring to group and partner companies from the company assets. It was decided that the value of a taxpayer's urban property is, in each case, the maximum of three available values: (a) the value declared on Form 714 (Boxes A, A1, C, D and M), the cadastral value of the property plus the value declared on Form 720 [37].

2. *Variable selection.* We collected and studied the descriptive statistics of candidate variables: income, assets, net worth, net worth of the net holders of wealth, etc. We then examined the limitations derived from technical taxation issues, such as exemptions from declaration at specified wealth intervals or special regulations,

and decided that the best theoretical option for gaining a clear understanding of reality would be total value of assets.

3. *Decision on the optimal binarization of the data*. We decided the optimal binarization strategy. The wealth range is $[10^0; 10^{10}]$ with a mean, $\mu$, of the order of $5 \times 10^5$ within which both high and low incomes have to be closely scrutinized.

4. *Analysis*. Although different distributions have been studied (see Section 1 above), biparametric distributions have been most often used in meta-analysis. Normal, lognormal, Pareto and, occasionally, gamma distributions have been used to compare the quality of the fits against distributions with more parameters.

   As a reminder, Kleiber and Kotz [38] studied at length the types of this family of distributions in which Mandelbrot [39] singled out a strong form, expressed in Equation (3), where $m_0$ is a scale factor, and the value of $\alpha$ is not determined.

   (a) Pareto I distribution

   $$\pi(m) = \begin{cases} (\frac{m}{m_0})^{-\alpha} & \text{when } m > m_0 \\ 1 & \text{when } m \leq m_0 \end{cases} = \int_m^\infty P(m)dm \qquad (3)$$

   or alternatively, using the survival function, which is the complement of the distribution function:

   $$S(x) = 1 - \left(\frac{m_0}{x}\right)^\alpha. \qquad (4)$$

   Pareto's law [40] considers the variable in the positive range (even in the knowledge that there are people who are net debtors, he studied wealth above the survival threshold), where the higher the Pareto index, the smaller the proportion of very rich. The value output by modern empirical studies is close to 2, bigger than in older studies [19,41].

   (b) Lognormal distribution
   This type of distribution is intuitively applicable and is endorsed by Gibrat's research, because the distribution of any manifestation of wealth is universally asymmetrical [42].

   From another point of view, note that Koch [17,25,43–48] observed that asymmetry is a possible exponential trace. The exponential increase in the wealth of the richest could be studied by analogy with biological phenomena [42].

5. *Analysis of the population mix*. Pareto [49] observed that "society is not homogeneous". This ledthem to think that inequality would be understood by studying the shape of the distribution of the total population. We now know that this intuition was true and that the Pareto index is an indicator of this form. While in a conventional exploratory study, we have to infer the number of subpopulations in the mix, knowledge of multiple covariates for each taxpayer, like age, activity, and income details, as well as expert knowledge, greatly facilitates the identification and choice of the "natural" subpopulations.

   In a mixture experiment, the experimenter selects a number of different mixtures and varies the proportions of two or more of their components. If we call the number of components $q$, and $x_i$ is the proportion of the $i$-th component, then the value of each component is between 0 and 1, and its sum is 1. Due to the constraints, the geometric description the factor space that contains the $q$ components consists of all the points within the bounds of a $(q - 1)$ simplex. The points inside the rectangle have possitive components, and its centroid describes the mixture with an equal proportion of their components.

   Having identified the factor space, we could ascertain which mixture of populations best explains the empirically observed distribution. If society is made up of $K$ sub-

populations and the density function of each subpopulation is $f_k(\gamma; \theta_k)$, where $\theta_k$ is a parameter vector, then the density function of the population is

$$f_k(\gamma; \theta) = \sum_{k=1}^{K} \pi_k f_k(\gamma; \theta_k), \tag{5}$$

where $\pi_k$ is the proportion of each population, with $(0 \le \pi_k \le 1)$, and its sum is 1. The population density of the model is approximated by estimating the parameter vector, without knowing to which of the groups each member belongs. If the number of subpopulations is known, the parameters can be output by maximum likelihood, where the most common options are the EM algorithm [50] or Bayesian methods.

## 3. Wealth Distribution in Spain

### 3.1. Data Collection

Data from the National Institute of Statistics (2020) indicate that Spain has a population of 46.94 million. The Spanish Tax Agency's taxpayer census contains data on 72.87 million taxpayers, of which 64.25 million are natural persons. The difference between the above figures is due to the fact that the taxpayer census includes legal persons belonging to one of the 24 existing categories, non-residents, and deceased whose taxes have not prescribed. The baseline of the research was the year 2015, when there were 59.733 million natural persons, including non-residents far outnumbering the number of households (18.47 million) in the Household Financial Survey used in other studies [51].

Table 1 shows some descriptive statistics divided into three blocks. The first block shows income (gross tax base, as defined in art. 47 of the Income Tax Law) of declarants with a positive tax base. The remainder, up to a total of 38.261 million taxpayers, had a negative or zero base or were not obliged to declare. The second block, taxpayer net worth (NW) (assets less liabilities of all kinds), is divided into two groups, taxpayers with a NW greater than zero and taxpayers with a NW less than zero (net debtors). The last three columns are the values of the assets deciles for all taxpayers holding assets, grouped according to total, direct, and indirect assets (that is, assets owned due to participation in non-listed companies).

**Table 1.** Descriptive statistics of wealth distribution in Spain (2015).

|  | INCOME | NET WORTH | | ASSETS > 0 | | |
|---|---|---|---|---|---|---|
|  | >0 | >0 | <0 | Total | Direct | Indirect |
| $\sum$ (M€) | 442,810 | 4,228,499 | −123,893 | 5,504,770 | 4,372,866 | 1,131,697 |
| $\mu$ | 20,593 | 114,205 | −27,335 | 145,959 | 115,974 | 21,936 |
| # | 21,472,065 | 37,025,366 | 4,532,469 | 37,714,346 | 37,714,346 | 1,458,186 |
| $\sigma$ | 59,448 | 5,142,106 | 202.669 | 5,423,997 | 4,216,853 | 3,321,528 |
| D1 | 2806 | 165 | −65,663 | 143 | 128 | 0 |
| D2 | 5267 | 1258 | −41,439 | 1202 | 1123 | 0 |
| D3 | 9326 | 5417 | −27,605 | 5996 | 5490 | 0 |
| D4 | 13,632 | 18,652 | −18,286 | 27,783 | 26,145 | 0 |
| Me | 16,655 | 39,897 | −11,892 | 54,621 | 53,137 | 0 |
| D6 | 20,088 | 64,136 | −7558 | 81,058 | 79,063 | 0 |
| D7 | 24,336 | 95,216 | −4393 | 114,745 | 111,579 | 0 |
| D8 | 30,210 | 142,423 | −1820 | 165,736 | 159,887 | 0 |
| D9 | 38,428 | 240,632 | −480 | 273,795 | 270,312 | 0 |

The rows show the following variables: total sum, mean, standard deviation, number of elements and distribution deciles in euros. For example, the value of variable income observed for the taxpayer ranked in the position 2,147,207, first decile, is 2,806 €. The first positive values for the last variable, indirect assets, are in percentile 93. Note that only 3.86% of asset owners hold assets in non-listed companies and the proportion of taxpayers who do not hold assets is 59.7 M − 37.7 M = 22.01 M (36.85%).

There are major changes in variability ($\sigma$) in the evolution of the value of assets over time, see Table 2, which could be due to macro errors in taxpayer declarations that have to be accounted for. In this research, we have used data as declared and do not process possible errors.

**Table 2.** Evolution of assets over time.

| Var. | 2015 | 2016 | 2017 | 2018 | 2016/2015 | 2017/2015 | 2018/2015 |
|------|------|------|------|------|-----------|-----------|-----------|
| $\sum$ | 5,504,811 | 6,522,811 | 5,925,214 | 6,297,723 | 118.49 | 107.64 | 114.40 |
| $\mu$ | 149,959 | 168,858 | 156,417 | 1190,390 | 115.69 | 107.17 | 130.44 |
| $\sigma$ | 4,916,808 | 126,005 | 127,548 | 3,972,678 | 2.56 | 2.59 | 80.80 |
| # | 37,714,346 | 38,806,625 | 37,880,766 | 38,072,958 | 102.90 | 100.44 | 100.95 |

*3.2. Variable Selection*

The term income is used to mean different things in studies taking a statistical approach, where it generally refers to a variable of interest in an interval, $Y = [\underline{y}, \overline{y}]$, constituting any manifestation of wealth. We have studied three variables:

1. *Income.* Some problems are inevitable:
   (a) Low incomes (<€22,000) are exempt from declaration.
   (b) The comparability of different studies is limited because these variables are defined legally, and there are special income attribution schemes in each country.
   (c) The tax base may be negative.
   (d) Tax fraud related to salaried income is easier to detect than for income of professionals or the extremely wealthy.
   (e) The data distribution may be contaminated due to the tendency to defraud [52].
   (f) The distribution mode for low incomes is sensitive to binning. For example, with a €100 grouping, the mode is situated in the interval (€) [13,800–13,900] (# 73,822 taxpayers). For a €1000 grouping, the mode is in the interval [10,000–11,000] (# 537,999). For a €10,000 grouping, the peak is between [50,000 and 60,000].
   (g) The use of income introduces a bias because workers may have mortgages, or assets, placing them in a completely different situation to others, with a similar salary, that are or are not homeowners.
   (h) Income tax declarations are sometimes filed individually and sometimes by household unit. Atkinson [20] and Atkinson, Piketty and Saez [53] have discussed the limitations of income tax data.

2. *Net worth.* Its associated difficulties are:
   (a) Values may be negative if there are net debtors.
   (b) Although possible, it is very difficult to calculate. The problems are:
      i. All the tax data on declared incomes and debts are required for its calculation.
      ii. Details contained in the corporate tax declaration that are not available in all countries are required to calculate indirect equity, that is, shares owned in unlisted investee companies.
      iii. All the assets of the companies owned by taxpayers have to be transferred to the current price value.
      iv. The declared corporate tax values have to be adjusted for each company due to different amortization rates that are acceptable under accounting rules, and the process is very complex [37].
      v. Access to all the direct and indirect liabilities (loans and debts) of each taxpayer, even if assumed by the company in which the taxpayer has a share, is required.

3. *Total assets.* This variable is more detailed than income, is always positive and does not have so many technical issues (from the viewpoint of accounting and data accessibility) as net worth.

Therefore, we decided to use total assets in most cases, with a detailed analysis of net worth on some specific occasions.

We use data for a period covering the last four years.

### 3.3. Decision on Binarization of Data

The decision on whether the "bins" should be of the same length, their position and, generally, their width is not straightforward. Silverman's rule [54] is:

$$\hat{h}_{OPT} = \min\left\{\hat{\sigma}, \frac{\hat{q}_3 - \hat{q}_1}{1.349}\right\}n^{-1/5}.$$

Its logic is that the variance is more sensitive to extreme values than the interquartile range. We find that, in this case, $\sigma$ is much greater, which is logical in a curve whose right tail describes the distribution of the assets of big fortunes:

$$0.9 * \min\{4, 123, 829, (136, 972 - 2658)/1.349\}37, 714, 346^{-1/5} = 3039.$$

The range of assets in € is $[0, 10^{10}]$. The $\hat{h}_{OPT}$ is € 3039. This binning can be used for the low ranges. However, it implies $10^{10}/3039 \approx 10^7$ intervals. As we will see, this level of granularity is unnecessary.

### 3.4. Analysis of Wealth

First, we updated previous studies providing a global metric of the effect of microdata. In the case of Spain, García Docampo [55] used different data sources (the National Statistics Institute (INE) Household Financial Surveys and the INE Household Panel) and output Pareto indicators between 0.8 and 0.9. We repeated the method and output an index of 1.85 in line with the current tendency to find values closer to 2.

3.4.1. Variable Distribution

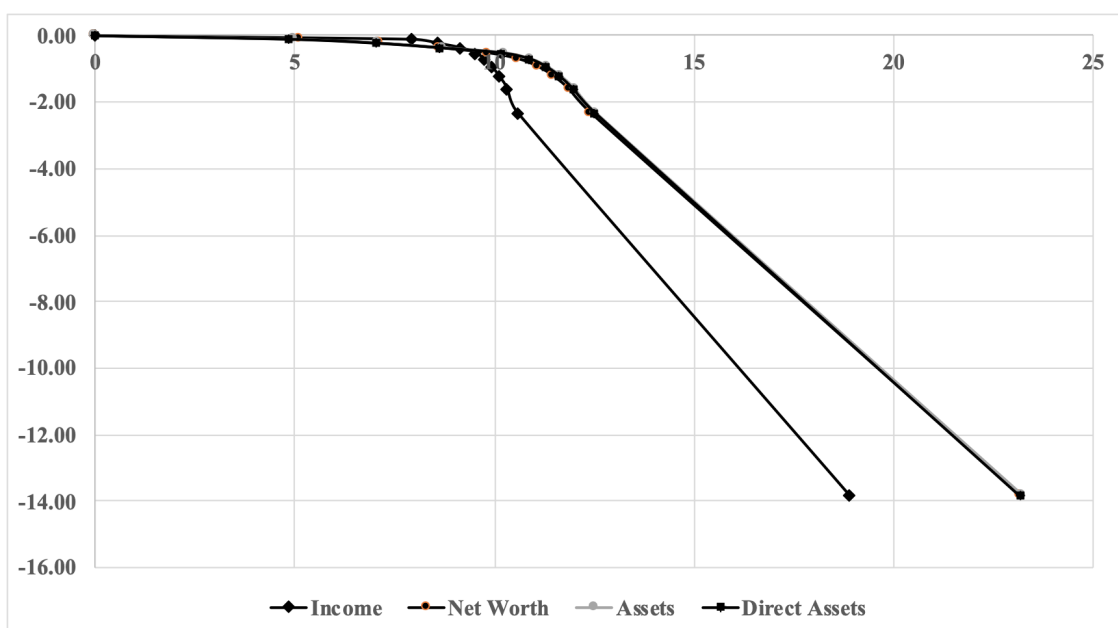Figure 2 shows the $ln - ln$ plots of survival functions for income, net worth and assets.



**Figure 2.** Distributions: $ln - ln$ of income, wealth and assets

The origin of the ordinate is always 0, because the value of 100% of the elements of each variable is greater than 0 (we have only plotted the positive part of the net worth variable whose magnitude is shown on the abscissa).

For example, only 10% of individuals will have a variable value greater than decile 9. Using the value $ln(1 - 0.9) = -2.30$ on the ordinate, we get the value of the variable considered on each curve, for example, 12.52 for assets, which is equivalent to €273,795 (Table 1).

We find that income has a singular behavior, especially in the lower region, because, given the declaration threshold, it is not universally declared, its range of variability is smaller than for wealth, and it is sometimes submitted for household units. The distribution for the other group of variables is similar. If we were to consider net worth, a transformation would be necessary because it has an interval with negative values. There are 4.53 million taxpayers with "negative assets", most of whom are paying back a mortgage on their home. Many of these taxpayers have income and are not accounted for only when the analyzed variable is "net assets > 0". Therefore, we concluded that the best option is to study total assets.

### 3.4.2. Definition of Subpopulations

We agree with Pareto that the population is composed of subpopulations with different characteristics. The use of microdata has the advantage that it is possible to ascertain how many subpopulations there are, as well as their characteristics. We hypothesize that the characteristics of a subpopulation will impose a specific pattern on the wealth interval in which it accounts for the numerical majority.

Figure 3 shows the number of taxpayers with assets. Figure 3a illustrates intervals from €$10^j$ to $10^{j+1}$ with $j = \{0, 1, 2, 3, \cdots, 9\}$. The first interval is $[10^0, 10^1)$. The abscissa represents the values of interval bounds $j : j + 1$. Figure 3b shows the details of the interval $[€10^4, €10^5)$ with a binning of 10,000. Figure 3a clearly shows that there is a subpopulation, probably minors, in the range €$[0, 10)$ without assets, and a population on the abscissa that is, on this scale, not normal, but could, however, be fitted to a lognormal distribution. The pattern shown in Figure 3b is similar, albeit on a natural and bimodal scale.
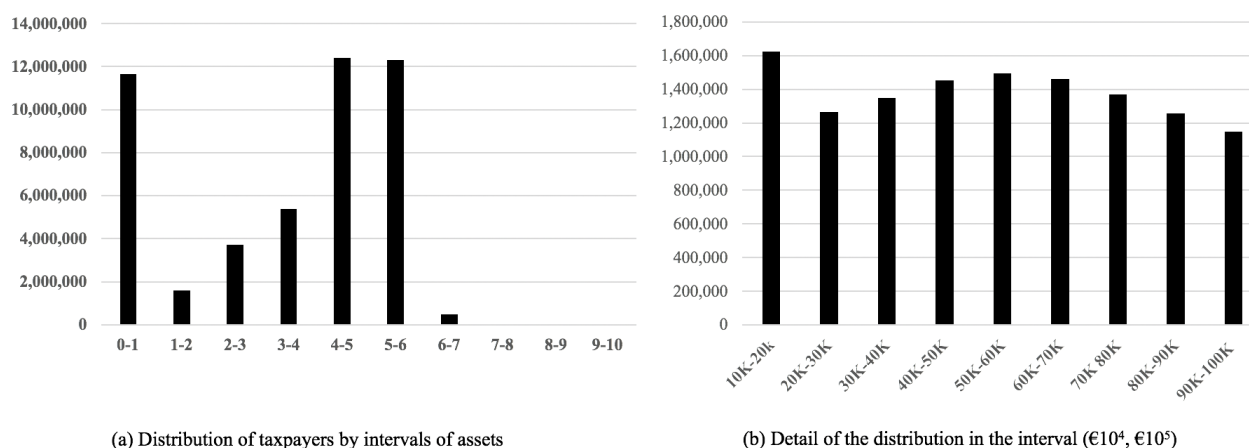


(a) Distribution of taxpayers by intervals of assets

(b) Detail of the distribution in the interval (€$10^4$, €$10^5$)

**Figure 3.** Assets by intervals.

There is a tendency to consider a *total* population but we should not forget that the observed data illustrate the emergence of a phenomenon originated by the coexistence of different wealth generation processes in different populations.

We tackle the problem by hypothesizing that there are five different sub-populations:

1.  Taxpayers without assets ($C_0$). Minors, net debtors, non-resident workers who receive income or subsidies, but do not have assets to their name. Taxpayers with assets ($C_a$) contains the following subgroups.
2.  Passive taxpayers ($C_p$). With assets received by donation or inheritance (minors) and attributed to them by third parties (their parents or financial institutions).
3.  Non-entrepreneurs ($C_n$). Salary earners and professionals whose main income is a salary. They do not hold shares in non-listed companies. The wealth generation process is governed by the accumulation of savings.
4.  Entrepreneurs ($C_e$). Owners or shareholders in unlisted companies. They are characterized by owning assets, shares in listed companies or participations in companies through other companies. The wealth generation process is a combination of the accumulation characteristic of savings (direct assets) and multiplicative effects (indirect assets), mediated in this case by random a variable (success of the non-listed companies) in a process that can be modelled based on Gibrat's ideas.
5.  Large fortunes ($C_f$) are, in some cases, exceptionally successful entrepreneurs but more often holders of large family fortunes. The wealth accumulation process is governed by other rules and is affected by inheritances and marriages.

### 3.4.3. Distribution of the Subpopulations

There is a population of 37,714,346 ($C_a$) that own assets and a group ($C_0$) of 22,019,068 not have assets to their name.

Figure 4 shows: (a) the distribution of the taxpayer census by birth year. The mode is 1977, and it illustrates the impact of sociological events, such as the mortality of the Civil War in Spain in 1936; (b) the distribution by birth year of the value of assets owned at three different wealth levels. The peak of the distribution of large fortunes (birth year 1962) corresponds with youngest people than lower fortunes (birth year 1947) and (c) distribution of wealth between minors and adults.
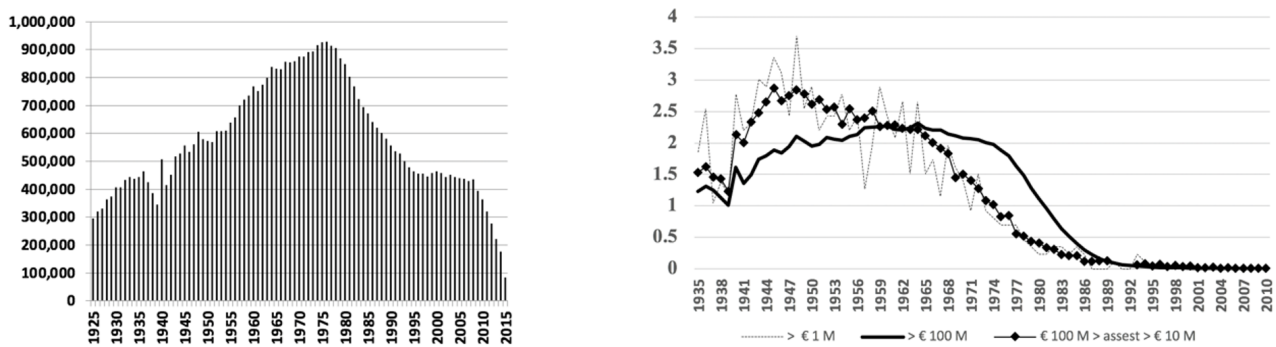


|          | Assets = 0 | Assets > 0 | #          |
|----------|-----------:|-----------:|-----------:|
| Minors   | 4,061,642  | 3,138,625  | 7,200,267  |
| Adults   | 17,957,426 | 34,575,721 | 52,533,147 |
| TOTAL    | 22,019,068 | 37,714,346 | 59,733,414 |

**Figure 4.** Net worth by year.

The low values of the distribution of assets could be considered to follow a normal distribution, assuming hypothetically that they are random donations or small gains without a regular pattern. Table 3 shows that, on the contrary, the asymmetry of the monotonically decreasing curves in all ranges (from €1 to €10, from €10 to €$10^2$, and up to €$10^4$–$10^5$) is similar.

In summary, donations to minors reflect the wealth of their parents.

**Table 3.** Frequency of assets in the range 1–100,000 €.

| Interval € | # | Interval | # | Interval | # | Interval | # |
|---|---|---|---|---|---|---|---|
| 1 to 10 | 1,845,563 | <100 | 3,522,730 | <1 K | 7,512,309 | <10 K | 13,411,653 |
| 10 to 20 | 419,723 | 100 to 200 | 858,755 | 1 K to 2 K | 1,791,396 | 10 K to 20 K | 1,520,055 |
| 20 to 30 | 264,339 | 200 to 300 | 606,925 | 2 K to 3 K | 1,035,198 | 20 K to 30 K | 1,447,770 |
| 30 to 40 | 205,114 | 300 to 400 | 508,197 | 3 K to 4 K | 744,920 | 30 K to 40 K | 1,520,055 |
| 40 to 50 | 161,698 | 400 to 500 | 430,310 | 4 K to 5 K | 564,586 | 40 K to 50 K | 1,614,721 |
| 50 to 60 | 186,280 | 500 to 600 | 390,268 | 5 K to 6 K | 465,260 | 50 K to 60 K | 1,639,613 |
| 60 to 70 | 133,745 | 600 to 700 | 347,597 | 6 K to 7 K | 397,574 | 60 K to 70 K | 1,591,818 |
| 70 to 80 | 113,784 | 700 to 800 | 309,361 | 7 K to 8 K | 336,237 | 70 K to 80 K | 1,483,085 |
| 80 to 90 | 104,315 | 800 to 900 | 281,055 | 8 K to 9 K | 300,479 | 80 K to 90 K | 1,353,336 |
| 90 to100 | 88,169 | 900 to 1K | 257,111 | 9 K to 10 K | 263,694 | 90 K to 100 K | 1,231,181 |
| $\sum$ | 3,522,730 | | 7,512,309 | | 13,411,653 | | 27,146,544 |

According to Pareto, we should find the value $m_0$ (mode), which serves to normalize the Pareto curve. A major issue in measuring the Pareto coefficient is the choice of threshold above which the distribution is assumed to adopt the Pareto shape. This is an arbitrary, but critical, decision because the value of $\alpha$ depends on the choice of the threshold. There are two approaches: (i) make a pragmatic decision, or (ii) use a statistical criterion to find a value, with the undesired collateral effect of yearly change. It is usual practice [20] to opt for a pragmatic approach, where an income threshold of £55,000 in the UK corresponded to the upper 5%.

We used different binnings for our analysis. With €1,000, the mode is €1,500, with 1,668,736 asset owners, with €10,000, the mode is in the interval [50,000–60,000) with 1,853,312 asset owners. If the binning is €100,000, the mode it is near the endpoint of the interval [0, 100,000). Applying Atkinson's criterion, we decided to take a pragmatic approach to facilitate comparison between years and countries, and we used €100,000, rounding to the value €101,000 for total assets used in previous studies [37]. From the pragmatic point of view, this suffices; it is placed between the 6th and 7th deciles.

The distribution of the group under the mode could be explained by a mixture of three populations: $C_0$, without assets (minors or adults), (36.86%), $C_p$ minors with assets, (5.25%), with a characteristic distribution similar in shape to that of their parents, and the least rich of the populations (57.89%), see Equation (5).

Our first hypothesis is that adults can be divided in two big subpopulations, *entrepreneurs*, including all individuals that have a holding in non-listed companies, and *non-entrepreneurs* (including farmers, professionals, civil servants, self-employed, etc.) We assume that there are rich and poor people in both groups, but we believe that their distribution, and, more importantly, their processes of wealth generation are different.

Figure 5a–c show the wealth distribution for non-entrepreneurs and entrepreneurs in three wealth intervals. Clearly, the curve of the total population is shaped by the combination of both subpopulations in the first interval, but the non-entrepreneur pattern is dominant only in the middle wealth interval.
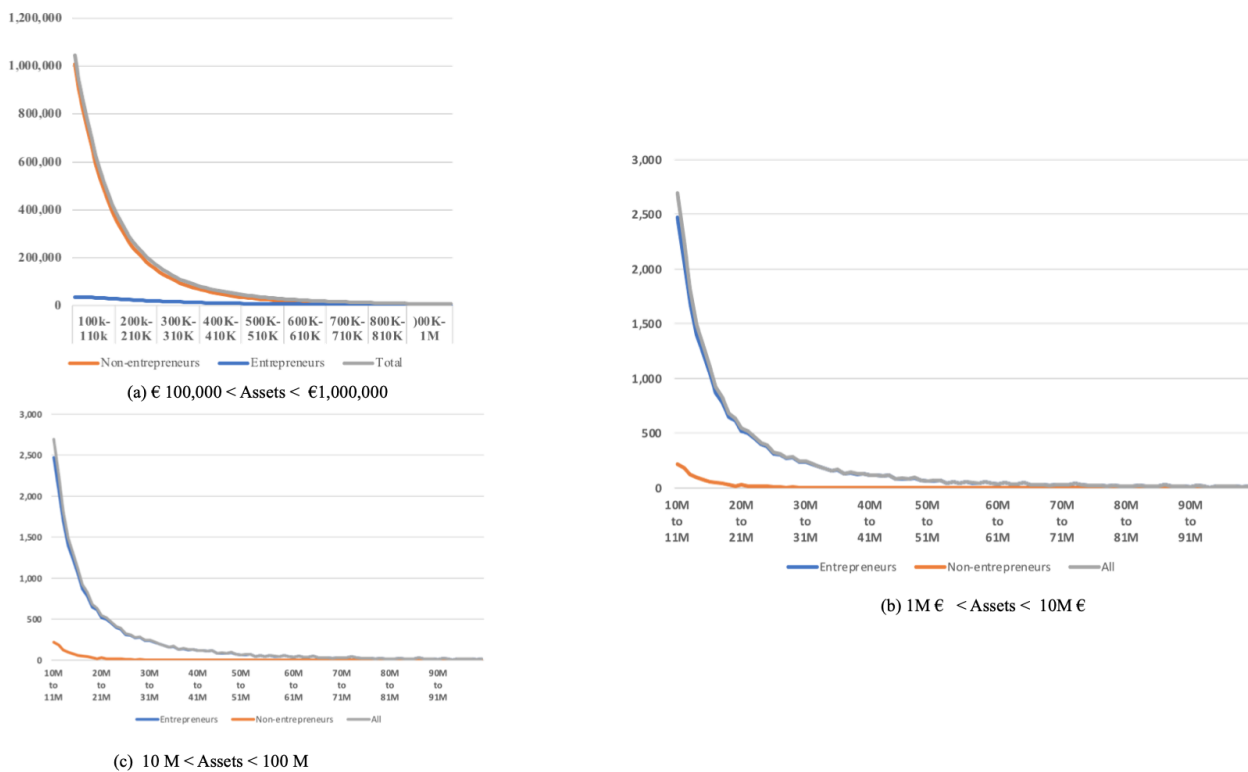
(a) € 100,000 < Assets < €1,000,000

(b) 1M € < Assets < 10M €

(c) 10 M < Assets < 100 M

**Figure 5.** Assets distribution.

Figure 6a shows that the two curves cross at the value 15.52 on the abscissa (approximately €5.5 million), and the entrepreneur distribution begins to reveal a different wealth accumulation process to non-entrepreneurs receiving a salary.
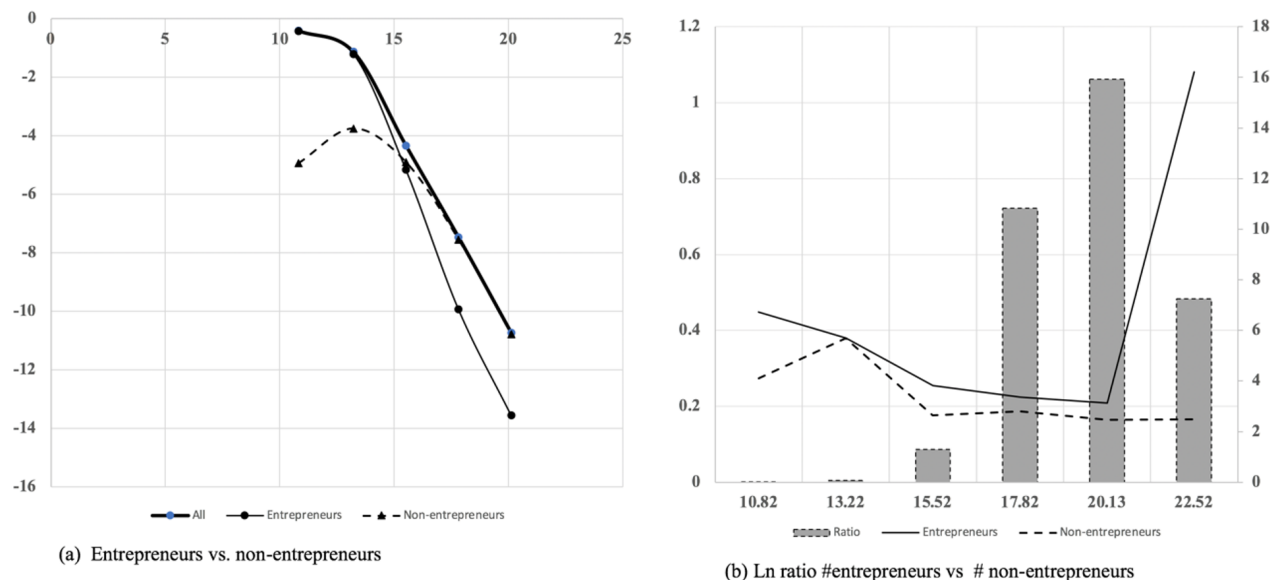


(a) Entrepreneurs vs. non-entrepreneurs

(b) Ln ratio #entrepreneurs vs # non-entrepreneurs

**Figure 6.** Assets with a value greater than mode.

This is studied in greater detail in Figure 6b with one continuous and one dotted curves (a ratio between entrepreneurs and non-entrepreneurs). The abscissa represents the $ln$ of wealth. The primary ordinate, on the left, shows the values of the $\mu$/Max variable. The left end of the curves indicates that the value of the ratio between the average and maximum (€100,000) values of the assets owned is 0.274 for non-entrepreneurs and

0.44 for entrepreneurs (whose average value is €50,000 and $ln(50,000) = 10.82$). Figure 6 shows transitions near to the abscissa values 13, 15 and 20, which are in the range of €500,000, €5,000,000, €500,000,000. In the proximity of these abscissa points, the slope of the accumulated curve changes quickly. The curve that fits the values observed in the range [10,13] does not fit the range [15,20], and the curve that fits this range does not fit the range of big fortunes. The observed curve could be considered as the "envelope" of the curves of different subpopulations.

This is compatible with the idea that the value of assets that workers can save is "limited" to €0.5 million or that is a natural limit to the process of accumulation of wealth, savings, of non-entrepeneurs.

The interval between €0.5 million and €5 million is characterized by a population that is a mixture of professionals and entrepreneurs, the interval [5 M, 500 M) by a population of successful entrepreneurs governed by a multiplicative process, and the interval [500 M, 10,000 M) is the range in which the accumulation process applicable to large fortunes holds.

The columns in Figure 6b shows the ratio between the number of entrepreneurs and non-entrepreneurs in each interval. On the far left, there are 270,266 entrepreneurs and 24,733,827 non-entrepreneurs, with a ratio of 0.01. The maximum ratio value of 15.94 (right secondary axis) is reached in the interval [10 M, 100 M) with 19,802 entrepreneurs and 1830 non-entrepreneurs. Looking at the two curves, we find that the shape of the curve is characterized by the parameters of the subpopulation of successful entrepreneurs above the threshold of €500 million.

Figure 7 shows the behavior of the different subgroups of the population. Entrepreneurs own more assets than non-entrepreneurs, and this is clearly illustrated by the shapes of their distributions which are very unalike, especially in the lower wealth intervals. Hence, the accumulation process must be different.
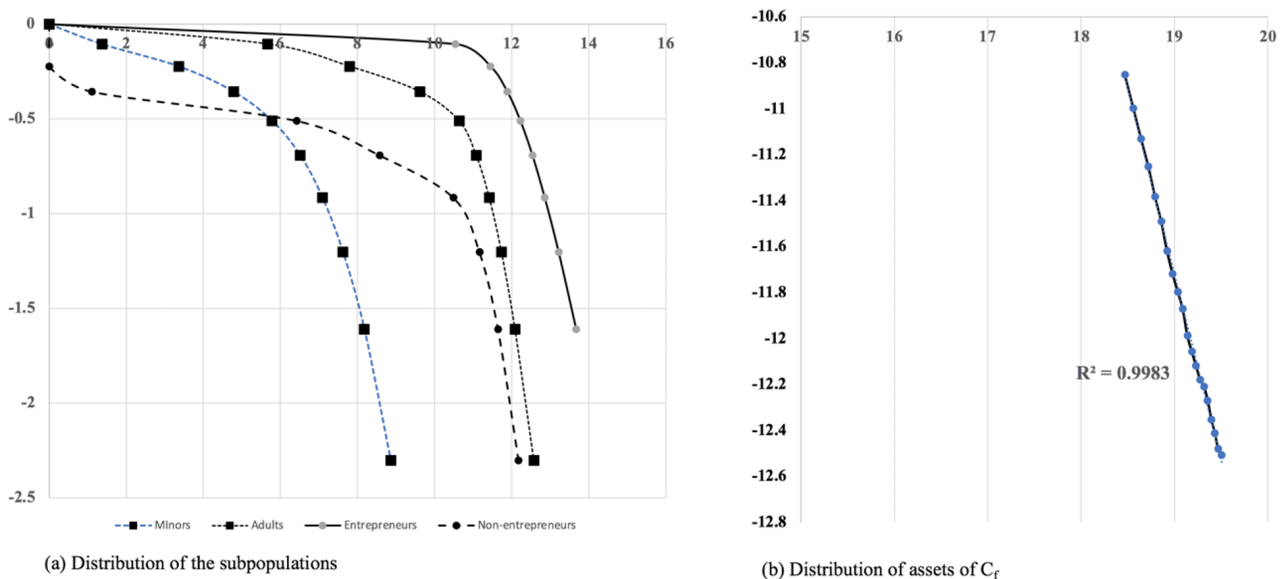


(a) Distribution of the subpopulations        (b) Distribution of assets of $C_f$

**Figure 7.** Behavior of subpopulations and large fortunes.

Figure 7b shows that the distribution of the richest ($C_f$), which can be fitted to a potential curve. It represents the 776 cases with assets between €100 million and €300 million. Using a power law, the fit is near perfect ($R^2 = 0.99$), also accounting for the 141 people with assets between €300 million and €10,000 million (whose data are not represented in the graph for privacy reasons).

Figure 8 illustrates the evolution of the distribution of large fortunes over the period 2015–2018, with, in all cases, very good fits and a Zipf coefficient of the order of $-1.3$.
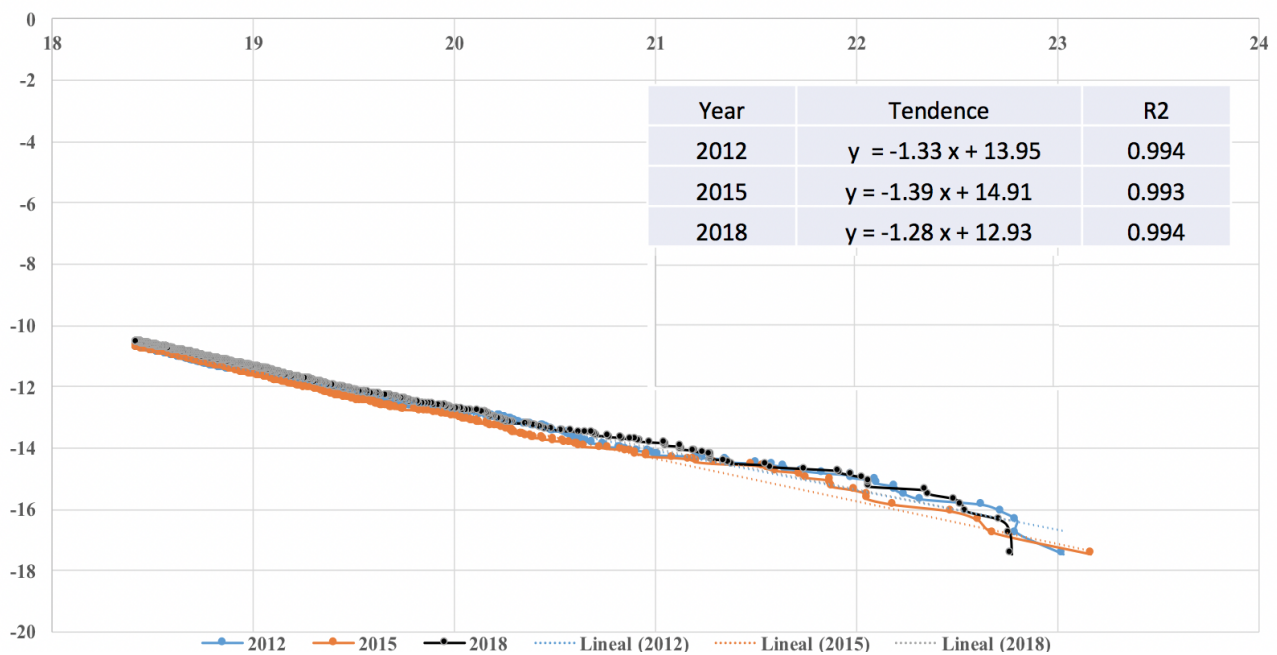
**Figure 8.** Evolution of the top wealth interval of over time.

### 3.5. Analysis of the Components of Wealth

Table 4 shows the components of the net worth of taxpayers by intervals of net worth (2015). Each row # contains the number of taxpayers in this range. There are fewer natural persons with positive net worth (37,006,939) than natural persons with positive assets (37,714,346) because some are indebted. The right-hand columns, as of liabilities, express the ratio of each variable to total assets, including in its last column financial assets.

**Table 4.** Distribution of net worth and its components by intervals.

| Interval | # | % WITH RESPECT TO TOTAL ASSETS | | | | |
|---|---|---|---|---|---|---|
| | | Liabilities | Indirect Assets | Real Estate | Current Accounts | FF.AA |
| Less 100 K | 29,965,372 | 63.59 | 9.25 | 73.15 | 14.03 | 3.57 |
| 100 K to 1 M | 9,629,289 | 11.01 | 8.24 | 70.35 | 15.27 | 6.14 |
| 1 M to 10 M | 364,357 | 17.63 | 40.23 | 37.59 | 7.19 | 14.99 |
| 10 M to100 M | 14,326 | 23.05 | 69.62 | 9.09 | 3.11 | 18.18 |
| 100 M to 1000 M | 471 | 25.45 | 79.47 | 2.03 | 1.74 | 16.76 |
| >1000 M | 19 | 26.43 | 79.44 | 0.23 | 0.50 | 19.83 |

Taxpayers at the bottom of the scale (whose wealth is below the mode of €100,000) owe 63.59% of the value of their assets, 73.15% of which are accounted for by real estate (urban property), 9.25%, by indirect assets held in non-listed companies, another 14.03% are held in current accounts and deposits, and 3.57% are financial assets.

We highlight three facts:

1. The growing importance of indirect assets (shares in non-listed companies) as we move up from the bottom interval (9.25%) to large fortunes (79.44%), where it far outweighs total financial assets (19.83%) (including shareholdings in IBEX-listed companies).

2.　The importance of the value of immovable property in the analysis. The share of real estate assets (first and second residences) that decreases with wealth. It represents 73.15% of the wealth of the poorest and accounts for an almost negligible fraction (0.50%) of large fortunes.

3.　Liabilities and indirect assets play an extremely important role in conducting a correct analysis. If the variable used does not take liabilities into account, the poor appear to be richer (inequality is attenuated). If indirect assets are not counted, the richest appear to be poorer than they really are.

　　It is of the greatest importance to highlight that, when models are fitted to income or direct assets, the presence of residuals is not a direct manifestation of model insufficiencies. These models are not fitting the real wealth.

　　Analyzing its components in Figure 9 (plotting *ln* values), we observe that the distribution of real estate assets almost perfectly fits the Pareto curve in the interval from €10,000 to €20,000,000, but differs from total assets and total wealth. It is also very clear that the distribution of total assets, which fits a Pareto distribution nicely, is not fully explained by real estate and financial assets, that is, these assets do not suffice to explain wealth.
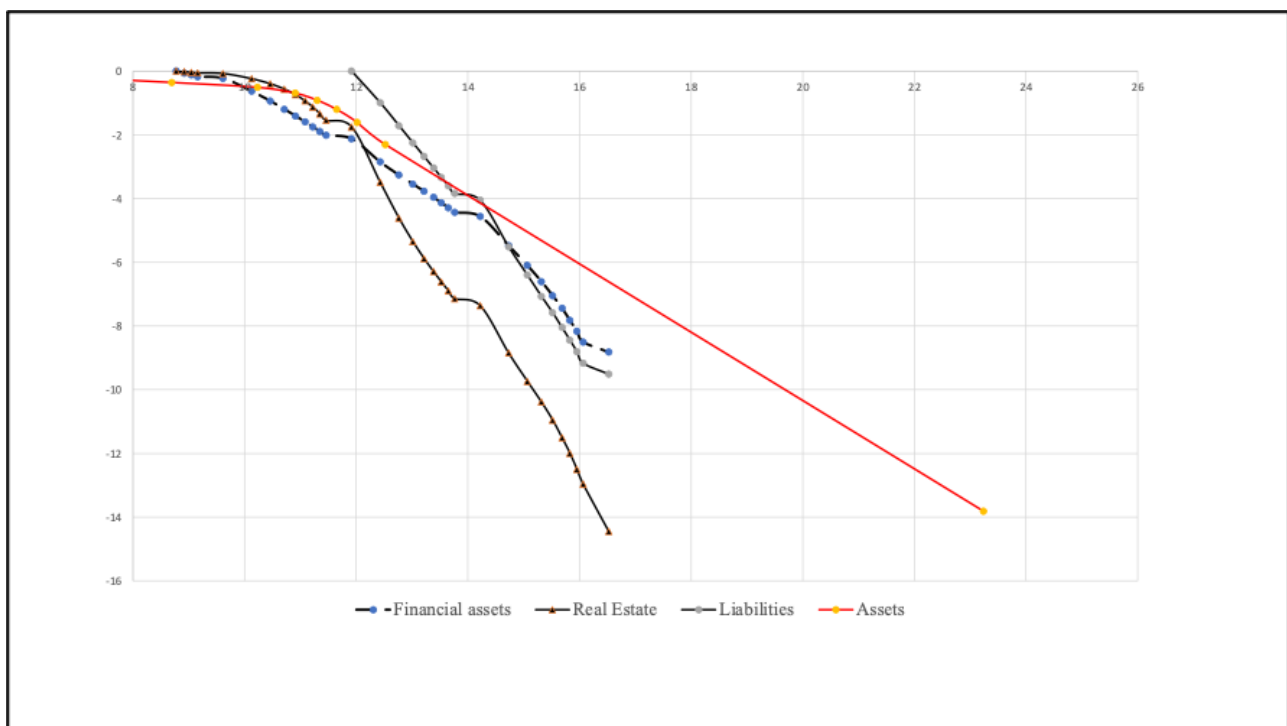


**Figure 9.** Distribution of the components of wealth.

　　Figure 10 shows that, globally, a gamma distribution provides a correct fit. However, it is not a matter of fitting empirical data to a curve: this problem has been solved.
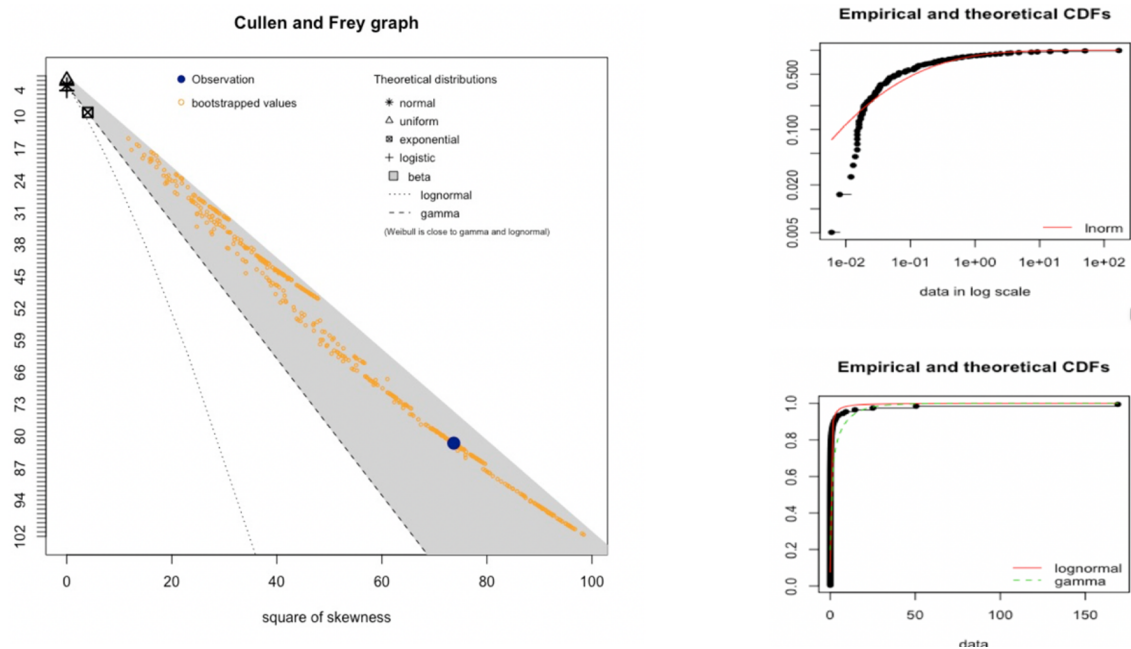
**Figure 10.** Cullen and Frey graph.

### 3.6. Characterization of the Nature of the Underlying Processes

In the light of this analysis, we describe some characteristics of the underlying economic processes. Table 5 shows the number of individuals, range, mean and standard deviation of the population subgroups. Entrepreneurs contains the population with indirect assets, including minors. For reasons of privacy regarding the biggest fortunes in Spain, we use the notation $a * 10,000$ M with $10 > a > 1$.

**Table 5.** Parameters of the subpopulation distributions.

|  |  | # | Range | $\mu$ | $\sigma$ |
|---|---|---|---|---|---|
| Total Population | | 59,733,414 | [0, 10,000 M] | | |
| WITHOUT ASSETS | $C_0$ | 22,019,068 | [0, 0] | 0 | 0 |
| Minors without assets | | 3,128,625 | [0, 0] | | |
| Adults without assets | | 18,890,443 | [0, 10,000 M] | | |
| ASSET OWNERS < 0 | | 4,532,569 | | | |
| ASSET OWNERS | $C_a = C_n + C_f$ | 37,714,346 | [0, $a * 10,000$ M] | | |
| Minors with assets | $C_p$ | 4,072,002 | [0; 130 M] | 5 | 165 |
| Adults with assets | | 34,585,721 | [[0, $a * 10,000$ M] | 159 | 4,318,838 |
| Non-entrepreneurs | $C_n$ | 35,256,160 | [0, 10,000 M] | 98,974 | 217 |
| Entrepreneurs | $C_e$ | 1,458,186 | [0; 100 M] | 890 | 3,031,214 |
| Big fortunes (>100 M) | $C_f$ | 917 | [100 M, $a * 10,000$ M] | 318,213,094 | 787,168,132 |
| Entrepreneurs | | 863 | [100 M, $a * 10,000$ M] | 314,059,377 | 785,624,254 |
| Other cases | | 44 | [100 M, $a * 10,000$ M] | 512,925,903 | 857,383,064 |

#### Generation of Additive and Multiplicative Effects

We explored the idea put forward by Limpert et al. [42], studying the mechanisms that induce lognormal distributions and the principles of additive and multiplicative effects. They used the example of an experiment with two ordinary dice. The addition of the two numbers leads to values from 2 to 12 with a mean of 7. Total range can be expressed as

$7 \pm 5$. The multiplication of the two numbers produces a skewed distribution with a range from 1 to 36. The range in this case can be expressed as $(x/c, x * c)$. In this specific case we obtain $c = 6$ and $x = 6$. In this case, symmetry has shifted to the multiplicative level, in a distribution with mean $\mu = 12.25$. They illustrated this point using natural models based on Galton's quincunx. In the first case, they used rows of equilateral triangles whose edges lay at $x + c$ and $x - c$ from the central vertex to get normal distributions with additive symmetry and a variant of quincunx with scalene triangles with edges at distances $x * c$ and $x/c$ to get lognormal distributions [56].

We took up this idea to investigate the uniformity of the wealth creation process in each interval of wealth. Each interval, $i$, except the first $[0, 10]$, has endpoints $[10^j, 10^{j+1}]$ with $j \in \{1, 2, \cdots, 9\}$. If, following Gibrat's ideas, the distribution happens to be correctly described by a lognormal, we could affirm that it reflects an underlying multiplicative process.

We know that:

1. In each interval $[10^j, 10^{j+1}]$, by construction, $\frac{x_{j+1} * c_{j+1}}{x_{j+1}/c_{j+1}} = \frac{10^{j+1}}{10^j} = 10$. For all the intervals, $c_{j+1} = (10)^{1/2} = 3.16, \forall j$, that is, $ln(3.16) = 1.15$.

2. If the underlying process is multiplicative, with the proposed intervals $[10^j, 10^{j+1}]$, it holds that there is a point $x_{j+1}$, in the interval

$$x_{j+1} * c_{j+1} = 10^{j+1}$$

$$\frac{x_{j+1}}{c_{j+1}} = 10^j$$

and, therefore,

$$x_{j+1} = 10^{(2j+1)/2}, \forall j. \tag{6}$$

If we compare the real value for both subpopulations, we would observe any difference with respect to the situation where the accumulative effect holds. We use for this observed value the notation $y$:

$$y_e = \frac{10^{(2j+1)/2}}{\mu_{entrepreneurs}}, \tag{7}$$

and

$$y_n = \frac{10^{(2j+1)/2}}{\mu_{non-entrepreneurs}}. \tag{8}$$

In Table 6, the columns containing the values of the group of entrepreneurs and non-entrepreneurs show the values of $\mu$, $\sigma$ and their ratio in the intervals $[10^j, 10^{j+1})$ from $j = 0$ to $j = 9$. For these groups, the $ln$ of $c = \mu/\sigma$ in each case and the excess with respect to the ideal behavior (1.15) of the multiplicative symmetry are shown in the right-hand column.

Figure 11 contains three graphics. In all of them we represent in abscissa the $ln$ of assets. In the upper left corner, we show the distribution of the ratio $(\mu/\sigma)$ for entrepreneurs, non-entrepreneurs and total population. In the lower part the graphic shows the ratio $ln(\mu)/ln(\sigma)$.

In the right side we express the difference with respect to the theoretical value of 1.15, that in our construction of the intervals of wealth, is the expression of a multiplicative process, that appears as a log normal distribution. We observe a symmetry that is expressed with the double arrow in the graphic, among the lines with the values of entrepreneurs and non-entrepreneurs.

**Table 6.** Distribution by wealth intervals and subpopulations.

| | ENTREPRENEURS | | | NON-ENTREPRENURS | | | E | N.E | E | N.E |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\frac{\mu}{\sigma}$ | $\mu$ | $\sigma$ | $\frac{\mu}{\sigma}$ | Ln() | Ln() | Exc. | Exc. |
| 0–1 | 3.34 | 3.01 | 1.11 | 0.02 | 0.10 | 0.20 | 1.09 | 1.70 | 0.06 | −0.55 |
| 1–2 | 52 | 25 | 2.08 | 42 | 26 | 1.62 | 1.23 | 1.15 | −0.08 | 0.00 |
| 2–3 | 569 | 261 | 2.18 | 450 | 263 | 1.71 | 1.14 | 1.10 | 0.01 | 0.05 |
| 3–4 | 4840 | 2543 | 1.90 | 3830 | 2471 | 1.55 | 1.08 | 1.06 | 0.07 | 0.09 |
| 4–5 | 54,091 | 27,005 | 2.00 | 53,382 | 25,653 | 2.08 | 1.07 | 1.07 | 0.08 | 0.08 |
| 5–6 | 364,262 | 225,982 | 1.61 | 199,751 | 156,081 | 1.28 | 1.04 | 1.02 | 0.11 | 0.13 |
| 6–7 | 2,510,687 | 1,780,143 | 1.41 | 1,686,501 | 1,100,573 | 1.53 | 1.02 | 1.03 | 0.13 | 0.12 |
| 7–8 | 22,482,106 | 15,811,009 | 1.42 | 16,394,325 | 10,320,486 | 1.59 | 1.02 | 1.03 | 0.13 | 0.12 |
| 8–9 | 206,604,515 | 143,069,745 | 1.44 | 158,398,773 | 86,480,258 | 1.83 | 1.02 | 1.03 | 0.13 | 0.12 |
| 9–10 | 3,233,259,800 | 2,871,975,594 | 1.13 | 2,286,954,185 | 732,488,355 | 3.12 | 1.01 | 1.06 | 0.14 | 0.09 |
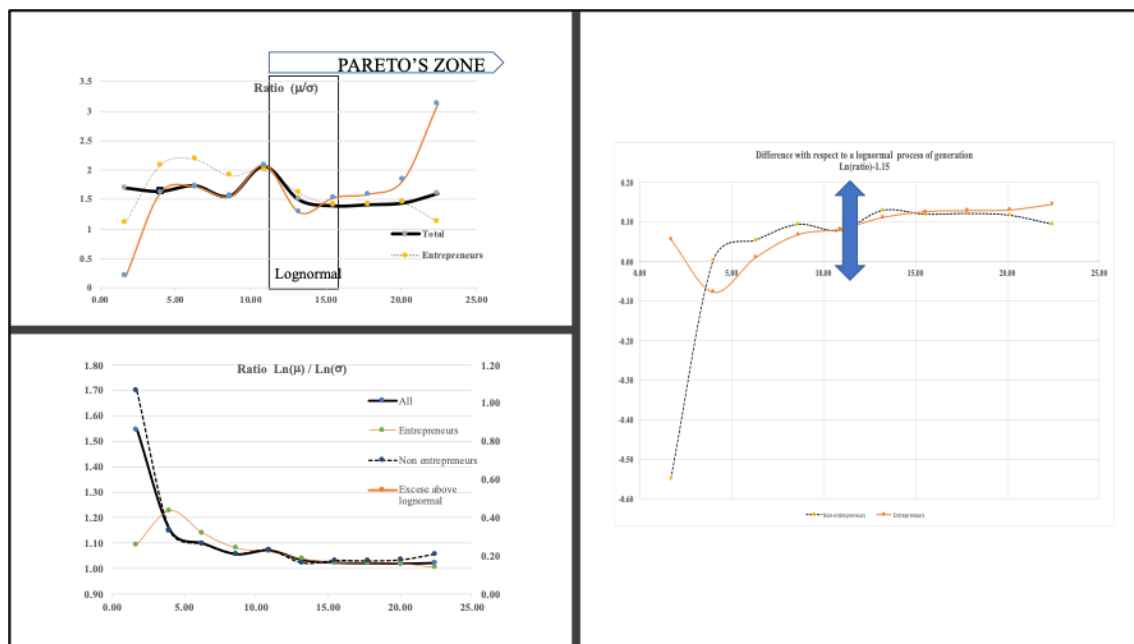


**Figure 11.** Analysis of the subpopulations in ranges of order 10,000.

In this graphic we can see that, over the intervals, there is an excess for total population with respect to the expected behavior of a multiplicative process.

We find that, throughout the region covered by Pareto's law (on the right-hand side of the mode) and even up to €10,000 on the left-hand side in the region of the 3rd to 4th deciles, it holds that the value $ln(\frac{\mu}{\sigma})$ is similar to the theoretical value of multiplicative accumulation (1.15). The excess with respect to the 1.15 value has an asymptotical behavior toward an excess of 0.1. We can see that this effect is generated by the mixture of two subpopulations with a symmetrical behavior with respect to the asymptotic and ideal value of 1.15.

There is a zone, defined by the process of accumulation of entrepreneurs that can be fitted with precision to a lognormal, but the upper wealth interval, big fortunes, is characterized by a different process of wealth accumulation, inheritances and marriages.

In Figure 12, we develop the idea expressed in Equations (6)–(8). We interpret that the observed values of $\mu$, greater that the values that should be expected in a multiplicative process, are the expression of the coexistence of one multiplicative process of accumulation of wealth with another that should be observed in absence of non-entrepreneurs, more similar to a Pareto's distribution.
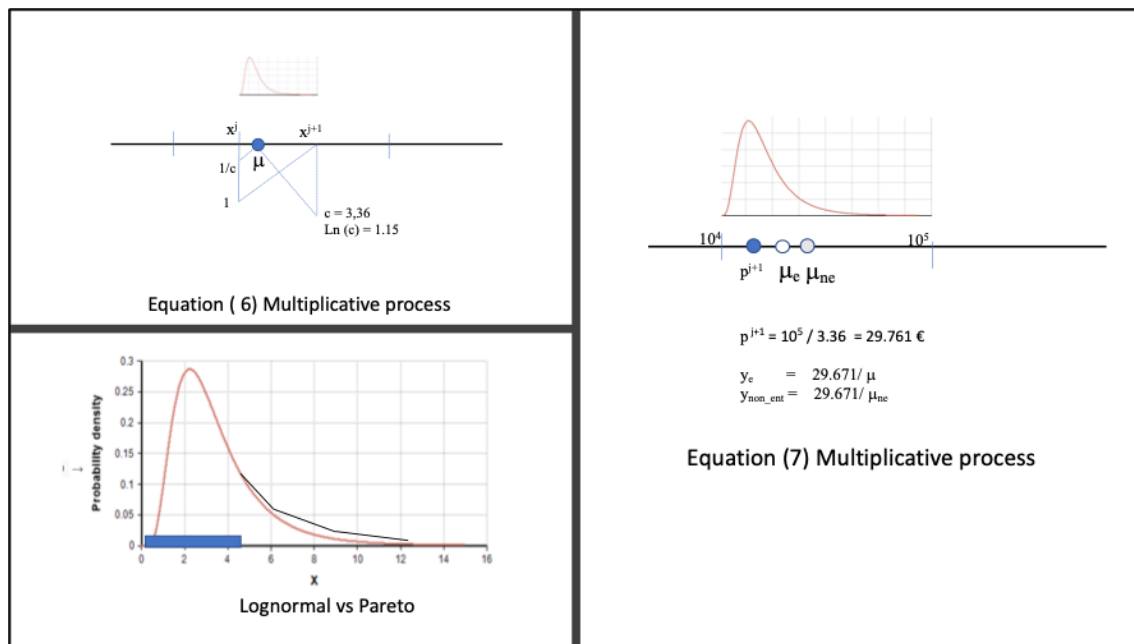
**Figure 12.** Interpretation of Equation (6)–(8).

Table 7 includes the distribution of the number of these groups and their evolution over the time.

**Table 7.** Distributions of entrepreneurs by wealth interval over the time.

| Year | 100 K < Actives < 1 M | | 1 M < Actives < 100 M | | 100 M < Actives < 500 M | | 1500 M < Actives < 15 KM | |
| | Entrep. | N-Entrep. | Entrep. | N-Entrep. | Entrep. | N-Entrep. | Entrep. | N-Entrep. |
|---|---|---|---|---|---|---|---|---|
| 2015 | 1,143,143 | 11,151,678 | 324,620 | 193,889 | 763 | 15 | 82 | 3 |
| 2016 | 1,183,178 | 11,330,440 | 333,804 | 206,028 | 796 | 19 | 92 | 1 |
| 2017 | 1,176,698 | 11,970,874 | 363,025 | 276,286 | 822 | 22 | 102 | 2 |
| 2018 | 1,192,844 | 11,225,120 | 375,442 | 281,007 | 877 | 20 | 99 | 10 |

It is clear that the regularity observed for the total population emerges from the mixture of two subpopulations but it is the sum of two regularities.

Table 8 shows the values obtained with (7) and (8) obtained to investigate if it is the case that there is a subjacent multiplicative process, using in this case the position of the mean in each interval of wealth.

**Table 8.** Deviations with respect to multiplicative process of the mean.

| Interval | $F = (2j+1)/2$ | 10F | $\mu$ (Entrepreneurs) | $y_e$ | $\mu$ (Non-Ent.) | $y_n$ |
|---|---|---|---|---|---|---|
| 0–1 | 0.5 | 3.16 | 3 | 1.05 | 4 | 1.26 |
| 1–2 | 1.5 | 31.62 | 52 | 0.61 | 42 | 1.33 |
| 2–3 | 2.5 | 316 | 569 | 0.56 | 450 | 1.42 |
| 3–4 | 3.5 | 3162 | 4840 | 0.65 | 3830 | 1.21 |
| 4–5 | 4.5 | 31.623 | 54,091 | 0.58 | 53,382 | 1.69 |
| 5–6 | 5.5 | 316,228 | 364,262 | 0.87 | 199,751 | 0.63 |
| 6–7 | 6.5 | 3,162,278 | 22,482,106 | 1.26 | 1,686,501 | 0.53 |
| 7–8 | 7.5 | 31,622,777 | 22,482,106 | 1.41 | 16,394,325 | 0.52 |
| 8–9 | 8.5 | 316,227,766 | 206,604,515 | 1.53 | 158,398,773 | 0.50 |
| 9–10 | 9.5 | 3,162,277,660 | 3,233,249,800 | 0.98 | 2,286,954,185 | 0.72 |

Figure 13 shows the ratio between the theorical $\mu$ and $\hat{y}$ observed for each interval and subpopulation. For example, the value 1.69 in the interval $[10^4, 10^5]$ expresses that the

value of $\mu$ (€53,382) for the subpopulation of non-entrepreneurs is 1.69 times the value of $10^{(2*4+1)/2}$, that is, €31,623.
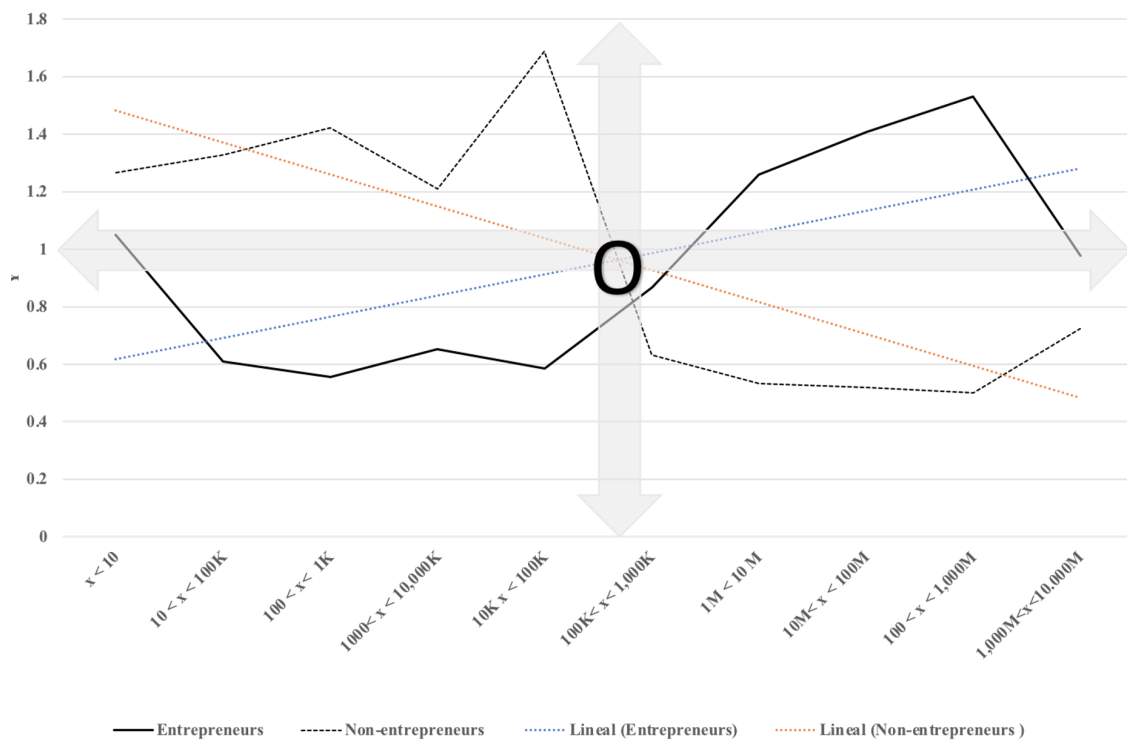


**Figure 13.** Analysis by subpopulations.

The total population is composed of two subpopulations (entrepreneurs and non-entrepreneurs) whose behavior in the graph is almost symmetric with respect to $y = 1$) and almost symmetric with respect to the threshold of Pareto, but the combined effect of the mixture of subpopulations is symmetric with respect to two axis, the ideal value of $y = 1$ and the Pareto threshold chosen. The lines of tendency of both subpopulations cross with a strange precision in O

Non-entrepreneurs save and invest in their homes or in the stock market. Until the limit of €0.5 million in the case of Spain they outnumber the entrepreneurs. Some of them have large fortunes, but the process of accumulation by saving and with investments in urban property has a decreasing rate of return, and presence of non-entrepreneurs in the right zone of the distribution is sparse.

In the interval between €0.5 M and €1000 M the dominant process is the activity of entrepreneurs with a different process of accumulation of wealth, multiplicative. There is a new change in the wealth interval of big fortunes that increase their wealth by investment, inheritance and marriage that reflects in a Pareto distribution.

Looking at the global trendlines, without the use of microdata, the observable total distribution of wealth can be fitted, by intervals, to two or three distributions one of them, when a threshold is reached, a Pareto distribution.

Figure 14 concludes the process initiated in Figure 2. There are three curves: total assets in the middle and indirect assets on the right and liabilities in the left. The methodology explained in Section 2 has found that indirect wealth, the value of the non-listed companies owned by entrepreneurs, accounts a 11.08% of the total wealth in Spain.

Studies that do not include indirect wealth, the wealth accumulated in non-listed companies or use the income declared in tax form clearly does not reflect the reality.
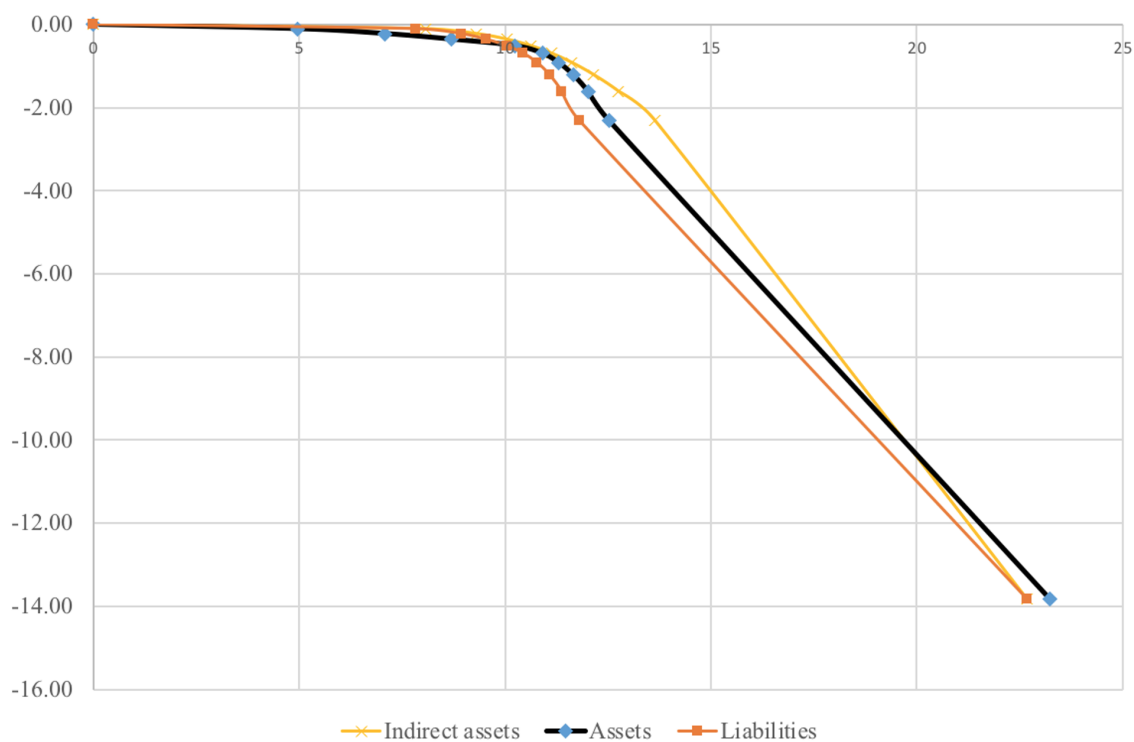
**Figure 14.** Analysis by subpopulations.

We have proved that there are two wealth accumulation processes—one for non-entrepreneurs (savings) and another for entrepreneurs (investments), the second is multiplicative.

There are five subpopulations, and the shares can be calculated from public statistics. There are two wealth accumulation processes (saving and investment). The evolution of the accumulation process can be studied using Equation (8). We suggest that this strategy could be used to fit data from sources with different granularity.

## 4. Conclusions

We have shown that:

1. The best option for analyzing wealth distribution regularities and studying inequality is total assets, including direct and indirect assets, because it uses more relevant data for accurately studying entrepreneur activity. The choice between the two available options of studying positive values or positive and negative values has to be made considering the objectives of the study and the availability of microdata on mortgages.

2. Analyzing the Spanish population with a wealth distribution in the range $[0, 10^{10}]$, we have found that the best threshold for Pareto analysis is at the point $10^5$ and the best binning is $10^4$.

3. We have identified five subpopulations with the following population shares: $C_0$ (without assets) accounts for 36.86%. Owner of assets is a mixture of four subpopulations, minors, non-entrepreneurs and entrepreneurs and big fortunes.

4. We have identified two types of accumulative processes (savings) and (investments), where the region of transition between the processes is $5.5 \times 10^7$. The process of generation of big fortunes is a mixture of both.

5. The distribution of the subpopulations with assets in the Pareto zone can be fitted to lognormal curves, but the process of accumulation proposed by Gibrat is clearer for entrepreneurs. Within the population that work for a salary, wealth accumulation is attenuated by the payment of mortgages. Table 4 shows that, on average, 73.15% of the value of assets owned by people that work for a salary takes the form of real

estate, that is, their homes, where liabilities (mostly mortgages) account for 63.59% of total assets. Their wealth accumulation is limited by interest payments and is, in some cases, driven by the increase of the property values. There is a huge difference with the average "private" liabilities of entrepreneurs (with total assets ranging from €100 K to €1 M), which are less affected by interest rates.

6. The wealth accumulation process is different for large fortunes, where it is relevant the wealth accumulated as shareholdings in non-listed companies.

7. Wealth models should consider that: (a) the subpopulation of minors and donees of wealth is distributed so as to reflect the wealth of the donors (and has a lognormal distribution), (b) the population of entrepreneurs (with corporate assets) should be differentiated from the population of employees and professionals (non entrepreneurs), which behave differently, and aggregate behavior with a level of multiplicative symmetry emerges only as a result of the mixture of these two populations.

8. We have explained the difference observed in empirical studies with respect to the law of Pareto by the effect in the wealth distribution of two processes of accumulation of wealth, one of them stationary and other characteristics of entrepreneurs that changes within each society with different uses of technology and with the growing importance of financial activities.

**Author Contributions:** Conceptualization, I.G.G. and Mateos; methodology, I.G.G. and A.M.C.; software, I.G.G.; validation, I.G.G. and A.M.C.; formal analysis, I.G.G.; investigation, I.G.G. and A.M.C.; resources, I.G.G.; data curation, I.G.G.; writing—original draft preparation, I.G.G. and A.M.C.; writing—review and editing, I.G.G. and A.M.C.; visualization, I.G.G.; supervision, A.M.C. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chandra Dash, K. *The Story of Econophysics*; Cambridge Scholar Publishing: Cambridge, UK, 2019.
2. Jovanovic, F.; Schinckus, C. The History of Econophysics Emergence: A new Approach in Modern Financial Theory. *Hist. Political Econ.* **2013**, *45*, 443–474. [CrossRef]
3. Savoiu, G.; Siman, I. History and Role of Econophysics in Scientific Research, Cap 1 en Econophysics: Background and Applications in Economics. *Financ. Sociophys.* **2013**, 3–16. [CrossRef]
4. Stanley, H.E. *Stanley on Econophysics*; Sage Publication: Sauzend Oaks, CA, USA, 2013; Volume 2, pp. 73–78.
5. Bouchaud, J.P. Econophysics: Still Fringe after 30 Years? Available online: https://arxiv.org/abs/1901.03691v1 (accessed on 8 February 2021).
6. Axtell, R. Zipf distribution of US firm sizes. *Science* **2001**, *293*, 1818–1820. [CrossRef] [PubMed]
7. Atkinson, A.B.; Voitchovsky, S. The distribution of top earnings in the UK since the Second World War. *Economica* **2010**, *78*, 440–459. [CrossRef]
8. Rawls, J. *A Theory of Justice*; Cambridge, M.A., Ed.; Harvard University Press: Cambridge, MA, USA, 1971.
9. Sen, A. *Poverty and Famines*; Oxford University Press: New Delhi, India, 1999.
10. Forbes, K.J. A reassessment of the relationships between inequality and growth. *Am. Econ. Rev.* **2000**, *90*, 869–887. [CrossRef]
11. Piketty, T. *21 Lessons for the 21st Century*; Harvard University Press: Cambridge, MA, USA, 2013.
12. Mantegna, R.N.; Stanley, H.E. *21 Lessons for the 21st Century*; Spiegel & Grau: New York, NY, USA, 2018.
13. Mantegna, R.N.; Stanley, H.E. *An introduction to Econophysics: Correlations and Complexity Finance*; Cambridge University Press: Cambridge, UK, 2001.
14. Chatterjee, A.; Yarlagadda, S.; Chakrabarti, B.K. (Eds.) *Econophysics of Wealth Distributions*; Springer: Milan, Italy, 2005.
15. Champernowe, D.G. A model of income distribution. *Science* **1953**, *63*, 318–351.
16. Gibrat, R. *Les Inegalités Economiques*; Librairie du Recueil. Sirey: Paris, France, 1931.
17. Santarelli, E.; Klomp, L.; Thurick, R. Gibrat's Law: An overview of the empiric literature. In *International Studies in Entreprenership*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 12, pp. 1–73.

18. Chakrabarti, B.; Chakraborti, A.; Chakrabarty, S.; Chaterjee, A. *Econophysics of Income and Wealth Distribution*; U.K. Cambridge University Press: Cambridge, UK, 2013.
19. Richmond, P.; Hutzler, S.; Coelho, R. A review of empirical studies and Models of Income Distributions in Society. In *Econophysics and Sociophysics. Trends and Perspectives*; Chakrabarti, B.K., Chakabrorti, A., Chaterjee, A., Eds.; Wiley-VCH: Berlin, Germany, 2016.
20. Atkinson, A.B. The distribution of top incomes in the United Kingdom 1908–2000. In *Top Incomes over the 20th Century*; Atkinson, A.B., Piketty, T., Eds.; Oxford University Press: Oxford, UK, 2007.
21. Yakovenko, V.M.; Rosser, J.B. Colloquium: Statistical mechanics of money wealth and income. *Econ. J.* **2013**, *81*, 703–725. [CrossRef]
22. Aoyama, H.; Souma, W.; Nagahara, Y.; Okazaki, M.P.; Takayasu, H.; Takayasu, M. Pareto's law for income of individuals and debt of bankrupt companies. *Fractals* **2000**, *8*, 293–300. [CrossRef]
23. Aoyama, H.; Souma, W.; Fujiwara, Y. Growth and fluctuations of personal and company's income. *Physica A* **2003**, *324*, 352–358. [CrossRef]
24. Aoyama, H.; Fujiwara, Y.; Fujiwara, Y. *Econophysics and Companies: Statistical Life and Death in Complex Business Networks*; Cambridge University Press: New York, NY, USA, 2011.
25. Shirras, F. The Pareto Law and the Distribution of Income. *Econ. J.* **1935**, *45*, 663–681. [CrossRef]
26. Atkinson, A.B. Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Econ. Spec. Issue Inequality* **2017**, *84*, 129–343. [CrossRef]
27. Benhabib, J.; Bisin, A. *Skewed Wealth Distributions: Theory and Empirics*; NBER Working Paper 21924; National Bureau of Economic Research: Cambridge, MA, USA, 2016.
28. Kleiber, C. Dagum vs. Singh-Maddala income distributions. *Econ. Lett.* **1996**, *53*, 265–268. [CrossRef]
29. Bandourian, R.; McDonald, J.B.; Turkey, R.S. A comparison of parametric models of income distribution across countries and overtime. *Estadística* **2003**, *55*, 135–152.
30. Montroll, E.W.; Schelesinger, M.F. On 1/f noise and other distributions with long tails. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3380–3383. [CrossRef]
31. Hogg, R.V.; Mckean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*; Pearson Education: Delhi, India, 2007.
32. Chatterjee, A.; Chakrabarti, B.K. Kinetic exchange models for income and wealth distributions. *Eur. Phys. J. B* **2006**, *60*, 135–149. [CrossRef]
33. Cowell, A.; Flachaire, E. Statistical Methods for Distributional Analysis. In *Handbook of Income Distribution*; Elsevier: Amsterdam, The Netherlands, 2015.
34. Chotikapanich, D.; Griffiths, W. Estimating income distributions using a mixture of gamma densities. In *Modelling Income Distributions and Lorenz Curves*; Chotikapanich, D., Ed.; Springer: New York, NY, USA, 2008; Chapter 16, pp. 285–302.
35. Pittau, M.G.; Zelli, R. Empirical evidence of income dynamics across EU regions. Italian evidence in the 1990s from kernel density estimates. *Empir. Econom.* **2006**, *29*, 415–430. [CrossRef]
36. AEAT 2020. Available online: www.aeat.es (accessed on 8 February 2021).
37. González, I.; Mateos, A. La distribución de la riqueza. Pareo deconstruido, Gibrat reconstruido. *Rev. Econ. Apl. Número Extraordin. Econophys.* **2019**, *37*, 22–40.
38. Kleiber, C.; Kotz, S. *Statistical Size Distributions in Economics and Actuarial Sciences*; John Willey and Sons: Hoboken, NJ, USA, 2003.
39. Mandelbrot, B.B. The Pareto Levy law and the distribution of income. *Int. Econ. Rev.* **1960**, *1*, 79–106. [CrossRef]
40. Arnold, B.C. *Pareto Distributions*, 2nd ed.; CRC Press, Taylor y Francis Group: Boca Raton, FL, USA, 2015.
41. Richmond, P.; Solomon, S. Stable power laws in variable economics; Lotka-Volterra implies Pareto-Zipf. *Eur. Phys. J. B* **2002**, *27*, 257–261. [CrossRef]
42. Limpert, E.; Stahel, W.A.; Abbt, M. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience* **2001**, *5*, 341–352. [CrossRef]
43. Koch, A.L. The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. *J. Theor. Biol.* **1966**, *23*, 276–290. [CrossRef]
44. Koch, A.L. The logarithm in biology. II. Mechanisms generating the log-normal distribution exactly. *J. Theor. Biol.* **1966**, *23*, 251–268. [CrossRef]
45. Coelho, R.; Neda, Z.; Ramasco, J.; Santos, A. A family-network model for wealth distribution in societies. *Physica A* **2005**, *353*, 515–528. [CrossRef]
46. Gabaix, X. Power laws in economics and finance. *Annu. Rev. Econ.* **2009**, *1*, 255–294. [CrossRef]
47. Montebruno, P.; Bennet, R.J.; Van Lieshout, C.; Smith, H. A tale of two tails: Do power Law and Lognormal models fit firm size distribution in the mid-Victorian Era? *Physica A* **2019**, *523*, 858–865. [CrossRef]
48. Solomon, S.; Richmond, P. Power laws of wealth, market order volumes and market returns. *Physica A* **2001**, *299*, 188. [CrossRef]
49. Pareto, V. *Cours d'Economie Politique*; Trad. "Manual of Political Economy"; Macmillan: Paris, France; London, UK, 1897; Volume 2.
50. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser.* **1977**, *39*, 1–38.
51. Brindusa, A.; Basso, H.; Bover, O.; Casado, J.M.; Hospido, L.; Izquierdo, M.; Kataryniu, I.; Lacuesta, A.; Montero, J.M.; Vozmediano, E. La desigualdad de la renta el consumo y la riqueza en España. *Doc. Ocas. Banco Espa.* **2018**, *6*, 1–49.
52. Chescher, A.; Schluter, C. Measurement error and inequality measurement. *Rev. Econ. Stud.* **2002**, *69*, 357–378. [CrossRef]

53.　Atkinson, A.B.; Piketty, T.; Saez, E. Top incomes in the long run oftheirtory. *J. Econ. Lit.* **2011**, *49*, 3–71. [CrossRef]

54.　Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman & Hall/CRC: London, UK, 1986.

55.　Garcia Docampo, M. Medición del análisis de las desigualdades en la distribución de la renta. *Empiria. Rev. Metodol. Las Cienc. Soc.* **2000**, *3*, 73–99.

56.　Gut, C.; Limpert, E.; Hinterberger, H. A Computer Simulation on the Web to Visualize the Genesis of Normal and Log-Normal Distributions. 2000. Available online: http://stat.ethz.ch/vis/log-normal (accessed on 8 February 2021).