Using Artificial Intelligence to Measure the Family Control of Companies

Finance Working Paper N° 950/2024
January 2024

Mario Daniele Amore
HEC Paris, CEPR, ICGS and ECGI

Valentino D'Angelo
Bocconi University

Isabelle le Breton-Miller
HEC Montréal

Danny Miller
HEC Montréal

Valerio Pelucco
Bocconi University

Marc van Essen
University of South Carolina

ECGI Working Paper Series in Finance

Using Artificial Intelligence to Measure the Family Control of Companies

Working Paper N° 950/2024
January 2024

Mario Daniele Amore
Valentino D'Angelo
Isabelle le Breton-Miller
Danny Miller
Valerio Pelucco
Marc van Essen

Abstract

Many studies have focused on family firms. Yet, grasping the nature of these organizations remains challenging because firms' familiness can take many forms, which are hard to trace with traditional data. We use AI to unravel the complex and intangible influence of families on firms in large datasets. Whereas it classifies family firms often consistently with equity criteria, AI is able to gauge families' legacy and values. As a result, using AI allows to detect more family firms in countries where families have a strong influence on firms even without large equity stakes. Importantly, AI distinguishes between family and lone-founder firms, and it assigns higher scores to firms that are eponymous, heir-led, and with multiple family directors. Finally, classifying family firms by using AI provides financially-relevant information to investors.

Mario Daniele Amore*
Professor
HEC Paris
Rue de Liberation 1, 79351
Jouy en Josas, France
e-mail: amore@hec.fr

Valentino D'Angelo
Researcher
Bocconi University
Via Roberto Sarfatti, 10
20136 Milano MI, Italy
e-mail: valentino.dangelo@phd.unibocconi.it

Isabelle le Breton-Miller
Professor and Chair in Succession and Family Enterprise
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec)
Canada H3T 2A7
phone: + 514 340-7315
e-mail: isabelle.lebreton@hec.ca

Danny Miller
Researcher
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec)
Canada H3T 2A7
phone: + 514 340-7315
e-mail: danny.miller@hec.ca

Valerio Pelucco
Researcher
Bocconi University
Via Roberto Sarfatti, 10
20136 Milano MI, Italy
e-mail: valerio.pelucco@phd.unibocconi.it

Marc van Essen
Professor
University of South Carolina
1014 Greene Street
Columbia SC 29208, USA
e-mail: Marc.vanessen@moore.sc.edu

*Corresponding Author

Using Artificial Intelligence to Measure the Family Control of Companies

Mario Daniele Amore
Valentino D'Angelo
Isabelle Le Breton-Miller
Danny Miller
Valerio Pelucco
Marc Van Essen

Abstract

Many studies have focused on family firms. Yet, grasping the nature of these organizations remains challenging because firms' familiness can take many forms, which are hard to trace with traditional data. We use AI to unravel the complex and intangible influence of families on firms in large datasets. Whereas it classifies family firms often consistently with equity criteria, AI is able to gauge families' legacy and values. As a result, using AI allows to detect more family firms in countries where families have a strong influence on firms even without large equity stakes. Importantly, AI distinguishes between family and lone-founder firms, and it assigns higher scores to firms that are eponymous, heir-led, and with multiple family directors. Finally, classifying family firms by using AI provides financially-relevant information to investors.

Keywords: Family Firms; AI; Management; Legacy; Values

Mario Daniele Amore; HEC Paris, CEPR, ECGI: amore@hec.fr
Valentino D'Angelo; Bocconi University: valentino.dangelo@phd.unibocconi.it
Isabelle Le Breton-Miller; HEC Montreal: isabelle.lebreton@hec.ca
Danny Miller; HEC Montreal: danny.miller@hec.ca
Valerio Pelucco; Bocconi University: valerio.pelucco@phd.unibocconi.it
Marc Van Essen; University of South Carolina: marc.vanessen@moore.sc.edu

1. Introduction

The question of who owns business firms has been paramount in management and finance research. Empirical works such as La Porta et al. (1999), Faccio and Lang (2002), and, more recently, Aminadav and Papaioannou (2020) suggest that families are among the most diffuse types of corporate owners. To classify family firms, the common approach in this literature has been to check whether founders or family members own certain thresholds of equity or voting rights in a given firm.

While this approach offers an intuitive and practical operationalization, it cannot capture important nuances across family firms. A firm in which family members own a 5% equity stake and do not exercise control of the business is clearly different from one in which the family owns the majority of shares and is heavily involved in management. Moreover, a firm led by a founder with no intention of passing control to future family generations is markedly different from a dynastic business led by a founder's siblings or nephews (Duran et al. 2016; Miller et al. 2011). To account for these differences, scholars have begun to extend the criteria used to operationalize family firms by employing many disparate definitions. Although ownership remains the most common criterion, Bennedsen et al. (2021) found 135 different definitions based on ownership, management, and governance criteria (or a combination).[1] While broadening the definition of family firms is beneficial, there remains some confusion regarding the critical attributes of family firms, and it is not clear whether one definition should be preferred to others.

This is an important issue as there remains much disagreement on the objectives, strategies, and performance of family firms (e.g., Miller et al. 2013; Belenzon et al. 2016), and

[1] They reviewed 112 articles which adopted 135 definitions of family firms. In about half of the cases, the definition is based on ownership, whereas the other half is based on combinations of ownership and management, governance or realized successions.

disparate family firm definitions may be one source of these debates.[2] Moreover, such conflicting characterizations can be partly due to family firms being a highly heterogeneous type of organization (Daspit et al. 2021) with different levels and types of family involvement. Thus, the nature and degree of families' involvement in the business and their social values and legacy objectives may have an important influence on how family firms behave and perform (Berrone et al. 2022). Some say that family firms are driven by social and emotional priorities like preserving generational control (Gomez-Mejia et al. 2007, 2011). Others, however, maintain that such priorities vary greatly across such firms, such that some pursue long-term stakeholder-friendly objectives (Miller and Le Breton-Miller, 2005; 2014) while others favor immediate economic and career benefits for the family (Bloom and Van Reenen, 2007; Bertrand and Schoar, 2006). Also, families with a purely financial stake may prioritize ongoing economic benefits and be strategically conservative, whereas those managing the firm and envisioning intergenerational family involvement may prioritize longer-term objectives and the types of innovative and expansive strategies to help achieve those objectives.

A broadly applicable, graduated, and fine-grained measure of *familiness* in family firms along with specific subscales on management involvement, values, and legacy priorities may inform and reconcile such disagreements in the family business literature, ultimately resulting in more cumulative research contributions. Given that family involvement in business varies in intensity, the ideal measure of family firms should be continuous. Moreover, it should be context-specific and be computed by leveraging the entire amount of information available on a given firm (not just data on equity or CEO but also historical and qualitative information on the evolution of the business, the intangible influence of families on decision-making, the family's values and legacy concerns, etc).

[2] Specifically, debates persist as to whether vis-à-vis other types of companies family firms are more or less innovative (Duran et al. 2016; Block et al. 2023), socially responsible (Mariani et al. 2023), fast growing (Miroshnychenko et al. 2022), and financially successful (Van Essen et al. 2015; Wagner et al. 2015).

In this paper, we take advantage of recent developments in AI to measure family control in a large sample of listed firms around the world. Specifically, we employ the AI-based service ChatGPT to use all available information on listed firms to quantify the extent to which they can be classified as family firms. Moreover, we use ChatGPT to develop specific sub-scores capturing the influence of family ownership, management, values, and legacy. We submit a set of queries multiple times for each firm, and then we take the average (or median) to compute the final metrics used in the empirical analysis. Our investigation provides several results that broaden current understandings of the essence of family firms, and how well these organizations perform. Moreover, it guides how scholars can use AI tools to measure corporate control around the world.

We start the analysis by comparing the country-level fraction of companies classified as family firms by AI against the fraction obtained by using equity holdings. We find that AI detects more family firms than equity criteria. On average, the most conservative share of family firms in our sample is 44% according to ChatGPT, and between 23% and 27% according to equity criteria. The AI classification is consistent with equity-based criteria in Anglo-Saxon countries like Australia, the US, and the UK as well as some Continental European countries like France. This suggests that being a family firm in such countries revolves around owning relevant equity stakes. By contrast, the difference is very large in Japan, where AI detects more than 50% of family firms while equity criteria yield a fraction as low as 4%. Studying Japan, Bennedsen et al. (2021) argue that legacy and cultural factors allow families to influence firms even without large equity holdings; several well-known firms such as Toyota, Casio, and Suzuki are commonly considered as family firms even if the founding families own little equity.[3] As a result, Bennedsen et al. (2021) argue that "family firms in Japan are more prevalent than the very low family ownership documented in extant studies would suggest".

[3] See, e.g., https://www.ft.com/content/b63b75c2-21b9-11ea-b8a1-584213ee7b2b

Our evidence, fully consistent with this insight, suggests that using AI allows scholars to capture the cultural elements of family control.[4]

To probe into this aspect, we explore how the AI family score derived from ChatGPT correlates with firms' leadership and governance attributes. Descriptively, the score is significantly higher for firms led by descendants or with multiple family members sitting on the board of directors. By contrast, the distribution of the family score is extremely similar for firms led by lone founders and non-family firms. These results are confirmed by our regression analysis, which keeps constant an array of other factors such as firm age, industry, country, etc. The AI family score is positively associated with a binary equity-based classification of family firms in the full sample, and the association remains statistically significant across different subsamples of firm size and geographic location. Yet, there are strong differences depending on family firms' leadership and governance: the AI family score is much higher for eponymous firms, firms with multiple family members on the board of directors, and firms led by family descendants. By contrast, it is statistically indistinguishable between lone-founder firms and non-family firms.

Next, we analyze the association between the binary equity-based classification of family firms and the AI sub-scores of family ownership, management, legacy, and values. All of these components are positively associated with being a family firm, though the effect is economically greater for family management, legacy, and values. Again, this evidence suggests that AI captures the nuanced influence of families on their businesses via managerial involvement, cultural influence, and legacy concerns above and beyond the mere effect of holding equity shares.

[4] With regards to the examples of Toyota, Casio and Suzuki discussed in Bennedsen et al. (2021), we found that ChatGPT would classify them as family firms (having a score corresponding to "agree") whereas equity criteria would classify them as non-family firms.

Before concluding, we ask whether using AI to classify family firms provides valuable information to investors. To this end, we collect data on stock returns for the subsample of US-listed firms, and compare the returns obtained by constructing different portfolios of family firms. The analysis provides some evidence that the portfolio of family firms constructed using an equity definition performs as well as the portfolio of all firms constructed using AI.

Our paper resonates with a nascent stream of works that employ AI to develop a variety of firm-specific measures related to investment policies (Jha et al. 2023), earnings expectations (Li et al. 2023a), and organizational culture (Li et al. 2023b). Leveraging a similar approach, we provide a number of contributions to the literature on family firms. First, we offer new insights into the long-running debate about how to operationalize the family control of firms. Dating back to Anderson and Reeb (2003), Miller et al. (2007), and Villalonga and Amit (2006), scholars have wrestled over questions like whether Microsoft (with its strong founder influence and founder equity) or Toyota (having low family equity but strong family legacy) are family firms. A consensual view among scholars is that family firms are a heterogeneous category because families differ widely in their economic and cultural attachment to the family business. Hence, as noted above, the ideal measure of family control should encompass family firms' values as well as intergenerational orientation, i.e. the propensity of current firm leaders to pass on control to next family generations. Yet, capturing these dimensions empirically has proven cumbersome given the intangible nature of family values and because intergenerational concerns can only be measured after they have manifested (see Berrone et al. 2022 for a recent approach). By processing large volumes of textual information available on the internet, the AI can differentiate family and non-family firms along a continuum of weak to strong family legacy, and weak to strong family values. These differentiations are helpful as they: (1) help to uncover a higher share of family firms in contexts where family influence does not require large equity holdings; (2) correlate with governance and leadership attributes (and hence capture meaningful concrete aspects of how family firms are managed). Operationally, AI-based measures of family firms can be derived for very large samples at a low cost, and they help to pinpoint the cultural nature of family control.

2. Data

2.1. Sample

To examine the relationship between traditional family firm metrics and the extent of family control measured by AI we used two main sources of data: NRG Metrics and Chat GPT. The former provides information on the main variables of family firms used in the prior literature, e.g. the family's ownership stake, whether or not the CEO is a family member, the number of family members sitting on the board, etc. The NRG Metrics's Family Firms dataset is created by a team of expert analysts who manually enter, review, and cross-check data with senior analysts, who perform frequent random audits. It is based on publicly available documents such as annual reports, corporate governance reports, firm presentations, SEC filings, and press releases. NRG Metrics has been used in both management and finance literature (e.g., Delis et al., 2019; Miroshnychenko et al., 2020; Marano et al., 2022; Pinelli et al., 2023; Gómez-Mejía et al., 2023). The dataset covers publicly traded (active and non-active) firms worldwide beginning in fiscal year 2007. Since we asked Chat GPT to provide a measure of the extent to which a particular firm could have been defined as a family business in the year 2011, we retained all information reported by NRG metrics in 2011 because the number of firms in the dataset increased in 2011 and remained relatively stable.

We merge the NRG dataset with Compustat Global (WRDS) by using ISIN codes to obtain the accounting variables necessary for the empirical analysis. When the ISIN in NRG metrics did not correspond to the ISIN in WRDS, a fuzzy match by firm name similarity was performed. When accounting variables in 2011 were missing in WRDS, we imputed the latest available information (i.e. 2010 or earlier).

Net of observations with missing values, our dataset comprises 3,864 unique firms across 43 countries. Table A1 provides the number of firms by country. The NRG dataset only covers industrial companies, excluding banks, real estate, insurance firms, and similar entities. Table A2 provides the number of firms by industry.

2.2. Data collection process

ChatGPT is an AI service developed by the US organization OpenAI and made available to the public in November 2022. It builds on a large language model "developed using three primary sources of information: (1) information that is publicly available on the internet, (2) information that we license from third parties, and (3) information that our users or our human trainers provide".[5]

In late June 2023, we accessed the OpenAI website to utilize the Playground API service, which proved ideal for handling substantial queries. This service offered the flexibility to incorporate web scraping and Python code for processing large datasets and repeated coding procedures. Our primary aim was to structure the data coding process to yield an output as closely aligned as possible with the content generated by ChatGPT. For this purpose, we employed the GPT-3.5 Turbo model to generate responses to our prompts. We fine-tuned the model's behavior by configuring certain parameters such as temperature (1), frequency penalty (0), presence penalty (0), and max tokens (4097).[6]

---

[5] For details, please refer to: https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed

[6] Temperature is a hyperparameter used in some natural language processing models, including ChatGPT, to control the level of randomness or "creativity" in the generated text. Frequency and presence penalty are parameters of the AI algorithm that discourage the generation of repetitive text. Max tokens refer to the memory size of input and output of each prompt/query.

Our approach commenced with the establishment of a "System" message, as outlined in Step 1 of Figure 1. This initial message set the context, framing it as business experts with comprehensive knowledge of diverse enterprises, particularly in historical data (i.e. not just information at the moment of data collection). This step was pivotal as the subsequent responses from the AI model are influenced by the content and tone established in the System message. Our objective was to craft a prompt that would allow the AI model to fully utilize its capabilities in categorizing firms based on their family business characteristics by drawing on all available information.

In the next stage of the prompt, we instructed ChatGPT to assess each firm in our sample using a one-to-seven Likert scale, gauging the extent to which it could be classified as a family firm. A score of 1 denoted a firm that was not a family business, while a score of 7 indicated a firm that was entirely a family business. Once we obtained these preliminary scores, we proceeded to refine the prompt by: (1) reiterating the system setting in Step 3, and (2) requesting the model to provide sub-scores regarding four distinct dimensions (in addition to the overall score, as described in Steps 4 and 5). These dimensions encompassed: i) Ownership structure, ii) Management structure, iii) Family values, and iv) Family legacy. In the next paragraph, we provide the theoretical rationale which led us to focus on these dimensions.

To round out the process, the prompt solicited historical information that could influence the evaluation of companies over time (Step 6). Subsequently, the scores were updated, taking into account the historical data recalled, with the aim of aligning them with presumed 2011 levels (Step 7). This step facilitated contemporaneity between the traditional ownership data and the AI-generated output. Figure 1 provides a breakdown of the prompt's details for each category. It is noteworthy that the selection of the four attributes was an iterative process, involving interactions with ChatGPT. Initially, the model generated a preliminary draft of the categories, which were then harmonized with constructs found in the literature. Overall, there was a substantial alignment between the categories generated by the model and the pertinent topics covered in the literature.

INSERT FIGURE 1 HERE

2.3. Unpacking family firms with AI

As discussed, ownership is a widely used criterion for identifying family firms. When asked about the reasons for that, ChatGPT generated the following response: "This attribute looks at the percentage of the company's ownership that is held by a family or families. It's a fundamental factor because family-owned firms typically have a significant family ownership stake. Ownership structure helps determine whether the family has a controlling interest in the company." These arguments notwithstanding, the theoretical relevance of this construct has been challenged by some researchers. This challenge arises from two main considerations: (1), ownership may not fully capture the significance of the firm for the family, and (2) it may not fully represent the importance of the family for the firm's creation and operation.

Conceptually, several works in the literature (e.g., Gomez-Mejia et al. 2011) have stressed that family firms differ from non-family firms due to the ties between family and business, which has an emotional basis arising from family history and memories, as well as a cultural basis in the family values permeating the organization. Also, using ownership as a classifying criterion for family firms fails to consider the family relationships within the firm. Using data on kinship relationships (blood, marriage, or adoption) among executives and directors of US-listed firms, Parise (2023) argues that "about half of the firms that are commonly classified as family firms do not report any family presence, whereas more than half of the firms that disclose widespread family links are not classified as family firms". Bennedsen et al. (2021) argue that several companies in Japan commonly considered family firms display a relevant family involvement while the family owns little to zero ownership (and hence will not be classified as family firms using an equity criterion). Even firms that do have relevant family ownership might not be classified as family firms whenever the family's equity holding is barely below a given threshold (e.g. 4.9% rather than 5%). These notions suggest that family ownership alone does not adequately characterize the nature of family firms. Firms with similar levels of family ownership can exhibit significant heterogeneity (e.g., Berrone et al. 2010; Gomez-Mejia et al. 2007), and firms with very little family involvement may be classified as family firms if one solely uses an ownership criterion (Parise 2023). Hence, other dimensions must be considered to gain a comprehensive understanding of family firms.

Family involvement in the top management of a company is important as it plays a pivotal role in shaping corporate policies and facilitating the alignment of business objectives with family goals. Villalonga and Amit (2006), among others, show that performance disparities between family and non-family firms vary as a function of different kinds of family leadership. In particular, their research shows that the highest level of financial performance is attained by founder-led firms rather than those led by family heirs. In this vein, Miller et al. (2007) find that lone founders, without the involvement of other family members, achieve the highest level of performance relative to all other types of firms (non-family firms, family firms led by family heirs, etc). When asked about the inclusion of management as a criterion to define family firms, ChatGPT replied: "Examining the management structure involves assessing whether family members are actively involved in the company's leadership and decision-making. This can include family members serving as executives or on the board of directors. Family involvement in management is a common characteristic of family firms." This criterion helps distinguish family firms based on the degree to which family members are engaged in key leadership roles within the company, providing valuable insights into their behavior and impact on performance.

The practitioner literature has stressed the importance of legacy in family firms. For instance, according to the *Harvard Business Review* “Legacy is an intangible, invisible force that affects the decision making of the next generation in the business, and in life in general. Legacy has also been called the “connective tissue” that links generations in a family business”.[7] The concept of family legacy, encapsulated as the "intention for transgenerational control", constitutes a central preoccupation for many family entrepreneurs (Chua et al., 2003 and 2004; Kiong, 2005; Zellweger et al., 2012), and hence represents a fundamental aspect of family firms. Bennedsen et al. (2021) emphasize the significance of family legacy, showcasing its broader implications beyond ownership. Their study of Japanese firms demonstrates that founding families maintain control of their companies through an array of strategic approaches that perpetuate the family legacy, independent of voting or cash-flow rights. Despite this importance, legacy is rarely used in empirical research on family firms due to the challenges of obtaining reliable data on a large scale.[8] By leveraging AI, we can derive a measure of family legacy to be employed in our empirical analysis. ChatGPT provides the following rationale for including family legacy as a criterion: "Family legacy is often considered important when defining family firms because it reflects the historical and cultural aspects that distinguish these businesses from other types of enterprises. Family legacy signifies the company's commitment to continuity and long-term sustainability. Many family firms are deeply invested in preserving the business for future generations. This long-term perspective can influence decision-making and strategies." These notions suggest that family legacy serves as a critical dimension for discerning family owners’ dedication to continuity, longevity, and the preservation of their firms’ historical and cultural heritage.

[7] See https://hbr.org/2022/09/wrestling-with-legacy-in-a-family-business
[8] A rare exception is the survey approach in Zellweger et al. (2012).

Values are pivotal when characterizing family firms, as they play a fundamental role in shaping the essence of these businesses. The influence of family values is a cornerstone criterion in defining family firms and has been explored from various theoretical perspectives, including via concepts like "familiness" (Zellweger et al., 2010) and "embeddedness" (Le Breton-Miller and Miller, 2009; Le Breton-Miller et al., 2011). ChatGPT offers the following rationale for including family values among the criteria used to identify family firms: "While not all family firms explicitly emphasize family values, some do. Assessing a company's values can help determine if there is an emphasis on family-oriented values such as continuity, tradition, and long-term sustainability." This underscores that firm values can offer valuable insights into whether an organization emphasizes family-centric values like perpetuity, heritage, and the enduring sustainability of operations. Thus, the values dimension can be critical in understanding and categorizing family firms.

3. Empirical results

3.1. Descriptive evidence

As shown in Panel A of Table 1, we are left with 3,864 unique firms after dropping observations with missing information. Of those, 21.1% are family businesses based on ownership of at least 5% of shares outstanding. Additionally, 16.4% of these firms are eponymous -- the name of the company is that of the family. Moreover,13.1% have a lone founder, i.e. a founder who is also an officer or director, or a large owner (5% or more of the firm's equity) but does not have other family members involved in the business. 12.2% of the firms have a descendent of the founder as CEO, chairman, or board member.

INSERT TABLE 1 HERE

We proceed to present the AI-based score of family control. For each company in our sample, we repeated the prompts discussed in the previous section 30 times. We inspected the distribution of the main family score obtained by averaging a different number of attempts (from 2 to 30). As expected, the distribution is less smooth when using 2 attempts, becoming far more so after 5 or more attempts. For our empirical analysis, we use the average based on 30 attempts.

Panel B of Table 1 provides the summary statistics on the main family score and the four sub-scores (i.e. ownership, management, legacy, and values), whereas Panel C provides the correlations among them, which are quite high. Consequently, we do not include multiple sub-scores in our regression analyses.

Table 2 compares country-level averages, focusing on the proportion of companies classified as family enterprises by ChatGPT. The operational definition of this classification relies upon a score above the threshold indicating consensus, specifically, a "somewhat agree" rating regarding the status as a family firm. A parallel examination evaluates the fraction of firms identified through conventional equity-based criteria, where the family owns 5% or more of the total outstanding shares (or any share at all). For comparison purposes, we also reproduce the share of family firms found in Aminadav and Papaioannou (2020).

INSERT TABLE 2 HERE

Table 2 reveals an appreciable discrepancy between the AI-based classification and existing paradigms. Significantly, ChatGPT detects a much higher prevalence of family businesses vis-à-vis the equity criterion, showing substantial variation across countries. In some countries (e.g., Australia, the US, the UK, and France), the proportion of family businesses identified by ChatGPT closely approximates that from conventional ownership metrics. However, in other contexts, the contrast is stark. In Japan, the disparity between the two methods reaches 46%. This aligns with Bennedsen et al. (2021), who indicate that legacy and cultural factors enable families to exert significant influence over Japanese firms, even when their equity holdings are modest. This indicates that ChatGPT can reveal culture-driven forces underlying family control in diverse contexts. For example, although founding family ownership is modest at Toyota, family members still shape strategic decisions and guide corporate governance. This underscores ChatGPT's proficiency in identifying family enterprises, while also capturing the complex interplay of cultural, legacy, and ownership dynamics that characterize family control across diverse national and corporate contexts. Denmark and Norway are other countries displaying a large difference between the proportion of family businesses identified by ChatGPT and that derived from conventional ownership-based metrics. Here, ChatGPT identifies between 26% and 57% more family firms, likely owing to the difficulty faced by traditional approaches in identifying the foundations owning companies there (Thomsen et al. 2018).

INSERT FIGURE 2 HERE

Figure 2 shows the distribution of the AI family score for firms classified as family or non-family based on a 5% equity criterion.[9] As expected, the distribution for family firms is shifted to the right. This evidence demonstrates ChatGPT's ability to recognize and quantify the influence of familial ownership within a firm.

However, ChatGPT's utility extends beyond the consideration of ownership percentages. The left part of Figure 3 reveals that it assigns significant weight to the presence

[9] When exploring higher ownership thresholds such as 10% or 25%, unreported data suggests an even stronger correlation, reinforcing the notion that an increase in family ownership corresponds to higher ChatGPT scores.

of founder heirs in top management positions. Businesses where heirs occupy key management roles receive higher scores. This highlights ChatGPT's capacity to discern the presence of key family executives within an organization. Another compelling aspect of AI's family dimension evaluation is its ability to distinguish between actual family firms and enterprises where the founder has a substantial equity share but lacks any family involvement – a lone founder firm (Miller et al., 2007). Google, Facebook, Amazon, and Berkshire Hathaway, are often misclassified as family businesses by studies using an ownership criterion but tend to be identified by ChatGPT as non-family entities (see the right panel of Figure 3). This level of discrimination demonstrates ChatGPT's ability to distinguish between lone-founder firms and family firms. ChatGPT also takes into account the composition of the board of directors. Figure 4 shows that an increase in the number of family members on the board corresponds to higher scores from ChatGPT, demonstrating the importance of family involvement in governance in the familial character of a business.

INSERT FIGURE 3 HERE

INSERT FIGURE 4 HERE

3.2. Relationship between AI score and equity classification of family firms

Following the compelling evidence presented in the preceding figures, we now employ regression analyses to account for various firm characteristics. In Table 3, the dependent variable is an indicator that takes on a value of one if a family possesses at least 5% of the outstanding shares in a company, and zero otherwise. Our focal independent variable is the aggregate average score derived from ChatGPT's assessment. We estimate linear probability models in which we progressively introduce control variables to capture distinct facets of firms. In particular, we control for: (1) the size of the firm using the natural logarithm of one plus the sales of the firm in $millions (Column 2), (2) the age of the firm using the natural logarithm of the age of the firm in years (Column 3), (3) the gearing ratio using the ratio between the common equity outstanding and the assets of the company (Column 4). These controls are helpful to ameliorate the concern that the AI score reflects differences in the amount of textual information available on some firms (e.g., there is likely to be more media material on larger or older firms) which in turn may be correlated with the family firm status. In Column 5, we further control for the industry where the firm operates including two-digit SIC dummies. Finally, in Column 6 we control for geographic heterogeneity by including country dummies. Again, these controls are useful to control for the differences in the availability of information across countries or industries, which in turn may correlate with the family control of firms. Depending on the specification, a one-point increase in the AI family score is associated with a 5.8 to 6.7 percentage points increase in the probability of being a family enterprise based on the 5% ownership criterion.

INSERT TABLE 3 HERE

In Table 4A we replicate the analyses reported in Column 6 of Table 3 separately for each continent. Even if the relationship between the ownership criterion and the family AI score is stronger in the US, the documented relationship remains significant worldwide. This underscores the validity and universality of AI in assessing family enterprises worldwide. The tool's versatility in recognizing familial attributes in companies from diverse cultural, regulatory, and economic contexts highlights its broad applicability and utility in the study of family firms on a global scale.

In Table 4B we perform a similar exercise by repeating the same analysis for firms belonging to different quartiles in terms of their size. The relationship between the AI family score and the family firm status is stronger for average-size firms.

INSERT TABLE 4 HERE

3.3. Relationship between AI score and family firms' attributes

In Table 5, we assess whether the AI family score incorporates elements long of interest to family firm scholars. In Column 1 of Table 5, the dependent variable is an indicator variable taking the value of one if the firm is eponymously named, and zero otherwise. Eponymy has been shown to increase the identification of the family with the firm and to magnify the image and reputational concerns (Belenzon et al. 2017; Deephouse and Jaskiewicz, 2013). We expected ChatGPT to assign a higher family score to eponymous firms.

Column 2 indicates the presence of a lone founder actively involved in leadership. The dependent variable is a dummy with a value of 1 if an individual is a founder with no other family members involved, and also an officer or director, or a large owner (5% or more of the firm's equity); zero elsewhere. Lone founders are those whose family members do not have a role in ownership and management, and who have not shown interest in passing control to other family members. We expected ChatGPT to assign a lower family score to lone founder firms.

Column 3 shows familial representation on the board. The dependent variable is a dummy with a value of one if at least one member of the family sits on the board; and 0 elsewhere. Sitting on the board is a way for family members to be involved and have a say in business affairs. Moreover, the board is a talent pool to draw the next family CEO. We expected ChatGPT to assign a higher family score to firms with family members sitting io the board.

Column 4 reflects the involvement of descendants of the founder in active management roles, such as CEO, chairman, or board member. Again an indicator variable is used to indicate where no such descendant engagement exists. The rationale is that having a family heir in these key corporate positions suggests an intention to keep control within the family. Hence, we expected ChatGPT to assign a higher family score to such firms.

Our analyses estimate linear probability models and incorporate the same control variables from Column 4 of Table 3. Our findings reveal that a unit increase in the AI family score is associated with a 2.4% increase in the likelihood of the firm being named after its founder. The same one-point increase correlates with an 8.1% increase in the probability of family board representation and a 6.3% boost in the likelihood of descendants actively participating in company management. Conversely, where the founders are major owners without the involvement of any family members (e.g., Amazon, Google, Microsoft), the firm is not counted as a family business. As shown in Column 2, the AI family score is not correlated with the lone founder variable. These results underline the range and versatility of AI to consider a diverse array of factors when generating its scores. Importantly, these factors align with attributes traditionally associated with family-owned businesses in the literature, thus suggesting the robustness and pertinence of ChatGPT assessments.

INSERT TABLE 5 HERE

3.4. AI subscores of family ownership, management, values and legacy

Table 6 replicates the analysis conducted in Column 4 of Table 3, substituting the overall family score variable with the subscores presented in Panel B of Table 1. It relates these scores to the family firm variable defined by 5% family ownership. Table 6 reveals that all the sub-indices correlated positively with the family firm variable. Notably, this holds in Column 3, relating family ownership and AI-derived family values, accentuating the significance of the individual components of the family score.

INSERT TABLE 6 HERE

3.5. Financial relevance of the AI score

Before concluding, we offer some descriptive insights on whether classifying family firms using AI provides a meaningful approach to financial investors. To this end, we extract daily stock return data on US-listed firms from CRSP and match them with our sample. We then construct the stock returns that an investor would have obtained by buying stocks of family firms on the first trading day of 2011 (recall that the AI classification is done as of 2011) and holding them for two or three full years (i.e. until the last trading day of 2012 or 2013). We subtract the industry median return over the same periods (calculated at the 2-digit SIC level) from the firm-level stock return to remove industry differences. Figure 5 provides the industry-adjusted stock return and relative 95% confidence interval. As shown, investing in a portfolio of family firms selected using a 5% equity criterion would yield an industry-adjusted return indistinguishable from zero. By contrast, investing in a portfolio of family firms chosen using AI generates a positive and statistically relevant return. The difference between the two average returns is close to significant in the upper figure ($p=0.108$) and statistically significant for the lower figure ($p=0.02$).

INSERT FIGURE 5 HERE

----

4. Directions for future research

Using AI to operationalize the family control of firms provides several useful and valuable opportunities to advance and address ongoing debates in the literature. These include examining their conduct and strategic outcome effects, performance implications, use in predictive typologies, and contextual and demographic explorations.

4.1. Corporate outcomes

The AI-based family scores and subscores can be used in studies of firm conduct, strategy, and performance to reconcile current disagreements in the literature. Tensions in that literature are revealed by reviews on family firms' corporate social responsibility (Mariani et al., 2023; Marques et al., 2014), innovation input and output (Duran et al., 2016; Block et al., 2023), firm growth (Miroshnychenko et al. 2022), financial returns (Wagner et al., 2015), and capital structure (Hansen and Block, 2021). The measures we introduce in this study can help inform these debates. For example, family firms with high value and legacy scores may demonstrate profiles consistent with a long-term stakeholder orientation – for instance, higher CSR, long-term capital expenditures, and secular growth. By contrast, firms defined according to family financial holdings may score lower along these variables, acting more like non-family firms and being more short-term oriented and economically focused. Similar explorations can be conducted to examine internal operations such as human resource policies, modes of organization, efficiency, and productivity associated with our measures.

4.2. Exploration of extreme outcomes

It has been argued that family firms are likely more extreme in their behavior than other firms (Miller and Le Breton-Miller, 2022). For example, because of the high levels of family discretion and the importance of, and contrasts in, their values and legacy focus, they are said to be found among the most and least socially responsible businesses, the most and least innovative, and the most and least corrupt. A high overall family AI score, or a high score on some of the subscores, may reflect a high degree of familiness making firms more subject to family influence, ethics, and preferences, and thus more likely to demonstrate extremes along the above behaviors and outcomes.

4.3 Performance studies

A perennial debate surrounds the issue of family firm performance and its superiority or inferiority vis-à-vis other types of companies along various performance measures (see Wagner et al. 2015 for a review). Again, having a measure reflecting the overall level of familiness of a firm and that of our subcomponents allows for a more discriminating assessment of family involvement and priorities, and their relationship to financial returns, growth, market share, efficiency, etc. Exploring the performance consequences of different levels and types of familiness also makes it possible to examine curvilinear effects (e.g. is too much and or too little legacy focus bad?), as well as the different industry, national, and technological contexts that may moderate these relationships.

4.4. Configuration, combination, and family firm types

Another potential path to explore is how the different combinations or "configurations" among the ownership, management, values, and legacy AI subscores affect firm behavior and performance. For example, bifurcation of the four scales for a family firm will produce multiple family firm profiles: for instance those high in ownership and management only, those high in values and legacy but lower in ownership or management, etc. These different types of family firms may exhibit contrasting strategic behavior and performance, providing insight into the sources and consequences of family firm heterogeneity (Daspit et al., 2021). Qualitative comparative analyses (QCA) using boolean or fuzzy set operationalizations of combinations of the AI subscores could be used to discover multiple (equifinal) paths to particular strategic or performance outcomes.

4.5. Contextual comparisons

Several studies in the field of family business employ samples of firms from different countries. Yet, most of the focus has been on the role of formal institutions, such as the legal system (La Porta et al. 1999) or inheritance law (Carney et al. 2014), while we know relatively little about informal, cultural institutions. International comparisons can be made to examine in detail how the AI-based measures of ownership, management, values, and legacy scores vary among family firms in different countries, revealing the culturally-linked nature of family firms. For example, family firms in most Western economies may be characterized mainly through large ownership stakes. In contrast, those from the East and in developing economies may demonstrate higher legacy and value scores. Those differences can have a significant effect on the behavior of those firms. As the scores we introduced are scaled from 1 to 7, they may also reveal how the overall "familiness" of firms differs across countries in degree and ownership, management, values, and legacy. The same can be done to examine the degree and nature of family firm presence in different industries within and across countries.

4.6. Underlying demographics

Finally, an area to explore is the demographics behind the different types of family firms: how old and large are these firms, what are the demographic characteristics of their owners and managers; and how many family versus non-family members are involved in different roles in the business and what are those roles.[10]

[10] There is a literature on both corporate age (e.g., Loderer et al. 2017) and executives' individual characteristics (e.g., Belenzon et al. 2019) but this is mostly in the realm of non-family firms or does not explicitly distinguish between family and non-family firms.

5. Conclusion

Despite a wealth of available theories and empirical evidence, family business scholars continue to wrestle with the definition of family firms. There is a consensus that family firms are very different from each other, and this heterogeneity is hard to operationalize with standard data. In this paper, we use recent developments in AI to develop a new measure of family control which can be quickly derived at a low cost. While the measure is consistent with standard definitions based on equity holdings, there are relevant differences across countries which serve to illustrate its specific properties. For instance, it uncovers a much higher share of family firms in countries like Japan, where families often influence firms without holding large equity stakes (Bennedsen et al. 2021). We also find that the AI-based family score captures cultural and legacy dimensions of family control. For example, the score is as low for lone founder firms as it is for non-family firms, whereas it is especially high for eponymous firms, firms managed by family heirs, and firms with family members on the board of directors.

Future scholars can employ this measure to study family firms in contexts where equity information is hard to access or to resolve conflicting findings in the literature. For instance, an important application would be to discover whether family firms perform better or worse than non-family firms. Evidence to date is mixed, and points to substantial heterogeneity across family firms depending on family involvement and priorities (Miller et al. 2007; Villalonga and Amit 2006), family generation (Bennedsen et al. 2007; Amore et al. 2021), and the institutional context (Banalieva et al. 2015; Miller et al. 2017). An AI-based measure of family control that explicitly accounts for family culture and legacy may shed light on which specific types of family firms perform better than non-family firms. Related applications concern R&D and investment policies, which too have generated mixed findings. Finally, scholars can use AI tools to parse the role of family influence on future governance and leadership decisions.

References

Aminadav, G., & Papaioannou, E. (2020). Corporate control around the world. *Journal of Finance*, *75*(3), 1191-1246.

Amore, M. D., Bennedsen, M., Le Breton-Miller, I., & Miller, D. (2021). Back to the future: The effect of returning family successions on firm performance. *Strategic Management Journal*, *42*(8), 1432-1458.

Anderson, R. C., & Reeb, D. M. (2003). Founding-family ownership and firm performance: evidence from the S&P 500. *Journal of Finance*, 58(3), 1301-1328.

Anderson, R. C., Mansi, S. A., & Reeb, D. M. (2003). Founding family ownership and the agency cost of debt. *Journal of Financial Economics*, 68(2), 263-285.

Anderson, R. C., & Reeb, D. M. (2004). Board composition: Balancing family influence in S&P 500 firms. *Administrative Science Quarterly*, 49(2), 209-237.

Banalieva, E. R., Eddleston, K. A., & Zellweger, T. M. (2015). When do family firms have an advantage in transitioning economies? Toward a dynamic institution-based view. *Strategic Management Journal*, *36*(9), 1358-1377.

Belenzon, S., Patacconi, A., & Zarutskie, R. (2016). Married to the firm? A large-scale investigation of the social context of ownership. *Strategic Management Journal*, *37*(13), 2611-2638.

Belenzon, S., Chatterji, A. K., & Daley, B. (2017). Eponymous entrepreneurs. *American Economic Review*, 107(6), 1638-1655.

Belenzon, S., Shamshur, A., & Zarutskie, R. (2019). CEO's age and the performance of closely held firms. *Strategic Management Journal*, *40*(6), 917-944.

Bennedsen, M., Nielsen, K. M., Pérez-González, F., & Wolfenzon, D. (2007). Inside the family firm: The role of families in succession decisions and performance. *Quarterly Journal of Economics*, *122*(2), 647-691.

Bennedsen, M., Mehrotra, V., Shim, J., & Wiwattanakantang, Y. (2021). Dynastic control without ownership: Evidence from post-war Japan. *Journal of Financial Economics*, *142*(2), 831-843.

Berrone, P., Cruz, C., Gomez-Mejia, L. R., & Larraza-Kintana, M. (2010). Socioemotional wealth and corporate responses to institutional pressures: Do family-controlled firms pollute less? *Administrative Science Quarterly*, 55(1), 82-113.

Berrone, P., Duran, P., Gomez-Mejia, L., Heugens, P., Kostova, T., & Van Essen, M. 2022. Impact of informal institutions on the prevalence, strategy, and performance of family firms: A meta-analysis. *Journal of International Business Studies*, 53, 1153-1177.

Bertrand, M., & Schoar, A. (2006). The role of family in family firms. *Journal of Economic Perspectives*, *20*(2), 73-96.

Block, J., Hansen, C., & Steinmetz, H. (2023). Are family firms doing more innovation output with less innovation input? A replication and extension. *Entrepreneurship Theory and Practice*, *47*(4), 1496-1520.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, *122*(4), 1351-1408.

Carney, M., Gedajlovic, E., & Strike, V. M. (2014). Dead money: Inheritance law and the longevity of family firms. *Entrepreneurship Theory and Practice*, *38*(6), 1261-1283.

Chen, G., Chittoor, R., & Vissa, B. (2021). Does nepotism run in the family? CEO pay and pay-performance sensitivity in Indian family firms. *Strategic Management Journal*, 42(7), 1326-1343.

Chua, J. H., Chrisman, J. J., & Sharma, P. (2003). Succession and non-succession concerns of family firms and agency relationship with nonfamily managers. *Family Business Review*, 16(2), 89-107.

Chua, J. H., Chrisman, J. J., & Chang, E. P. (2004). Are family firms born or made? An exploratory investigation. *Family Business Review*, 17(1), 37-54.

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.

Deephouse, D. L., & Jaskiewicz, P. (2013). Do family firms have better reputations than non-family firms? An integration of socioemotional wealth and social identity theories. *Journal of Management Studies*, *50*(3), 337-360.

Delis, M. D., Hasan, I., & Ongena, S. (2020). Democracy and credit. *Journal of Financial Economics*, 136(2), 571-596.

Duran, P., Kammerlander, N., Van Essen, M., & Zellweger, T. (2016). Doing more with less: Innovation input and output in family firms. *Academy of Management Journal*, *59*(4), 1224-1264.

Faccio, M., & Lang, L. H. (2002). The ultimate ownership of Western European corporations. *Journal of Financial Economics*, *65*(3), 365-395.

Gómez-Mejía, L. R., Haynes, K. T., Núñez-Nickel, M., Jacobson, K. J., & Moyano-Fuentes, J. (2007). Socioemotional wealth and business risks in family-controlled firms: Evidence from Spanish olive oil mills. *Administrative Science Quarterly*, 52(1), 106-137.

Gomez-Mejia, L. R., Makri, M., & Kintana, M. L. (2010). Diversification decisions in family-controlled firms. *Journal of Management Studies*, 47(2), 223-252.

Gomez-Mejia, L. R., Cruz, C., Berrone, P., & De Castro, J. (2011). The bind that ties: Socioemotional wealth preservation in family firms. *Academy of Management Annals*, *5*(1), 653-707.

Gómez-Mejía, L. R., Sanchez-Bueno, M. J., Miroshnychenko, I., Wiseman, R. M., Muñoz-Bullón, F., & De Massis, A. (2023). Family Control, Political Risk and Employment Security: A Cross-National Study. *Journal of Management Studies*, forthcoming.

Hansen, C., & Block, J. (2021). Public family firms and capital structure: a meta-analysis. *Corporate Governance: An International Review*, *29*(3), 297-319.

Jha, M., Qian, J., Weber, M., & Yang, B. (2023). ChatGPT and corporate policies. Working paper.

Li, E. X., Tu, Z., & Zhou, D. (2023). The promise and peril of generative AI: Evidence from ChatGPT as sell-side analysts. Working paper.

Li, K., Mai, F., Shen, R., Yang, C., & Zhang, T. (2023). Dissecting corporate culture using generative AI – Insights from analyst reports. *Working paper.*

Loderer, C., Stulz, R., & Waelchli, U. (2017). Firm rigidities and the decline in growth opportunities. *Management Science*, *63*(9), 3000-3020.

Miller, D., & Le Breton-Miller, I. (2021). Family firms: A breed of extremes? *Entrepreneurship Theory and Practice*, *45*(4), 663-681.

Miller, D., Le Breton-Miller, I., Lester, R. H., & Cannella Jr, A. A. (2007). Are family firms really superior performers? *Journal of Corporate Finance*, 13(5), 829-858.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *Journal of Finance*, *54*(2), 471-517.

Le Breton-Miller, I., & Miller, D. (2009). Agency vs. stewardship in public family firms: A social embeddedness reconciliation. *Entrepreneurship Theory and Practice*, 33(6), 1169-1191.

Le Breton-Miller, I., Miller, D., & Lester, R. H. (2011). Stewardship or agency? A social embeddedness reconciliation of conduct and performance in public family businesses. *Organization Science*, 22(3), 704-721.

Mariani, M. M., Al-Sultan, K., & De Massis, A. (2023). Corporate social responsibility in family firms: A systematic literature review. *Journal of Small Business Management*, *61*(3), 1192-1246.

Marques, P., Presas, P., & Simon, A. (2014). The heterogeneity of family firms in CSR engagement: The role of values. *Family Business Review*, *27*(3), 206-227.

Miller, D., Minichilli, A., & Corbetta, G. (2013). Is family leadership always beneficial? *Strategic Management Journal*, *34*(5), 553-571.

Miller, D., Le Breton-Miller, I., Amore, M. D., Minichilli, A., & Corbetta, G. (2017). Institutional logics, family firm governance and performance. *Journal of Business Venturing*, *32*(6), 674-693.

Miroshnychenko, I., De Massis, A., Miller, D., & Barontini, R. (2021). Family business growth around the world. *Entrepreneurship Theory and Practice*, 45(4), 682-708.

Parise, G. (2023). Revisiting family firms. *Journal of Financial and Quantitative Analysis*, forthcoming.

Pinelli, M., Chirico, F., De Massis, A., & Zattoni, A. (2023). Acquisition relatedness in family firms: Do the environment and the institutional context matter? *Journal of Management Studies*, forthcoming.

Thomsen, S., Poulsen, T., Børsting, C., & Kuhn, J. (2018). Industrial foundations as long-term owners. *Corporate Governance: An International Review*, *26*(3), 180-196.

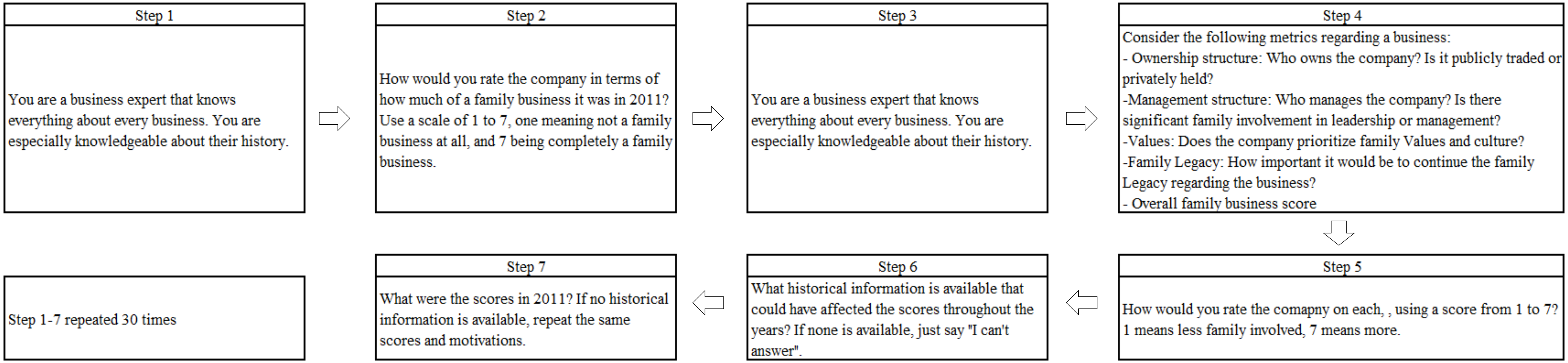
Villalonga, B., & Amit, R. (2006). How do family ownership, control and management affect firm value? *Journal of Financial Economics*, 80(2), 385-417.

Wagner, D., Block, J. H., Miller, D., Schwens, C., & Xi, G. (2015). A meta-analysis of the financial performance of family firms: Another attempt. *Journal of Family Business Strategy*, *6*(1), 3-13.

Zellweger, T. M., Eddleston, K. A., & Kellermanns, F. W. (2010). Exploring the concept of familiness: Introducing family firm identity. *Journal of Family Business Strategy*, 1(1), 54-63.

Zellweger, T. M., Kellermanns, F. W., Chrisman, J. J., & Chua, J. H. (2012). Family control and family firm valuation by family CEOs: The importance of intentions for transgenerational control. *Organization Science*, 23(3), 851-868.

Figure 1.
Data collection process



Data was collected from June 27th to July 7th, 2023 using chat completions API in GPT 3.5 Turbo. Answers to each question were provided on a Likert scale (1-7).

Figure 2.
AI family score for family and non-family firms

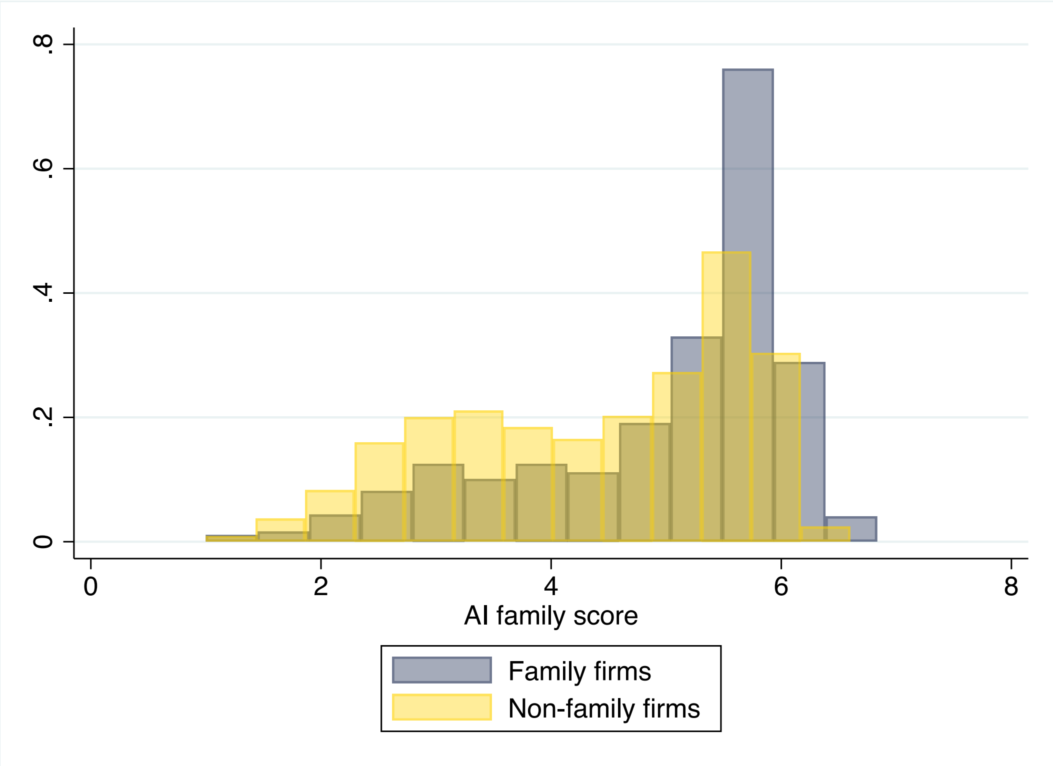


Figure 3.
AI family score by family firm leadership

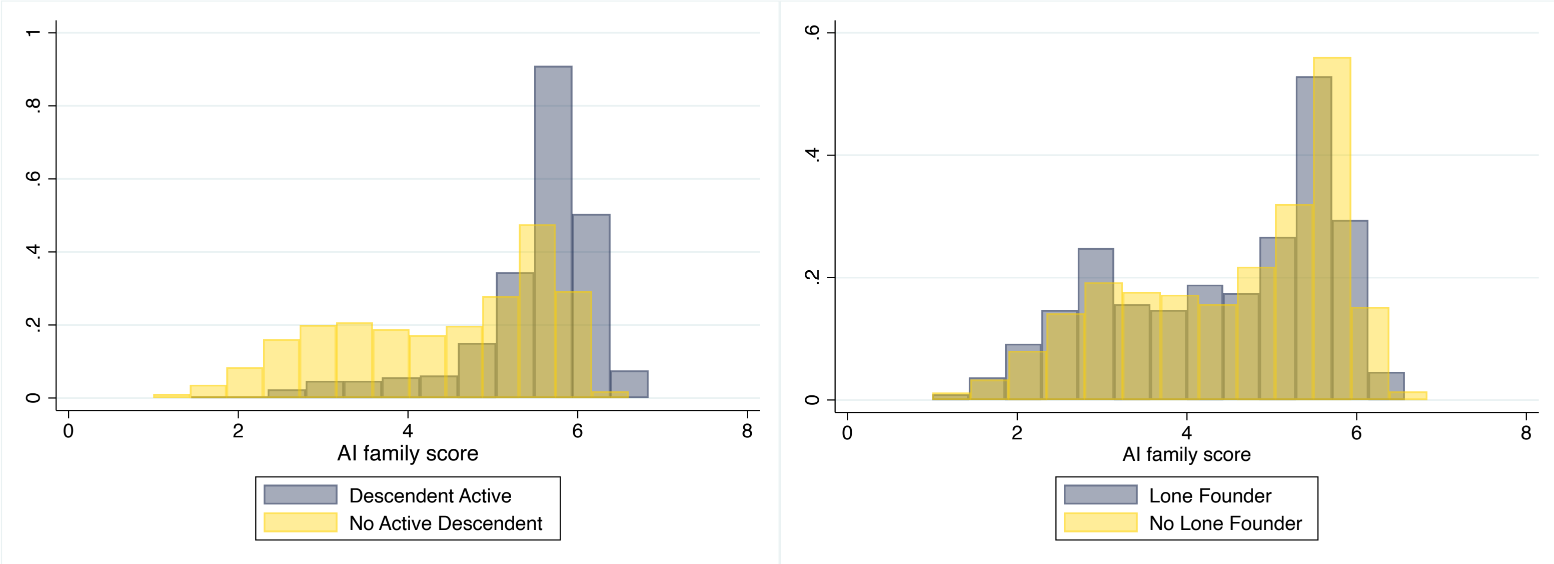


Figure 4.
AI family score by degree of family involvement in governance

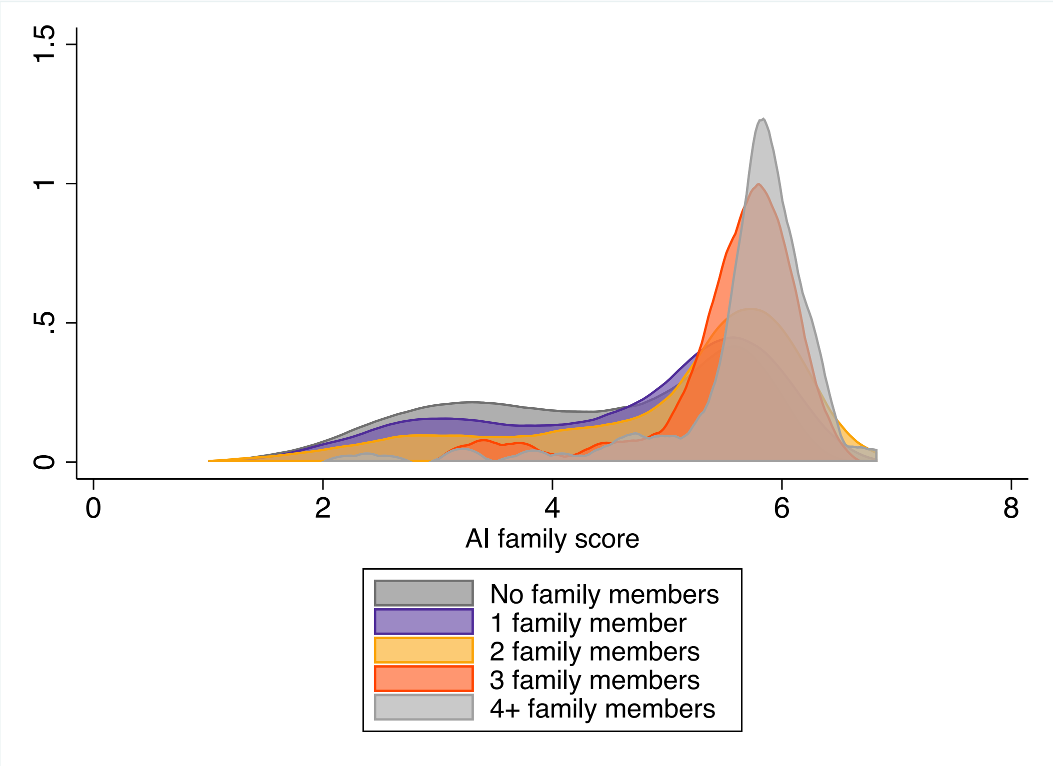


Figure 5.
Family firm classifications and financial returns

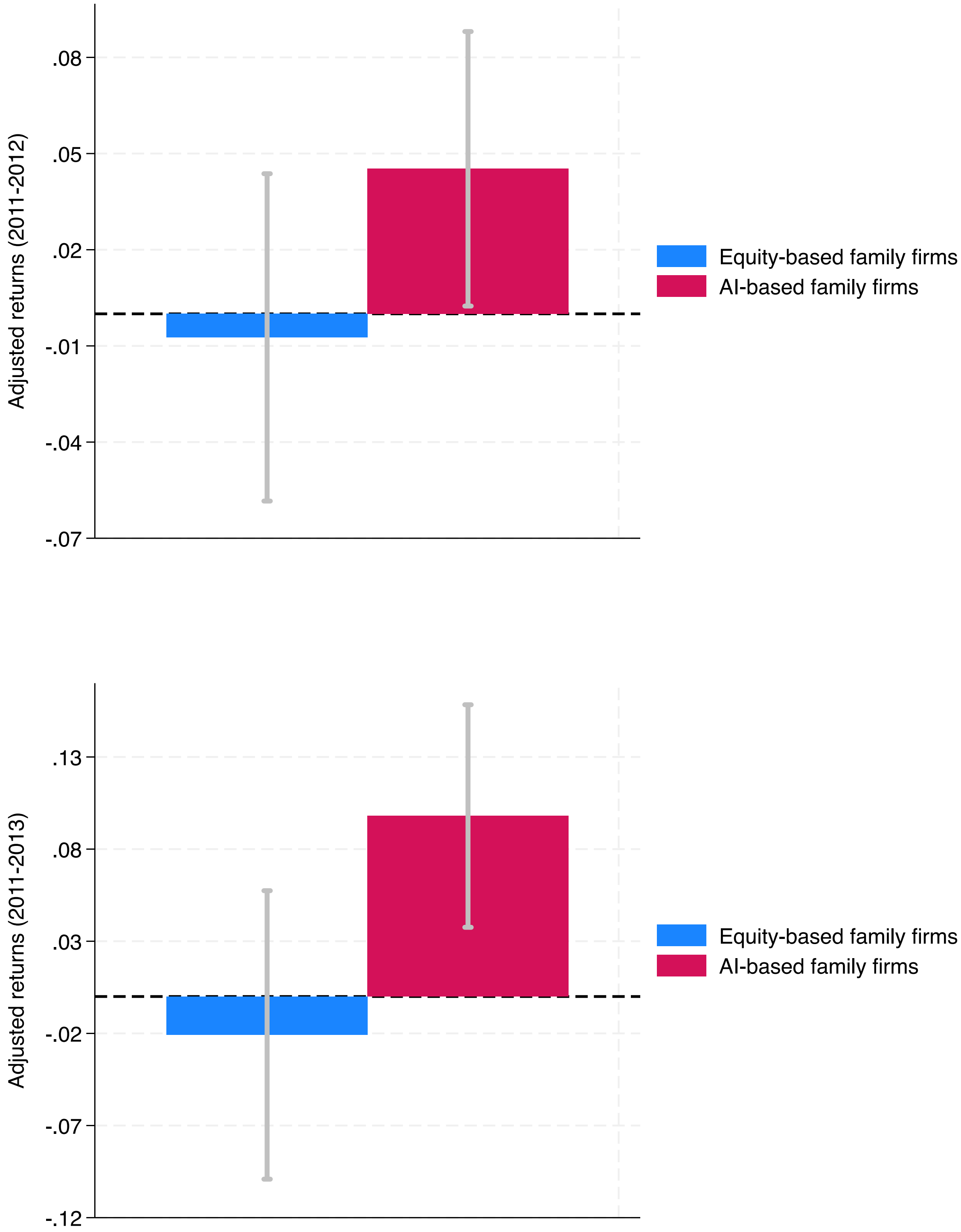


Table 1.
Descriptive statistics

Panel A. Firm characteristics	N	Mean	s.d.	Min	Max
Family firm	3,864	0.211	0.408	0	1
Family ownership	3,864	0.074	0.173	0	0.980
Eponymous firm	3,864	0.164	0.370	0	1
Lone founder	3,864	0.131	0.338	0	1
Family directors	3,864	0.314	0.464	0	1
Descendant active	3,864	0.122	0.328	0	1
Sales	3,864	7.725	2.688	0	18.907
Firm age	3,864	3.686	0.921	0	5.991
Debt to assets	3,864	0.457	0.203	0	0.999

Panel B. AI scores	N	Mean	s.d.	Min	Max
AI family score	3,864	4.532	1.262	1	6.833
AI family ownership score	3,864	2.523	0.776	1	6.233
AI family management score	3,864	2.827	0.978	1	6.500
AI family values score	3,864	3.360	0.750	1.333	6.033
AI family legacy score	3,864	2.708	1.505	1	7

Panel C. Correlations	1.	2.	3.	4.	5.
1. AI family score	1.000				
2. AI family ownership score	0.672	1.000			
3. AI family management score	0.720	0.913	1.000		
4. AI family values score	0.675	0.831	0.922	0.879	
5. AI family legacy score	0.706	0.766	0.913	0.975	0.885

The table includes summary statistics on the sample of firms in 2011. *AI Family Score* is the score computed by Chat GPT about the extent to which it would consider the firm as a family firm on a scale of 1 to 7. The procedure has been repeated 30 times, and *AI Family Score* is the average between these computations. A similar procedure was used to compute *AI family ownership score*, *AI family management score*, *AI family history score*, and *AI family values score*. Refer to Table 1 for the specific queries used to derive the scores. *Family firm* is a dummy with a value of 1 if a family owns at least 5% of the firm's equity; 0 elsewhere. Family ownership is the continuous equity stake owned by the family. *Eponymous firm* is a dummy with a value of 1 if the firm is eponymously named; 0 elsewhere. *Lone founder* is a dummy with a value of 1 if an individual is one of the company's founders with no other family members involved and is also an insider (officer or director) or a large owner (5% or more of the firm's equity); 0 elsewhere. *Family directors* is a dummy with a value of 1 if at least a member of the family sits on the board of the firm; 0 elsewhere. *Descendent active* is a dummy with a value of 1 if a descendent of the founder is actively involved in the management of the company as a CEO, chairman, or member of the board; 0 elsewhere. *Sales* is the natural logarithm of one plus the sales of the firm. *Firm age* is the logarithm of one plus the age of the firm (in years). *Debt to assets* is the ratio between common equity and assets. If accounting variables in 2011 were missing, we used the latest available year.

Table 2.
Share of family firms by country

Country	Aminadav & Papaioannou	5% family ownership	Any family ownership	AI classific-ation (mean)	AI classific-ation (median)	Difference (4)-(3)	Difference (5)-(3)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Australia	0.05	0.15	0.24	0.25	0.12	0.11	-0.12
Belgium	0.17	0.18	0.21	0.60	0.48	0.39	0.27
Denmark	0.07	0.09	0.10	0.67	0.57	0.57	0.47
Finland	0.09	0.18	0.18	0.37	0.26	0.19	0.08
France	0.29	0.39	0.43	0.41	0.28	0.02	-0.15
Germany	0.26	0.32	0.36	0.51	0.38	0.15	0.02
Greece	0.52	0.55	0.56	0.96	0.93	0.40	0.37
Hong Kong	0.21	0.33	0.38	0.73	0.69	0.35	0.31
India	0.21	0.25	0.31	0.76	0.68	0.45	0.37
Italy	0.36	0.44	0.50	0.86	0.70	0.36	0.20
Japan	0.04	0.02	0.04	0.70	0.50	0.66	0.46
Netherlands	0.11	0.12	0.15	0.44	0.32	0.29	0.17
Norway	0.13	0.18	0.18	0.53	0.44	0.35	0.26
Singapore	0.19	0.19	0.25	0.65	0.55	0.40	0.30
Spain	0.22	0.26	0.26	0.70	0.56	0.44	0.30
Sweden	0.13	0.16	0.19	0.37	0.21	0.28	0.02
Switzerland	0.20	0.22	0.24	0.57	0.45	0.33	0.21
USA	0.16	0.19	0.28	0.27	0.16	-0.01	-0.08
UK	0.10	0.17	0.22	0.29	0.20	0.07	-0.02

Column (1) reports of the share of family firms from Aminadav and Papaioannou (2020). Column (2) reports the share of family firms in our sample by country (focusing on those countries with at least 50 firms); in this column, family firms are those firms in which the family owns at least 5% of the firm's equity. Column (3) reports the share of family firms in our sample by country (focusing on those countries with at least 50 firms); in this column, family firms are those firms in which the family owns any equity. Column (4) reports the share of firms for which the average family score across the 30 attempts in ChatGPT is higher than 5 (which is the value corresponding to "somewhat agree" that the firm is a family firm). Column (5) reports the share of firms for which the median family score across the 30 attempts is higher than 5 (which is the value corresponding to "somewhat agree" that the firm is a family firm). Columns (6) and (7) report the differences between the shares of family firms as defined by ChatGPT (Columns 4, 5) and the share of firms with any family ownership (Column 3).

Table 3.
AI family score and family firm classification

DV: Family firm						
	(1)	(2)	(3)	(4)	(5)	(6)
AI family score	0.061***	0.058***	0.062***	0.062***	0.067***	0.064***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Sales		-0.024***	-0.021***	-0.020***	-0.016***	-0.005
		(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Firm age			-0.036***	-0.035***	-0.038***	-0.038***
			(0.007)	(0.007)	(0.007)	(0.007)
Debt to assets				0.036	0.027	0.085**
				(0.033)	(0.034)	(0.035)
Observations	3,864	3,864	3,864	3,864	3,864	3,864
Industry fixed effects	No	No	No	No	Yes	Yes
Country fixed effects	No	No	No	No	No	Yes

This table presents the results of OLS regressions. The main variables are described in Table 1. Columns 5 and 6 include industry fixed effects (2-digit SIC). Column 6 includes country fixed effects. Robust standard errors are reported in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4a.

AI family score and family firm classification – by geographic area

DV: Family firm	Europe	North America	Asia	Rest of the world
	(1)	(2)	(3)	(4)
AI family score	0.058***	0.075***	0.059***	0.066***
	(0.010)	(0.011)	(0.016)	(0.022)
Sales	-0.007	0.011	-0.003	-0.009
	(0.006)	(0.009)	(0.008)	(0.012)
Firm age	-0.043***	-0.056***	0.022	0.002
	(0.011)	(0.014)	(0.023)	(0.028)
Debt to assets	0.083	0.197***	0.010	0.012
	(0.054)	(0.065)	(0.089)	(0.134)
Observations	1,923	1,151	542	216
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes

This table presents the results of OLS regressions. The main variables are described in Table 1. All Columns include industry fixed effects (2-digit SIC) and country fixed effects. We run separate regressions for different macro areas (see Table A1). Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 4b.

AI family score and family firm classification – by firm size

DV: Family firm	Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)
AI family score	0.037**	0.091***	0.079***	0.048***
	(0.016)	(0.013)	(0.012)	(0.011)
Sales	0.002	0.058	0.006	0.018**
	(0.011)	(0.036)	(0.029)	(0.009)
Firm age	-0.069***	-0.037**	-0.046***	-0.017
	(0.021)	(0.015)	(0.014)	(0.013)
Debt to assets	-0.016	0.186***	0.071	0.059
	(0.070)	(0.070)	(0.080)	(0.070)
Observations	957	954	957	959
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes

This table presents the results of OLS regressions. The main variables are described in Table 1. All Columns include industry fixed effects (2-digit SIC) and country fixed effects. We run separate regressions for different quartiles of firm size. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 5.
AI family score and family firm attributes

DV:	Eponymous firm	Lone founder	Family directors	Descendent active
	(1)	(2)	(3)	(4)
AI family score	0.024***	0.007	0.081***	0.063***
	(0.002)	(0.005)	(0.007)	(0.005)
Sales	-0.002	-0.011***	-0.005	0.007**
	(0.002)	(0.003)	(0.004)	(0.003)
Firm age	-0.011***	-0.074***	-0.079***	0.016***
	(0.003)	(0.006)	(0.008)	(0.006)
Debt to assets	0.005	0.106***	0.130***	-0.022
	(0.014)	(0.031)	(0.039)	(0.025)
Observations	3,864	3,864	3,864	3,864
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes

This table presents the results of OLS regressions. The main variables are described in Table 1. All Columns include industry fixed effects (2-digit SIC) and country fixed effects. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 6.
AI family sub-scores and family firm classification

DV: Family firm	(1)	(2)	(4)	(5)
AI family ownership score	0.048***			
	(0.010)			
AI family management score		0.071***		
		(0.008)		
AI family values score			0.096***	
			(0.011)	
AI family legacy score				0.071***
				(0.005)
Sales	-0.009**	-0.006	-0.007*	-0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Firm age	-0.029***	-0.036***	-0.038***	-0.051***
	(0.008)	(0.008)	(0.008)	(0.008)
Debt to assets	0.091***	0.092***	0.088**	0.088**
	(0.035)	(0.035)	(0.035)	(0.034)
Observations	3,864	3,864	3,864	3,864
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes

This table presents the results of OLS regressions. The main variables are described in Table 1. All Columns include industry fixed effects (2-digit SIC) and country fixed effects. Standard errors are heteroskedasticity-adjusted. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table A1.
Geographic distribution of firms

Country	N. Firms	Percentage
Australia	136	3.52
Austria	36	0.93
Belgium	62	1.6
Brazil	46	1.19
Croatia	6	0.16
Cyprus	3	0.08
Czech	7	0.18
Denmark	69	1.79
Finland	84	2.17
France	157	4.06
Germany	227	5.87
Greece	137	3.55
Hong Kong	64	1.66
Hungary	6	0.16
India	75	1.94
Indonesia	29	0.75
Ireland	31	0.8
Israel	11	0.28
Italy	116	3
Japan	184	4.76
Malaysia	22	0.57
Mexico	11	0.28
Netherlands	68	1.76
New Zealand	11	0.28
Norway	122	3.16
Philippines	25	0.65
Poland	19	0.49
Portugal	25	0.65
Qatar	8	0.21
Romania	5	0.13
Russia	43	1.11
Singapore	63	1.63
Slovenia	12	0.31
South Africa	29	0.75
South Korea	20	0.52
Spain	66	1.71
Sweden	151	3.91
Switzerland	99	2.56
Taiwan	19	0.49
Thailand	29	0.75
Turkey	41	1.06
USA	1155	29.89
United Kingdom	335	8.67
Total	3,864	

Europe includes Austria, Belgium, Croatia, Cyprus, Czech, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom. North America includes USA and Canada. Asia includes Hong Kong, India, Indonesia, Israel, Japan, Malaysia, Philippines, Qatar, Singapore, South Korea, Taiwan, Thailand. Rest of the world includes Australia, Brazil, Mexico, New Zealand, South Africa.

Appendix Table A2.
Industry distribution of firms

Country	N. Firms	Percentage
Automobiles & Parts	79	2.04
Basic Resources	240	6.21
Chemicals	134	3.47
Construction & Materials	213	5.51
Food & Beverage	218	5.64
Health Care	317	8.2
Industrial Goods & Services	852	22.05
Media	148	3.83
Oil & Gas	276	7.14
Personal & Household Goods	256	6.63
Retail	244	6.31
Technology	429	11.1
Telecommunications	101	2.61
Travel & Leisure	178	4.61
Utilities	179	4.63
Total	3,864	

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html