



## SAFER Methodology: A Proposal for the Identification of Family Firms in Spain Based on the SABI Database

Jose Carlos Casillas<sup>1\*</sup>, Alejandro Escribá-Esteve<sup>2</sup>, María Elena Gómez-Miranda<sup>3</sup>, M. Concepción López-Fernández<sup>4</sup>, Daniel Lorenzo-Gómez<sup>5</sup>, Ignacio Requejo<sup>6</sup>, Alfonso A. Rojo-Ramírez<sup>7</sup>

<sup>1</sup>Universidad de Sevilla, <sup>2</sup>Universitat de València & Ivie, <sup>3</sup>Universidad de Granada, <sup>4</sup>Universidad de Cantabria, <sup>5</sup>Universidad de Cádiz, <sup>6</sup>Universidad de Salamanca, <sup>7</sup>Universidad de Almería

Research article. Received: 2024-02-07; accepted: 2024-05-13

JEL CODE  
L16, L22, M21

KEYWORDS  
Family Firm, Identification, Definition, Employment, Gross Value Added

CÓDIGO JEL  
L16, L22, M21

PALABRAS CLAVE  
Empresa Familiar, Identificación, Definición, Empleo, Valor Añadido Bruto-miliares, Valor del trabajo

**Abstract** Although family firms represent a large percentage of all companies worldwide, methodological contributions regarding their identification are very scarce. This paper offers a new methodology (labelled as “SAFER methodology”) for identifying family versus non-family firms using the SABI database (Spanish version of Orbis) as a source of information, primarily considering ownership and corporate governance variables. The SAFER methodology can be applied without the need for massive data extraction, which contributes to its accessibility and ease of use by researchers. Furthermore, the proposed methodology has been tested using a sample of 500 companies, with a classification error of less than five percent. This methodology has also been discussed and assumed by the IEF (*Instituto de la Empresa Familiar*), becoming standard in Spain. The value-added of this work is to provide researchers with a widely accepted strategy for the selection of family firm samples. The methodology can also be used with equivalent databases that cover other specific countries (e.g., AIDA for Italian company data or Markus in Germany) or that have an international coverage (e.g., Orbis). Our results represent an important step forward in facilitating the work of professionals and policy makers in the development of reports on family businesses through a rigorous process of identification and classification.

**Metodología SAFER: Una propuesta de identificación de las empresas familiares en España a partir de la base de datos SABI**

**Resumen** A pesar de que las empresas familiares representan un gran porcentaje del conjunto de compañías a nivel mundial, las aportaciones metodológicas para identificarlas son muy escasas. El presente trabajo ofrece una metodología (denominada “metodología SAFER”) de identificación de empresas familiares versus no familiares utilizando como fuente de información la base de datos SABI (versión española de Orbis), considerando fundamentalmente variables de propiedad y de gobierno corporativo. La metodología SAFER permite su aplicación sin necesidad de realizar una extracción masiva de datos, lo que contribuye a su accesibilidad y facilidad de uso por parte de los investigadores. Asimismo, la metodología propuesta ha sido contrastada utilizando una muestra de 500 empresas, con un error de clasificación inferior al cinco por ciento. El valor añadido de este trabajo es ofrecer a los investigadores una estrategia ampliamente aceptada para la selección de muestras de empresas familiares. La metodología también puede utilizarse con bases de datos equivalentes en otros países (p.ej., AIDA para datos de empresas italianas o Markus en Alemania) o bases de datos que tienen una cobertura internacional (p.ej., Orbis). Nuestros resultados representan un avance importante al facilitar la labor de los profesionales y responsables políticos en el desarrollo de informes sobre empresas familiares a través de un proceso riguroso de identificación y clasificación.

<https://doi.org/10.24310/ejfb.14.1.2024.18965>

**Author contribution:** Authors contributed equally to the work

Copyright 2024: José Carlos Casillas, Alejandro Escribá-Esteve, María Elena Gómez-Miranda, M. Concepción Lopez-Fernández, Daniel Lorenzo-Gómez, Ignacio Requejo, Alfonso A. Rojo-Ramírez

European Journal of Family Business is a Diamond Open Access journal published in Malaga by UMA Editorial under CCBY-NC-ND license. ISSN 2444-8788 ISSN-e 2444-877X

\*Corresponding author:

E-mail: Casillas@us.es

## 1. Introduction

The study of family firms (FFs) has evolved rapidly in recent decades (Evert et al., 2016; Short et al., 2016), but there are still some important unanswered questions such as the estimation of more accurate numbers on the relevance of FFs in national economies based on the existing databases. Several academic papers have attempted to address this question. A first group of studies have used data from listed firms (e.g., Claessens et al., 2000; Faccio & Lang, 2002; La Porta et al., 1999; Villalonga & Amit, 2010), but these works offer only a limited view of the real importance of FFs. There have also been some attempts to obtain a global cross-national idea of the relevance of FFs (IFERA, 2003), although aggregating data of diverse nature and reliability comes with its own challenges. Finally, a smaller number of academic papers have tried to obtain a comprehensive estimate of the relevance of FFs in specific economies. In particular, we find the fundamental work of Shanker and Astrachan (1996), updated in Astrachan and Shanker (2003), for the case of the U.S. economy. Other works have similarly attempted to measure the presence of FFs in the United States (e.g., Chang et al., 2008). In Europe, Bjuggren et al. (2011) estimate the relevance of FFs for the Swedish economy. In addition, there has been some projects launched at governmental or institutional level (e.g., Flören et al., 2010) including the action “Statistics for family businesses” carried out with the support of the Program for the Competitiveness of Small and Medium-sized Enterprises (2014-2020) (COSME) (European Commission) by seven European countries (Bulgaria, Denmark, Finland, Italy, Malta, the Netherlands, and Poland). The aim of this action has been to identify FFs in order to assess their relevance, scope and nature.

Three main reasons explain the lack of accurate estimates of FFs' importance at an economic level (Astrachan & Shanker, 2003; Bjuggren et al., 2011; Chang et al., 2008; Shanker & Astrachan, 1996). First, there is lack of a generally accepted definition and operationalization of FF (European Commission, 2009; 2015; Sarkar et al., 2014). Second, we do not have any government statistics or secondary data sources that directly identify FFs as such, whatever the definition used (Bjuggren et al., 2011; Chang, et al., 2008). Third, and related to the latter, there is no generally accepted methodology for calculating indicators that measure the weight of FFs across countries (Franks et al., 2015).

These are the reasons why previous works employ different methodologies adapted to the type of data available in each country. In the case of the United States, data from various sources (e.g.,

various statistics and censuses) have been used, which have allowed scholars to calculate the contribution of FFs to gross domestic product (GDP) and employment (Astrachan & Shanker, 2003; Shanker & Astrachan, 1996). Survey data (Chang et al., 2008) have also been used to estimate the relationship between economic development and the prevalence of FFs. In Sweden, data collected by the Swedish government to identify and tax business owners, in addition to data from listed companies, enable researchers to calculate the contribution of FFs to the country's GDP and total employment (Bjuggren et al., 2011). Flören et al. (2010), in their report for the Dutch Ministry of Economic Affairs, applied a telephone survey to a representative sample of Dutch companies and identified FFs using the GEEF (European Group of Owner-Managed and Family Enterprises) and FBN (Family Business Network) definition of a family business, which follows the recommendation of the European Commission (2015). Finally, the seven European countries involved in the Statistics for Family Businesses Action have used the GEEF/FBN definition but operationalized it in different ways (ownership being the main used criterion); the authors of this report follow different measurement approaches to identify FFs, analyze their geographical and/or sectoral distribution, identify their characteristics and, in most cases, assess their economic importance.

In Spain, the Instituto de la Empresa Familiar (IEF), with the invaluable support of the existing Network of Family Business Chairs in most Spanish universities, promoted a study to directly estimate the reality of FFs in Spain in general and in each of the regions (i.e., autonomous communities) in particular (Casillas et al., 2015). The report had several objectives. Firstly, the authors wanted to estimate the weight of Spanish companies in the national economy, in terms of gross value added (GVA) and employment, considering the whole population of Spanish enterprises and using a data source that is relevant, rigorous, systematic and regularly updated, even if not provided by the government. Secondly, the study identified the defining characteristics of FFs, broadening the knowledge of business and family management. This IEF study was based on the SABI (Sistema de Análisis de Balances Ibéricos) database, an extensive version of Orbis for Spain, developed by Informa-Bureau van Dijk. The 2015 report promoted and conducted by the IEF, which took the pioneering work by Rojo-Ramírez et al. (2011) in Spain as a reference point, represents a relevant qualitative step in the estimation of the importance of Spanish family businesses.

This study, almost ten years later, aims to improve the methodology used in the work by Rojo-Ramírez et al. (2011), incorporating the ideas de-

veloped by other researchers who have used the same data source for similar purposes (Arosa et al., 2010; Diéguez-Soto et al., 2015; López-Delgado & Diéguez-Soto, 2015). The current paper is the result of the work conducted by a commission created within the *Spanish Academy of Family Enterprise Researchers* (SAFER) who has been collaborating for almost a year on the development of the present methodology. We recognize that the proposal we are making is not perfect and we are aware that there could be classification errors in both possible directions (i.e., classifying truly family companies as non-family and vice versa, classifying truly non-family companies as family). These errors are due, on the one hand, to the fact that there is no clear boundary between the two types of companies, with firms that could be put in a kind of "gray zone", and, on the other hand, because the quality of the information available in SABI is not always the desirable one. However, these errors are less than five percent, according to tests carried out on a sample of 500 companies, which are detailed in Section 5. Our methodology has also been discussed and adopted by the IEF (*Instituto de la Empresa Familiar*), which represents the majority of Spanish family businesses, thus becoming standard in Spain.

The SAFER methodology will contribute to the development of studies on family businesses, helping both researchers that wish to conduct academic research, and consultants and business managers, as well as professionals and policy makers, in the preparation of reports on FFs through a rigorous process of identification and classification. The SAFER methodology provides simple, objective and rigorous criteria to distinguish family from non-family businesses, and can be applied without the need for massive data extraction, which contributes to its accessibility and ease of use by researchers and other potential users. This carefully thought and designed methodology for the selection of samples of FFs can be transferred to similar databases in other countries (such as AIDA in Italy or Markus in Germany) or applied with the global Orbis database. In this sense, despite being developed for a country such as Spain, the SAFER methodology has broader applicability and its potential goes beyond the national scope in which it has been formulated and tested.

## 2. Identification of Spanish Family and Non-Family Businesses

### 2.1. Source of information and analysis process For the development and application of the FF

identification methodology (from now on, the SAFER methodology), the SABI data source developed by Informa-Bureau van Dijk has been used, which can be regarded as the Spanish version of the global Orbis database. SABI contains the firms' financial statements deposited in the Commercial Registry for more than three million companies. It is therefore information at the company level, with both quantitative and qualitative historical data. Orbis (and SABI) is a reliable and widely used database (Ahmad et al., 2018; Diéguez-Soto et al., 2015; Rojo-Ramírez et al., 2011), although it has some biases that are worth mentioning (Martínez-Romero & Rojo-Ramírez, 2017; Pindado & Requejo, 2015). In particular, its coverage is higher in companies with at least 10 employees, with Spain (SABI) being one of the countries with a better coverage rate internationally (Bajgar et al., 2020). Specifically, SABI incorporates information on approximately one million active companies with financial and qualitative data (1998-2022), of which more than 120,000 are companies with at least 10 employees in 2021.

Based on this source, we propose a process divided into four stages to estimate the weight of FFs in the Spanish economy: (i) selection of the population of firms with available data; (ii) classification of these firms into family and non-family; (iii) extrapolation of the data resulting from the classification to the rest of the firms; and (iv) segmentation of FFs into different types, thus considering the widely accepted heterogeneity among FFs (Arteaga & Basco, 2023).

These four stages are developed in the following sections. The application of the SAFER methodology to a set of companies of any type and legal form (trading companies, cooperatives, listed companies, etc.) will result in a classification into three groups: family, non-family and unclassified (due to lack of information). The treatment of companies according to their ownership structure will give rise to a specific analysis of some firms, such as listed corporations, which is discussed in Section 2.3.4. The fourth stage, which consists of the segmentation of family companies, makes it possible to capture their heterogeneity, according to criteria such as size, sector, geographical scope, age, ownership structure or corporate governance structure. The result will be a double list of family and non-family companies that can be classified by different criteria, thus considering their heterogeneity.

### 2.2. Selection of the population of companies with available data

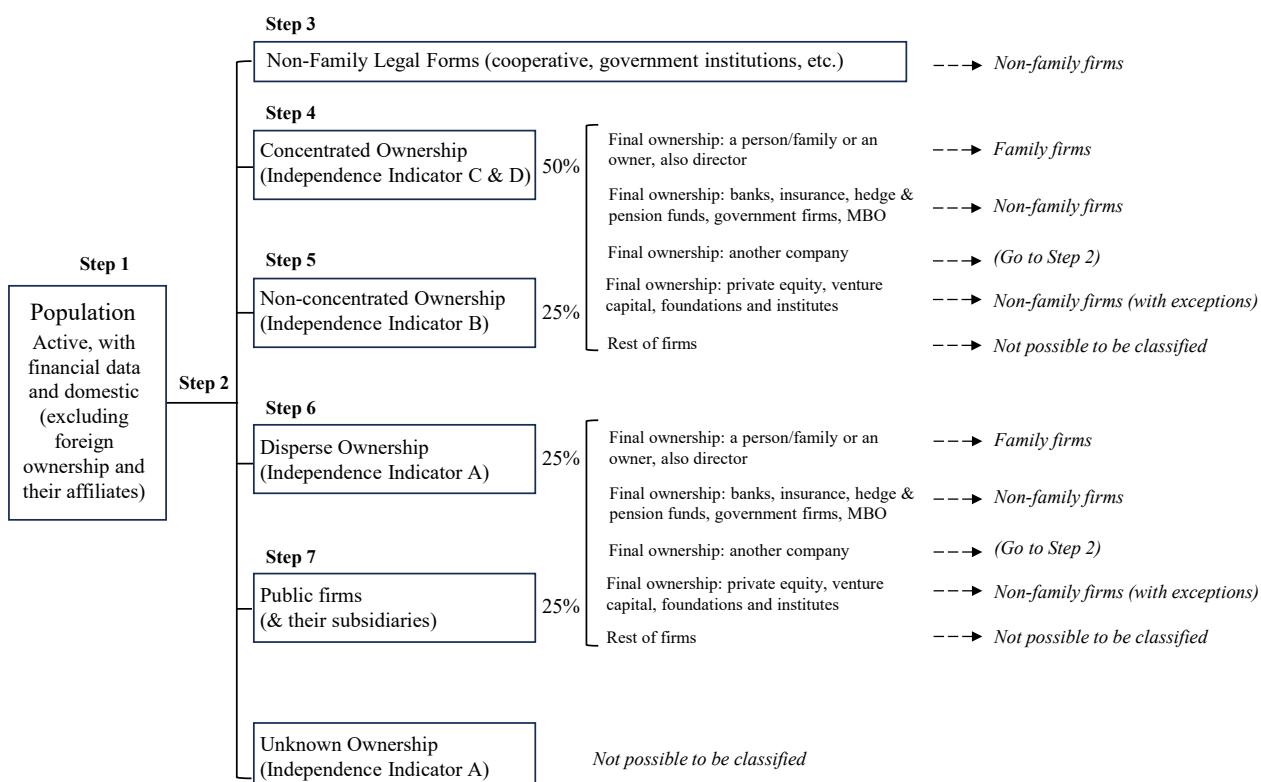
As previously mentioned, the source of information to be used is SABI. Despite having a very

broad coverage, the incorporation of data in the database is not homogeneous for all the firm variables it includes. Being aware of this limitation, the first step in the proposed methodology consists of identifying the population of companies to be classified into two large groups: family and non-family.

In this sense, we propose to establish a series of initial filters, summarized in Figure 1. The first of these refers to the status of the company, i.e., whether the company is active or not. Not being active implies that the company is in a process of

suspension of payments, bankruptcy, insolvency proceedings, closure of its registration, extinct, untraceable, etc. Inactive companies (e.g., due to dissolution, extinction or ex officio deregistration) are not considered. This information is updated by SABI on a regular and prompt basis as it does not depend on the registration of the annual accounts (in other words, the “active/inactive” status does not suffer from the usual delay derived from the filing and recording of the annual accounts in the database, which takes approximately two years).

**Figure 1.** Process for classifying family and non-family businesses



The second initial criterion refers to the availability of annual accounts in SABI. To this aim, we include all companies that have submitted their annual accounts, whether consolidated (C1), non-consolidated (U1) or both types of accounts (C2/U2). In short, the objective is to exclude those companies whose accounts have not been filed or are not available.

The third delimitation criterion consists of the exclusive consideration of family companies whose ultimate owner is Spanish. The aim is to remove

foreign-owned companies for which SABI does not present sufficient information. To this end, the criterion established is the exclusion of those companies in which a foreign parent company is the shareholder (definition of parent company: minimum percentage of the chain of control from the focal company to its parent company greater than 50%)<sup>1</sup>.

Finally, in order to ensure the availability of information in the annual accounts, we recommend to impose the requirement that the company has

1. For more information, see the SABI Help Guide. It is available at [https://help.bvdinfo.com/mergedProjects/122\\_ES/Home.htm](https://help.bvdinfo.com/mergedProjects/122_ES/Home.htm)

available data on general accounting items in the years to be analyzed as a fourth criterion. As a general rule, a good criterion may be to require operating income to have a minimum value of, for example, 10,000€ and, in any case, always greater than zero. The more historical years are requested and the higher the amount of accounting data (e.g., operating income), the higher and better the quality of the data will be, since

larger companies provide higher quality information ([Bajgar et al., 2022](#)). In this sense, it should be noted that the use of these filters could bias the sample towards companies that provide better information and over a longer period of time. Table 1 shows the degree of representativeness of SABI by employee stratum. It can be seen that the database covers almost all companies with more than 10 employees, but far fewer of those with less than 10 employees.

**Table 1.** Comparison between data available in SABI and the DIRCE business census (Spanish *Office for National Statistics*, INE)

Type of company (by size, considering the number of employees)	Number of companies in the INE census	Number of companies present in SABI	Number of active companies with financial data present in SABI
Without employees	1,942,319	0	0
Micro-firms (2-9 employees)	1,340,792	376,602	368,713
Small firms (10-49 employees)	122,838	100,592	99,414
Medium-sized firms (50-249 employees)	19,994	15,659	15,374
Large firms (250 or more employees)	4,720	3,299	3,266

Note 1: Data from year 2022.

Note 2: In this work, it has been assumed that the most appropriate criterion for classifying companies by size is the number of employees. However, an approximation to size could be adopted based on the value of operating income, which would allow us to start from a larger initial population.

### 2.3. Classification of companies into family and non-family businesses

Once the initial filters have been established to ensure the quality and reliability of the information to be used, we begin considering the legal form of the companies as a criterion for identifying non-family firms. We automatically classify some specific legal forms as non-family companies. The most relevant group is the one formed by cooperatives, to which we must add other less frequent types such as professional associations and mutual funds, among others.

Once this step has been completed, we address one of the fundamental issues in establishing a practical process for identifying FFs. Companies with a highly dispersed ownership structure and those with a highly concentrated structure should be examined separately. We do not believe it is

appropriate to apply the exact same criteria to all companies. In this regard, SABI provides interesting information on the ownership structure of companies through what it calls the “independence indicator”. According to this indicator, it is possible to differentiate between three major groups of companies<sup>2</sup>:

*Companies with a concentrated ownership structure:* these companies are classified in SABI with an independence indicator C (companies with a registered shareholder with a total percentage or calculated total ownership percentage greater than 50%) and D (an identified shareholder with direct ownership greater than 50%).

*Companies with a non-concentrated ownership structure (independence indicator type B):* these companies have known shareholders without an ownership percentage (direct, total or calculated total) of more than 50%, but with one or more

2. In addition to the indicators listed below, there is the U indicator, which refers to companies with insufficient information on their ownership structure.

shareholders with an ownership stake of more than 25%.

*Companies with a dispersed ownership structure (independence indicator type A):* in these companies, there is no single shareholder accumulating more than 25% of ownership, direct or total.

### 2.3.1. Classification of companies with concentrated ownership

Companies with concentrated ownership are those in which there is a single ultimate shareholder (non-group) that holds at least 50.01% of the ownership. These companies are classified by SABI with the independence indicator codes D (known shareholder with a direct shareholding of over 50%) and C (registered shareholder with a total shareholding - known or calculated - of over 50%). It should be clarified that the ultimate owner is being considered and not the direct shareholder, i.e., we account for the existence of a chain of control through intermediate companies between the company to be classified and its final owner.

When classifying a company with concentrated ownership as family-owned, we propose to consider the combination of two criteria. The first is that the ultimate owner of the company (maintaining 50.01% as the ownership control threshold) is of the type “a single person or family” (in SABI: Financial ties. owned by a parent. (50%): one or more individuals or families). However, companies in which a shareholder is also a director who owns at least 50.01% of the company shares should be considered as FF as well. Both criteria are combined with the operator “or” in the “Boolean search” of the SABI search strategy, which in practice implies the combination of both sets of companies.

In a second step, we proceed to identify those companies that can be classified as “non-family” using a criterion similar to the previous one. To this end, we propose to consider as non-family companies those in which the ultimate owner (owning at least 50.01% of the shares) is one of the following types: “banks and financial companies”, “insurance companies”, “hedge funds”, “investment and pension funds / nominees / trusts / trustees”, “public authorities, states and governments” and “employees / administrators / directors (management buy outs)”.

In addition to the previous two groups, a third group to be classified consists of those companies whose ultimate owner is another company (always based on a chain of control of more than 50.01%). This set of companies requires a more detailed analysis, which will be explained in Section 2.4.

Finally, there are three other possible types of ultimate owners, according to the classifica-

tion provided by SABI, which are “private equity firms”, “venture capital” and “foundations/research institutes”. These three groups include a small percentage of cases and are difficult to classify automatically. In these groups, it is possible to find some foundations related to a family business, as well as some family offices, although in general terms, they should be classified as non-family, given their nature and objectives. For this reason, we propose to classify them manually using other criteria that will be detailed in Section 2.5.

### 2.3.2. Classification of companies with non-concentrated ownership

We label all firms in which the main ultimate owner holds between 25.01% and 50% of the shares as companies with non-concentrated ownership. In other words, no single shareholder ultimately owns more than half of the company, but there are owners with significant stakes, who can be assumed to exercise effective control over the company. SABI classifies these companies with the type B independence indicator. For these companies, we apply a process similar to the previous one, although the threshold required in the chain of control is reduced to 25.01% in this case. Thus, we propose to classify as family companies with non-concentrated ownership those whose ultimate owner is of the “one or more individuals or families” type. Moreover, following a process similar to the one described above, we also consider as family companies those with a shareholder who is also a director and who additionally owns a minimum of 25.01% of the shares and less than 50%. Finally, after various tests, we have found that this process would exclude companies in which there is a direct shareholder who owns more than 50% of the company and falls in the “individual and family” category. Therefore, together with the two previous criteria, we incorporated a third criterion to be added (operator “or”) to the previous ones as follows: there is a direct shareholder with more than 50.01% of the ownership in the hands of a shareholder of the “one or more individuals or families” type. The difference between this criterion and the two previous ones is that we refer to the direct shareholder, not the ultimate owner (in which case there are intermediate companies), nor do we consider the participation in management. Following the same process as for closely held companies, the next step is to identify companies that can be classified as “non-family”. The proposal is to follow the same criteria as above, only modifying the ownership control threshold, now set at 25.01% (“banks and financial companies”, “insurers”, “hedge funds”, “investment and pension funds / nomenclatures / trusts /

trustees”, “public authorities, states and governments” and “employees / administrators / directors (management by outs)”.

Within the group of companies that we analyze, there are those in which, having a non-concentrated capital, the largest shareholder is another company. Once again, we find a group of firms that are difficult to classify, insofar as they are firms owned by a parent company (also a company), with an ownership control of more than 25% but less than 50%. For these companies, we suggest a specific analysis that will be detailed in Section 2.4.

Finally, a small group of companies remains unclassified after the previous classification process. They are firms in which the ultimate owner (considering the aforementioned interval of between 25.01% and 50%) is one of the following types: “private equity companies”, “venture capital” and “foundations/research institutes”. As mentioned above, the companies belonging to this group should be studied according to other criteria on a case-by-case basis, although, in general terms, they could be classified as non-family.

### *2.3.3. Classification of companies with dispersed ownership*

Companies in which there is no ultimate owner with more than 25% of the capital are classified by SABI with the type A independence indicator. When considering FFs in this group of companies, we cannot identify their ultimate owner given that the shareholdings are too low (below 25%). However, it is possible to identify companies in which there is a direct shareholder of the “one or more individuals or families” type that owns a minimum of 50.01% of the shares. These may be classified as family companies.

Once these are classified, we propose to consider as non-family those companies in this group in which at least 50.01% of their direct shareholding is a shareholder of the type “banks and financial companies”, “insurance companies”, “hedge funds”, “investment and pension funds / nominees / trusts”, “public authorities, states and governments” and “employees / administrators / directors (management buy outs)”. Similarly, when the largest direct shareholder of a company with an A independence indicator is another company, with an ownership control of more than 25%, it is necessary to conduct a more in-depth analysis of the parent company, in order to determine whether it is a family or non-family company (see Section 2.4). Finally, as in the previous cases, a small group of companies will remain at the end of this classification process, which are those whose ultimate owner is not one of the types mentioned above, so that they need to be analyzed individually (see Section 2.5).

### *2.3.4. Classification of listed companies and their subsidiaries*

In the case of listed companies, in which ownership is usually much more dispersed, our proposal incorporates the definition of the GEEF/FBN for this type of company, according to which listed companies could be considered family-owned if the person who founded or acquired the company, or his or her relatives or descendants, hold 25% of the voting rights deriving from their participation in the capital. Following this recommendation, we classify a listed company as family-owned when it meets one of the following three conditions: (1) it is owned by a parent company (ultimate owner) of the “one or more individuals or families” type with a shareholding of more than 25.01%; (2) a managing shareholder owns at least 25.01% of the shares; or (3) there is a direct shareholder of the “one or more individuals or families” type and they own at least 25.01% of the shares.

On the other hand, we will classify as “non-family” those listed companies in which either their ultimate parent company or their direct shareholder (with an ownership of more than 25%) falls in one of the following types: “banks and financial companies”, “insurance companies”, “hedge funds”, “investment and pension funds / nominees / trusts”, “public authorities, states and governments”, and “employees / administrators / directors (management buy outs)”.

In the case of those listed companies that are actually firms owned by another company (more than 25% ownership threshold), it will be necessary to analyze the latter. To properly classify the remaining firms, a more detailed and individual analysis is required.

Finally, we still need to analyze companies that are owned by listed firms. It must be taken into account that listed companies are usually large groups that control other companies. Our proposal consists of classifying the investees of listed companies (with control greater than 50.01%) in the same group as their parent company, so that we will regard the subsidiaries of family companies as family-owned and vice versa.

### *2.3.5. Companies in which their ownership structure is unknown*

SABI assigns a type U independence indicator to all companies in which their ownership structure has not been identified. In this case, it does not seem to make sense to try to analyze them manually, since there is simply not enough information. For these companies, our proposal is to exclude them from the analysis when the objective is simply the identification or classification as family or non-family.

On the contrary, if the objective of the analysis

goes beyond identification/classification and you wish to estimate the weight of FFs in the country's economy based on certain variables such as employment, GDP, etc., our proposal consists of conducting a process of extrapolation of the data of these remaining companies using the information on the previously classified firms. However, this extrapolation process must be implemented considering the greatest possible number of identifying variables related to the demographic characteristics of the firms. In Section 2.5, we make a specific proposal for the extrapolation process.

#### **2.4. Classification of companies owned by another company (parent company)**

In the previous steps, we have not been able to classify a set of companies because their known ultimate owner, according to a certain ownership threshold (25% or 50%), is another company. In this case, the proposal we make involves carrying out a new analysis of the ultimate owner (company), according to the same criteria proposed in the previous steps. Thus, those companies whose parent company has been assigned an independence indicator A will be classified as family or non-family in line with the criteria set out for this type of company; a similar approach will be followed for companies with an independence indicator of B, C, D and U. Once the parent company has been classified as family, non-family or unclassifiable, the initial investee company will also be put in the same category.

In this second classification stage we have just explained, there could be some companies whose ultimate owner is again a company; this could lead to a new round of classifications, which could be repeated indefinitely. In order to maintain some parsimony in the process, we understand that new rounds will not lead to significant improvements, making the implementation of the classification process overly complicated. In this case, and after reviewing a sample of companies that fit in this situation, we propose that companies whose ultimate owner is another company in the second round should be classified as non-family.

#### **2.5. Individual classification of companies with complex structures**

In the processes described above, it is really risky to classify some companies as family or non-family, not due to lack of information but because they have complex ownership structures (circular participations, extreme fragmentation of shareholders, etc.). In these cases, we propose to carry out an individual analysis of the companies since the number of cases is manageable. We suggest to basically analyze two aspects: (1) their ownership structure and (2) their adminis-

trators/board of directors.

Regarding the analysis of shareholders, a company should be considered as family when shareholders with the same or similar last names have effective control of the business. By effective control we mean that the sum of the shares of shareholders with the same last name exceeds that of any other individual shareholder. Regarding the similarity of last names, the aim is to see if a pattern can be identified that allows us to reasonably assume that they are members of the same family. In Spain, there is the advantage that people use two family names, which makes this process easier and more precise.

When the previous step does not give clear results, a second analysis consists of focusing on the company's administrators (board of directors). In this case, we will classify a company as family when the majority of the directors appear to have a family connection, again based on their last names. Likewise, if those who appear to be members of the same family are not the majority but are in turn owners of a significant percentage of the business and hold the highest responsibility position (e.g., president, CEO), we propose to classify the company as family.

Finally, the rest of the companies will be classified as "non-family" or otherwise as "not classifiable" if there is insufficient information to classify them.

### **3. Extrapolation**

The SABI database has a very broad coverage in relation to the financial information of companies, but coverage is relatively lower when it comes to the ownership structure data and the firm governance structures. For this reason, a significant percentage of companies lack reliable and complete information regarding these aspects, making it virtually impossible to classify them, not even with an acceptable error margin. This situation is not too relevant if the objective is to classify companies, but it is important when it comes to estimating the contribution of FFs to a country's economy. To this goal, we must add the contribution of all companies, family and non-family, so it is necessary to also make an estimate of those companies that could not be classified following the previous steps.

Our suggestion consists of making an estimate of the rest of the unclassified companies based on an extrapolation process by using the information of the previously classified companies. In this sense, it is interesting to apply a method similar to the one proposed by [Raghunathan et al. \(2001\)](#) for estimating missing values. In our case, we seek to estimate the classification of companies into two groups; for example, think of a dummy

variable as follows: 1 = family; 0 = non-family. It would be feasible to use a logistic regression model that allows us to estimate parameters based on known variables for the vast majority of companies, for those already classified and for those that cannot be classified (Crespi-Cladera et al., 2016). We recommend to first estimate the parameters with the population of companies already classified, leaving a sub-sample of companies that allow the predictive validity of the model to be verified.

There are two alternative ways for the selection of the variables. The first is to work only with data included in SABI. In this case, the list of variables could include, among others: employees, assets, operating income, sector (dummy variables), etc. After the estimation of the predictive model, the estimated parameters can be used to classify companies that could not be classified in the previous steps. An alternative way would be to try to apply the model to the group of companies in the DIRCE (*Central Directory of Companies*) elaborated by the INE (Spanish Office for National Statistics). In this case, the model could only include variables included in the DIRCE for the segmentation of companies: employee stratum, legal form and sector. Although this option might reduce the explanatory power of the estimation model, it allows the model to be applied to all existing companies.

#### 4. The Heterogeneity of Family Businesses

Family businesses do not constitute a homogeneous group. There is currently high agreement in assuming the heterogeneity of family businesses (Chua et al., 2012; Daspit et al., 2021; Jaskiewicz & Dyer, 2017; Neubaum et al., 2019). Therefore, beyond the possible media interest of having a global percentage of FFs, what is really relevant is to identify different types of FFs, giving the possibility of making comparisons between them. In this sense, we must recognize that, in this work, we will only refer to those variables that allow family businesses to be segmented using the SABI database in a relatively acceptable way when working with the set of companies it contains; that is, we do not contemplate the possibility of carrying out manual segmentations or analyses company by company. With the options currently offered by SABI in its online version, it is not feasible to establish segmentation criteria based on the people who form the firm's ownership structure or who are among the list of administrators/directors. It is only possible to use certain criteria related to the percentages of ownership in the hands of the first shareholder or the type of existing administrative body (sole director versus board of directors).

Specifically, based on the previous considerations, we propose to establish the following segmentation criteria for family businesses identified through the steps detailed above:

**Firm size:** mainly using the number of employees, its turnover and its assets. According to this criterion, it will be possible to differentiate between large, medium, small and micro FFs, making it feasible to compare them with each other and with the corresponding non-family businesses.

**Industry:** according to the CNAE 2009 (*National Classification of Economic Activities*). In this sense, it is possible to establish different groups of companies, simply based on the large branches of activity to which they belong - primary, industry, construction, commerce, services - as well as through the analysis of certain specific sectors, for example, sectors intensive in technology or circular economy sectors.

**Geography:** beyond the segmentation of companies by region (i.e., autonomous community) or province, it is interesting to address the differentiation between large cities (or metropolitan areas) versus small municipalities (or rural areas).

**Age as a proxy for the generational stage of the FF:** there are studies that estimate the generation in charge based on the age of the company. Most of them approximately establish a rule that assigns 25 years to each generation (Arrondo-García et al., 2016; Blanco-Mazagatos et al., 2016; López-Delgado & Diéguez-Soto, 2015). However, within the first 25 years, it would be interesting to distinguish between different types to the extent that a newly created company, closer to a start-up, is not comparable to a company with more than two decades, which will be closer to its first generational change. Therefore, we propose to differentiate the following types of companies: family start-ups (< 3 years old); young FFs (between 3 and 10 years old); FFs in the hands of the founder (between 10 and 25 years old); FFs in the hands of siblings (between 25 and 50 years old); FFs in the hands of cousins (between 50 and 75 years old); long-lived FFs (between 75 and 100 years); centennial FFs (> 100 years).

**Ownership structure:** in this sense, it is possible to establish different types of segmentations, among which we propose the following: (1) listed versus unlisted companies; (2) companies with concentrated ownership versus companies with non-concentrated or dispersed ownership; (3) classification based on the firm's legal form (SA versus SL, using the Spanish terminology); (4) different types of family businesses based on the percentage of family ownership.

**Governance structure:** it is possible to differentiate between companies with and without a board

of directors.

## 5. Control of Errors

The evaluation of the quality of the process to classify companies as family and non-family through the SAFER methodology is based on a sample of 500 companies. As in any binary classification (only two categories: family versus non-family businesses), it is possible to make two types of errors, which we will call type 1 error—companies that are really family-owned but that the methodology classifies as non-family—and type 2 error—companies that are really non-family but that the methodology classifies as family. For the quality control, we have followed a double strategy: (a) starting from the real data (companies for which there is external information on their family or non-family nature) and checking whether or not they have been correctly classified or not according to the SAFER methodology; and (b) starting from the classification conducted using the filters of the methodology, we have analyzed each company individually, carefully investigating its ownership and management structure, thus allowing us to check whether the firm has been correctly classified.

For the implementation of strategy (a), 250 companies have been selected: 200 family-owned and 50 non-family-owned. The selection of the 200 family businesses has been carried out through the websites of those family business associations

that provide lists of members. We have considered a proportional distribution of the sample by autonomous community. To this aim, we have started from the total number of companies classified by the SAFER methodology and calculated their distribution by autonomous community. Regarding the 50 non-family companies, since there are no lists, we have turned to companies and entities that are known and whose non-family nature is beyond any doubt. In this case, it has not been possible to maintain geographical proportionality. Regarding strategy (b), we have started from the classification carried out by the SAFER methodology developed in this article and, maintaining the same proportion, we have individually analyzed 200 companies classified as family and 50 as non-family. We have implemented the process of selecting the FFs using a random number extraction method. The same procedure has been followed for non-family businesses.

The result from the application of these two strategies shows an error of approximately 4.4% when using the SAFER methodology (see Table 2). However, it must be noted that the error increases as the size of the companies rises. This phenomenon is fundamentally due to the greater complexity of corporate and ownership structures (intermediate companies that hide the real ultimate ownership of the companies); in light of this finding, our recommendation is to take extra care if you want to/need to work with larger companies.

**Table 2.** Control of errors

		Information extracted from factual data					
		Strategy (a)		Strategy (b)		Joint analysis	
		Family	Non-Family	Family	Non-Family	Family	Non-Family
Methodology	Family	191	9	194	7	385	16
	Non-Family	0	50	6	43	6	93
Methodology	Family	76.4%	3.6%	77.6%	2.8%	77.0%	3.2%
	Non-Family	0.0%	20.0%	2.4%	17.2%	1.2%	18.6%

## 6. Conclusions, Limitations and Future Lines of Research

Although FFs represent a large percentage of all companies worldwide, estimates of their importance in terms of wealth and employment are scarce. In this work, we address the main barriers to identify FFs and measure their economic relevance with the goal of proposing an appropriate methodology for the Spanish case, called

SAFER methodology. The FF definition is based on the proposal of the [European Commission \(2015\)](#) and the data come from the SABI database, which contains information about the financial situation of companies and their corporate governance structures.

Using the information available in the database, we propose a set of automatic filters (based on the definition of the GEEF and FBN) to identify companies as family or non-family. The suggested

methodology has a margin of error of less than five percent, although this error increases with the size of the company.

This study contributes to both academic literature and professional work regarding the classification of large samples of family and non-family businesses, without the need to download the companies from the original source (SABI). We are aware that massive downloads of data would allow for more granular analyses of ownership and governance structures, using, for example, the last names of owners and/or managers to identify family ties (Amore et al., 2023). However, this option entails different types of contractual and operational inconveniences that the SAFER methodology avoids; nonetheless, we recognize that the relative simplicity of the SAFER methodology entails some limitations.

The methodology that we propose expands the options to investigate the relevance and idiosyncrasy of FFs, as well as their heterogeneity. SABI is distributed by Informa-Bureau van Dijk, but this provider also distributes other products such as Orbis (global) that can be used in international studies. The application of the SAFER methodology could be adapted to different international contexts, considering cultural or national specificities, which can open an interesting debate among scholars willing to delve more deeply into the heterogeneity of FFs (Jaskiewicz & Dyer, 2017). Once FFs have been objectively identified through automated filters, further comparisons can be made with non-FFs, looking for differences based on sector, size, performance, business strategies, etc. The availability of data also allows for longitudinal studies. The application of the SAFER methodology facilitates academic research in the family business field by providing simple, objective and rigorous criteria. In addition, the methodology can also be very useful for family business professionals and consultants, as well as policy makers and national and regional associations of family businesses, by making the identification of potential new partners more affordable. In a nutshell, the objectivity that characterizes the SAFER methodology will be useful for new international and longitudinal studies that aim to describe the situation and evolution of FFs. It will be equally helpful to identify different FF typologies based on various criteria that reflect their heterogeneity.

### **Author contribution statement**

*The authors contributed equally to the work.* All the authors have made substantial contributions to the conception or design of the work, drafting the work and reviewing it critically for important

intellectual content. They also gave the final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### **Conflict of interest statement**

There is not competing interest in this research.

### **Ethical statement**

*Not applicable.*

### **Declaration on the use of generative AI in the writing process**

*Not applicable.*

### **Funding**

This manuscript has been financially supported by the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033 (Grants PID2021-126358NB-I00, PID2022-1404130A-I00, PID2022-139222NB-I00 and PID2022-136496NB-I00).

### **Acknowledgment**

The authors thank the IEF (*Instituto de la Empresa Familiar*) and SAFER (Spanish Academy of Family Enterprise Researchers) for the suggestions provided in the process of preparing the article.

### **Data availability statement**

*Not applicable.*

### **References**

- Ahmad, S., Oliver, S., & Peters, C. (2018). *Using firm-level data to compare productivities across countries and sectors: possibilities and challenges.* Economics Working Paper Series, Working Paper 2018-07-A. U.S. International Trade Commission.
- Amore, M. D., D'Angelo, V., Le Breton Miller, I., Miller, D., Pelucco, V., & Van Essen, M. (2023). *Using artificial intelligence to measure the family control of companies.* Electronic copy available at: <https://ssrn.com/abstract=4683794>
- Arosa, B., Iturrealde, T., & Maseda, A. (2010). Ownership structure and firm performance in non-listed firms: Evidence from Spain. *Journal of Family Business Strategy*, 1(2), 88-96. <https://doi.org/10.1016/j.jfbs.2010.03.001>
- Arrondo-García, R., Fernández-Méndez, C., & Menéndez-Requejo, S. (2016). The growth and performance of family businesses during the global financial crisis: The role of the generation in control.

- Journal of Family Business Strategy*, 7, 227-237. <https://doi.org/10.1016/j.jfbs.2016.11.003>
- Arteaga, R., & Basco, R. (2023).** Disentangling family firm heterogeneity: Evidence from a cross-country analysis. *European Journal of Family Business*, 13(2), 162-181. <https://doi.org/10.24310/ejfb.13.2.2023.17638>
- Astrachan, J. H., & Shanker, M. C. (2003).** Family businesses' contribution to the U.S. economy: A closer look. *Family Business Review*, 16(3), 211-219. <https://doi.org/10.1177/08944865030160030601>
- Bajgar, M., Berlingieri, G., Calligaris, S., Criscuolo, C., & Timmis, J. (2020).** Coverage And Representativeness of Orbis Data, OECD Science, Technology and Industry Working Papers, 2020/06, <https://dx.doi.org/10.1787/c7bdaa03-en>
- Bjuggren, C. M., Johansson, D., & Sjögren, H. (2011).** A note on employment and gross domestic product in Swedish family-owned businesses: A descriptive analysis. *Family Business Review*, 24(4), 362-371. <https://doi.org/10.1177/0894486511420138>
- Blanco-Mazagatos, V., de Quevedo-Puente, E., & Delgado-García, J. B. (2016).** How agency conflict between family managers and family owners affects performance in wholly family-owned firms: A generational perspective. *Journal of Family Business Strategy*, 7, 167-177. <https://doi.org/10.1016/j.jfbs.2016.07.003>
- Casillas, J. C. et al. (2015).** *La empresa familiar en España* (2015). IEF y Red de Cátedras de Empresa Familiar. Retrieved from <https://www.iefamiliar.com/publicaciones/la-empresa-familiar-en-espana-2015/>
- Chang, E. P. C., Chrisman, J. J., Chua, J. H., & Kellermanns, F. W. (2008).** regional economy as a determinant of the prevalence of family firms in the United States: A preliminary report. *Entrepreneurship Theory and Practice*, 32(3), 559-573. <https://doi.org/10.1111/j.1540-6520.2008.00241.x>
- Chua, J. H., Chrisman, J. J., Steier, L. P., & Rau, S. B. (2012).** Sources of heterogeneity in family firms: An introduction. *Entrepreneurship Theory and Practice*, 36(6), 1103-1113. <https://doi.org/10.1111/j.1540-6520.2012.00540.x>
- Claessens, S., Djankov, S., & Lang, L. H. P. (2000).** The separation of ownership and control in East Asian corporations. *Journal of Financial Economics*, 58, 81-112. [https://doi.org/10.1016/S0304-405X\(00\)00067-2](https://doi.org/10.1016/S0304-405X(00)00067-2)
- Crespi-Cladera, R., Martín-Sánchez, M., & Infantes-Sánchez, P. M. (2016).** *La Empresa Familiar en Baleares* (2016). Càtedra Banca March de l'Empresa Familiar.
- Daspit, J. J., Chrisman, J. J., Ashton, T., & Evangelopoulos, N. (2021).** Family firm heterogeneity: A definition, common themes, scholarly progress, and directions forward. *Family Business Review*, 34(3), 296-322. <https://doi.org/10.1177/08944865211008350>
- Diéguez-Soto, J., López-Delgado, P., & Rojo-Ramírez, A. A. (2015).** Identifying and classifying family businesses. *Review of Managerial Science*, 9(3), 603-634. <https://doi.org/10.1007/s11846-014-0128-6>
- European Commission (2009).** Overview of family-business-relevant issues. Report-Expert group on family business between 2007 - 2009. Retrieved from <https://ec.europa.eu/docsroom/documents/10388/attachments/1/translations>
- European Commission (2015).** *Family business in Europe*. European Parliament resolution of 8 September 2015 (2014/2210(INI)). Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015IP0290>
- Evert, R. E., Martin, J. A., McLeod, M. S., & Payne, T. (2016).** Empirics if family business research: Progress, challenges, and the path ahead. *Family Business Review*, 29(1), 17-43. <https://doi.org/10.1177/0894486515593869>
- Faccio, M., & Lang, L. H. P. (2002).** The ultimate ownership of Western European corporations. *Journal of Financial Economics*, 65, 365-395. [https://doi.org/10.1016/S0304-405X\(02\)00146-0](https://doi.org/10.1016/S0304-405X(02)00146-0)
- Flören, R., Uhlaner, L., & Berent-Braun, M. (2010).** *Family business in the Netherlands. Characteristics and success factors: A report for the Ministry of Economic Affairs*. Center for Entrepreneurship, Nyenrode Business Universiteit, Breukelen, the Netherlands, 25.
- Franks, J., Mayer, C., Volpin, P., & Wagner, H.F. (2012).** The life cycle of family ownership: International evidence. *The Review of Financial Studies*, 25, 1676-1712. <https://doi.org/10.1093/rfs/hhr135>
- IFERA (2003).** Family businesses dominate: International Family Enterprise Research Academy (IFERA). *Family Business Review*, 16(4), 235-240. <https://doi.org/10.1177/08944865030160040201>
- Jaskiewicz, P., & Dyer, W. G. (2017).** Addressing the elephant in the room: Disentangling family heterogeneity to advance family business research. *Family Business Review*, 30(2), 111-118. <https://doi.org/10.1177/0894486517700469>
- La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999).** Corporate ownership around the World. *Journal of Finance*, 54, 471-517. <https://doi.org/10.1111/0022-1082.00115>
- López-Delgado, P., & Diéguez Soto, J. (2015).** Lone founders, types of private family businesses and firm performance. *Journal of Family Business Strategy*, 6(2), 73-85. <https://doi.org/10.1016/j.jfbs.2014.11.001>
- Martínez-Romero, M. J., & Rojo-Ramírez, A. A. (2017).** Socioemotional wealth's implications in the calculus of the minimum rate of return required by family businesses' owners. *Review of Managerial Science*, 11(1), 95-118. <https://doi.org/10.1007/s11846-015-0181-9>
- Neubaum, D. O., Kammerlander, N., & Brigham, K. H. (2019).** Capturing family firm heterogeneity: How taxonomies and typologies can help the field move forward. *Family Business Review*, 32(2), 106-130. <https://doi.org/10.1177/0894486519848512>
- Pindado, J., & Requejo, I. (2015).** Family business performance from a governance perspective: A review of empirical research. *International Journal of Management Reviews*, 17(3), 279-311. <https://doi.org/10.1111/ijmr.12040>
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001).** A multivariate tech-

- nique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27, 85-95.
- Rojo-Ramírez, A. A., Diéguez-Soto, J., & López-Delgado, P. (2011).** Importancia del concepto de empresa familiar en investigación: utilización de la base de datos SABI para su clasificación. *European Journal of Family Business*, 1(1), 53-67. <https://doi.org/10.24310/ejfbejfb.v1i1.5034>
- Sarkar, S., Hernández-Linares, R., Cobo, M.J. (2014).** Winnowing wheat from the chaff: The evolution of the family business concept. *74th Annual Meeting of the Academy of Management*. Philadelphia, Pennsylvania, USA. <https://journals.aom.org/doi/10.5465/ambpp.2014.14200abstract>
- Shanker, M. C., & Astrachan, J. H. (1996).** Myths and realities: Family businesses' contribution to the US economy – A framework for assessing family business statistics. *Family Business Review*, 9(2), 107-123. <https://doi.org/10.1111/j.1741-6248.1996.00107.x>
- Short, J. C., Sharma, P., Lumpkin, G. T., & Pearson, A. W. (2016).** Oh, the places we'll go! Reviewing past, present, and future possibilities in family business research. *Family Business Review*, 29(1), 11-16. <https://doi.org/10.1177/0894486515622294>
- Villalonga, B., & Amit, R. (2010).** Family control of firms and industries. *Financial Management*, 39(3), 863-904. <https://doi.org/10.1111/j.1755-053X.2010.01098.x>