The cultural origins of family firms *

Jian Xie[†]
Southern University of Science and Technology

Song Yuan[‡]
Zhejiang University

December 10, 2024

Abstract

What determines the prevalence of family firms? In this project, we investigate the role of historical family culture in the spatial distribution of family firms. Using detailed firm-level data from China, we find that there is a larger share of family firms in regions with a stronger historical family culture, as measured by genealogy density. The results are further confirmed by an instrumental variables approach and a matching approach. Examining mechanisms, we find that entrepreneurs in regions with a stronger historical family culture: i) tend to have family members engage more in firms; ii) are more likely to raise initial capital from family members; iii) are more willing to pass on the firms to their children. Historical family culture predicts better firm performance partly due to a lower leverage ratio.

Keywords: Capital; China; Cultural Origins; Entrepreneurship; Family Culture; Family Firms; Firm Ownership; Firm Performance; Genealogy; Kinship Networks

JEL classification: D2, D02, G3, L2, M1, Z1

*We thank the editor, two anonymous reviewers, Robert Akerlof, Jitendra Aswani, Manuel Bagues, Sascha Becker, Gustavo Cortes, Jana Fidrmuc, Bishnupriya Gupta, Eric Hilt, Clement Imbert, Ruixue Jia, Sharun Mukand, Chicheng Ma, Debin Ma, Yancheng Qiu, Carol Shiue, John Thanassoulis, Melanie Meng Xue, and participants in the CAGE AMES seminar, Warwick Business School, Peking University, Applied Young Economist Webinar, China Financial Research Conference, LSE Economic History Graduate Seminar, Economic History Society Conference, the Annual Meeting of Economic History Association, OWL workshop, NEUDC conference, and Annual Symposium on Quantitative History, for their helpful comments. We are grateful to James Fenske and Daniel Sgroi for their guidance and support. Jian Xie acknowledges financial support from the National Natural Science Foundation of China (Grant No. 72394391) and the Social Economics Teaching and Research Center of SUSTECH (SJZLGC202310). Song Yuan thanks financial support from Zhejiang University Real Estate Investment Research Center (Grant No. EI22003) and Special Project Grant at Academy of Social Sciences at Zhejiang University (Grant No. S20230033) . All authors contribute equally to the paper. All errors are our own.

[†]Corresponding author. College of Business, Southern University of Science and Technology. Email: xiej@sustech.edu.cn

[‡]School of Economics, Zhejiang University. Email: song.yuan@zju.edu.cn

1. Introduction

Family firms have attracted increasing interest and attention due to their economic importance over the past years. Around 65% to 80% of all firms in the world are family firms (Davis et al., 1997; Cai et al., 2013). In the U.S., around 47% of the largest 2000 industrial firms retain founding family ownership (Anderson et al., 2009). Among publicly listed international firms, 45% are family-controlled (La Porta et al., 1999; Ellul et al., 2010). Family businesses are common in Latin America, Africa, the Middle East, and parts of western Europe and Asia, though there is a lot of heterogeneity across regions (Bertrand and Schoar, 2006). A growing body of research has studied the differences in behavior and performance between family firms and their non-family counterparts, which could be crucial in understanding local economic outcomes. What determines the prevalence of family firms, an important aspect of local industry structure?

Weber (1904) argues that strong family values deter the development of individual forms of entrepreneurship, which requires trust towards strangers. If this assertion holds, it raises the question: do strong family values facilitate or hinder the development of family firms and further influence the composition of firms within a region? In this paper, we explore the role of historical family culture in the prevalence or the share of modern family firms in the context of China.[1] Family culture is the "cultural patterns of family loyalties, allegiances and authorities" and is usually defined as the individual values that stress the link between family members and loyalty to the family (Algan and Cahuc, 2007; Alesina and Giuliano, 2014).[2] Family culture is documented to influence economic outcomes, individual behaviors and likely the establishment of family firms. For instance, entrepreneurs with strong family ties are more likely to raise initial capital from their family members and might assign important positions in their firms to their relatives due to obligation or trust. They may also be more inclined to build a family legacy to ensure survival and family control (Bertrand and Schoar, 2006).[3] Therefore, regions with a stronger family culture are more likely to have a larger share of family firms. However, the empirical evidence on this question remains scarce.

China is an ideal setting for studying this topic. First, family firms are highly prevalent and play an essential role in the Chinese economy, especially within the private sector. In 2010, around 50% of private firms in China are family firms.[4] Importantly, the private sector was shut down during the socialist transformation in China from 1953 to 1956, and almost all current private firms were established after the reform and opening policy in

[1] We use historical family culture in this project because the proxy measure of family culture is not contemporary but prior to 1949.

[2] Family culture and family ties are highly similar. They are interchangeable in our paper.

[3] Bertrand and Schoar (2006) finds that family ties positively correlate with establishment size and the fraction of self-employed at the country level. However, the fraction of family firms is regarded as the ideal measure of industry structure outcome, which does not exist across countries.

[4] The number is calculated from the Chinese Private Enterprise Survey in 2010 based on our baseline definition of family firms.

1978. Chinese culture also has a long history of emphasizing family ties with regional variations, and the measure of historical family culture adopted was predetermined before the emergence of the contemporary private economy. This offers a unique opportunity to study the roots of variation in the prevalence of family firms since these firms are entirely new for all regions. Second, focusing on a single country at a disaggregated level helps us to exclude country level and sub-national level confounding factors in the estimation. It also addresses the practical issue that the key measure of family firm prevalence, the fraction of family firms, does not exist (or is not comparable) in many countries.

To investigate the historical and cultural roots of the prevalence of family firms, we gather detailed data on firms and historical family culture at the county level in China. The analysis focuses on unlisted private firms, which comprise most of the Chinese economy, face fewer strict regulations, and have fewer outside investors than listed firms. In addition to the standard balance sheets, the firm data contains extensive information on the resumes of the entrepreneurs and their family members' engagement in the respective firms in 2010. This allows us to plausibly distinguish between family[5] and non-family firms and test the underlying mechanisms. We create our dependent variable based on the detailed firm dataset, the fraction of family firms in each county.

The empirical strategy to test the hypothesis requires a plausible measure of historical family culture. We make use of historical genealogy (1368AD-1949AD) to proxy the historical family culture.[6] Specifically, we construct the culture measure – historical genealogy density – by normalizing the number of genealogy books using the population in 1953, which helps us to address the reverse causality issue in the identification and provides more statistical power with the disaggregated county level data.[7] Genealogy describes family history and relationships among lineage members across and within generations. It usually contains basic information about each family member, important events in family history and information on family properties or graves. It reflects family culture because it involves living in an extended family and shows that people are interested in their family origins, lineage continuity and intra-lineage member connections (Shiue, 2016). It serves the purposes of worshipping ancestors and uniting family members (Feng and Chang, 2008). It has practical functions such as land allocation, providing public goods, organizing defences, and other joint activities. It indicates that individuals within the extended family rely on the family as a provider of services, insurance and other resources, which is a form of strong family culture (Alesina and Giuliano, 2011). As noted in *The Law of the Clan* (1020AD-

[5]Our baseline definition for the family firm is that a family owns at least 50% of the equity and at least another family member (other than the founder) engages in the firm (either holds positive equity shares or a board member position or the CEO position).

[6]The interpretation of genealogy varies across the literature. For instance, Greif and Tabellini (2017) firstly use genealogy books to measure the number and origins of clans in China. Chen et al. (2022) use genealogy books to measure the clan culture emphasizing the influence of Confucian culture. In line with existing literature, we focus more on the contemporary implications of historical clans, since large clans gradually decomposed, and households and families became the norm.

[7]When calculating historical genealogy density, the county level is the smallest feasible level in China.

1077AD),[8] without genealogies, the extended families do not know their origins and cannot be kept together very long.

Furthermore, we perform a validation test to show that historical genealogy density has strong predictive power for people's beliefs and behaviors related to family culture, using a large scale detailed individual level survey in China. Specifically, people in regions with a higher genealogy density tend to believe that family is important, and have closer family relationships and more interactions with other family members. In particular, we show that their correlation is robust with controlling for other cultural norms, which might be captured by genealogy density, such as trust, clan culture, and collectivist culture.

We find that historical genealogy density is positively correlated with the share of family firms in counties. Empirically, a one standard deviation increase in genealogy density leads to a 5.2 percentage points increase in the fraction of family firms. The effect is economically significant, considering that the average fraction of family firms is around 50%. We use several strategies to prove that omitted variables, such as historical economic development and institutions that might promote both genealogy book compilation and family firms, do not drive our results. First, we control for an extensive set of confounders, including geographic variables, measures of historical prosperity, and contemporary socioeconomic variables. Second, we adopt an instrumental variable, based on the influence of Confucian academies during the Song dynasty (960AD-1279AD) within 200 km. These academies primarily promoted a new family clan system, and we do not find that they have persistent effects on other contemporary economic outcomes. These findings are also confirmed by a matching approach comparing neighbor counties, assuming that neighboring counties share similar unobserved characteristics. This assumption is plausible in our setting as the counties are usually formed around rivers, and the county boundaries are usually natural borders such as rivers and mountains. By exploiting the spatial variation of genealogy distributions across neighboring counties, the matching estimation finds similar effects of historical culture as the baseline results. We also use an analogous matching approach based on GDP per capita or average nighttime light to show that family culture drives the establishment of family firms in counties with similar economic development levels.

We demonstrate additional heterogeneous analysis to explore how family culture affects the prevalence of family firms. Since the measure of historical family culture – genealogy density – covers different periods, we first show that older genealogies display a larger impact on the fraction of family firms, indicating the persisting effect of family culture. Further, we show that the impact of family culture is larger in regions with a lower education level, implying that family culture may compensate for the lack of talents and provide human capital support for the family business. We also find that the impact of family culture is greater in regions with lower quality public administration or higher levels of marketization. This suggests that family culture protects entrepreneurs from political abuse by local officials

[8] *Zong Fa* (宗法), by Zhang Zai

and that family networks complement market forces, facilitating the establishment of more family firms.

We test three mechanisms through which family ties might drive the presence and prevalence of family firms. In families with strong ties, firm owners would take account of family members' participation in firms in their utility, and have a stronger desire to maintain control within the family and build a family legacy. In light of these concerns, family firms would, therefore, have family members own more equity shares or occupy more managerial and director positions, making family firms persist over generations. Empirically, counties with one standard deviation higher genealogy density see 2.1% more equity shares owned by family members, and the fractions of board members, CEOs, and managers who are family members increase by 2.6%, 1.0% and 1.7%, respectively. Another channel is that family ties could substitute for formal institutions regarding financing and investor protection due to the high trust, altruism, reciprocity or sense of duty among family members. In our setting, we show that in counties with stronger family ties, family firms are more likely to raise the initial capital from family members instead of modern banks, which could further lead to the family members' participation in family firms through ownership of equity shares or by occupying managerial positions. Furthermore, a stronger family culture indicates more incentives to build a family legacy and ensure family control over generations. The findings suggest that entrepreneurs exposed to stronger historical family culture have stronger intentions to pass down firms to their children, and their children, in turn, are more willing to inherit the business, shedding light on the source of the persistence of family firms.

We further study the firm performance implications of historical family culture since it has strong predictive power for corporate governance. The empirical results demonstrate that, on average, firms in counties with a stronger historical family culture perform better. The mediation analysis indicates that a lower leverage ratio, which may improve the firm performance,[9] serves as one mechanism since firms in these regions are generally more risk averse and have less debt financing.

This research project is related to several strands of literature. The first one is regarding the spatial distribution of family firms, and more broadly, the factors influencing industry structure and patterns of business ownership within a region.[10] Previous firm-level studies have shown that the gender of a departing CEO's firstborn child (Bennedsen et al., 2007), contract enforcement (Lu and Tao, 2009), social capital (Amore, 2017) and collective culture (Fan et al., 2022) [11] could impact family control and engagement in firms. Yet, little is known about the spatial distribution of family firms across different regions, as measured

[9] See studies by Campello (2006); Dalci (2018); Lang et al. (1996); Opler and Titman (1994)

[10] Cetorelli and Strahan (2006); Pierce and Snyder (2020)

[11] Fan et al. (2022) explore the role of collectivist culture that prioritizes the group over the individual in comparison to individualism in family control among listed firms at the province level. We study a different culture at the county level and are the first to investigate the formation and diffusion of family firms. Moreover, features of family culture, such as a strong intention of building family legacies, are essential for the persistence of family firms over generations.

by the share of family firms at the county level, and the underlying mechanisms. Regional-level analyses on the distribution of family firms could offer deeper insights into the industry structure and patterns of business ownership of a region. Mueller and Philippon (2011) show that family firms are more common in countries where labor relations are hostile, based on around 30 observations. Without a good measure of the prevalence of family firms across countries, Bertrand and Schoar (2006) point out that cultural norms could impact industry structure measured by average establishment size and the fraction of self-employed, which they believe are related to the fraction of family firms. We contribute to this literature by systematically analysing the effect of historical family culture on the spatial distribution of family firms by using the share of family firms as the measure of industry structure at a more disaggregated level.

The second strand of literature to which we contribute is on the impact of family culture. The relationship between family culture and economic outcomes has been discussed extensively in the literature.[12] The majority of the studies claim that stronger family ties impede economic development. In addition, family culture is proven to be correlated with labor market outcomes (Alesina and Giuliano, 2010; Alesina et al., 2015), participation in politics (Alesina and Giuliano, 2011), the emergence of pension systems (Galasso and Profeta, 2018) and cooperative behaviors (Enke, 2019). Relatedly, Chen et al. (2022) adopt a similar genealogy measure and document that historical clan leads to fewer modern banks in China.[13] Zhang (2020) observes that genealogy density correlates with entrepreneurial propensity and the proportion of the economy in the private sector. We differ by studying family firms featuring the dominance of a single family in firms, which has profound implications for firm behaviours and performance. Moreover, we perform the analysis at the firm level and provide new insights into how family culture affects the management, financing, and inheritance of family firms.[14]

The third is the performance and behaviors of family firms. A large group of studies have tried to explore whether family firms perform better or worse than their non-family counterparts.[15] The results seem to be conditional on whether the founders are CEOs (Villalonga and Amit, 2006), whether current CEOs are related to founders (Mehrotra

[12] Banfield (1958); Coleman (1990); Greif (2006); Bertrand and Schoar (2006); Alesina and Giuliano (2014); Greif and Tabellini (2017).

[13] We differ from Chen et al. (2022) by directly documenting that historical family culture could increase intra-lineage financing due to altruism and trust among family members, contributing to the structure formation of the private economy. In particular, we focus on the behaviour and management of private firms while studying the implications of historical family culture from a more micro perspective.

[14] This project is also broadly related to the literature studying the impact of culture on economic outcomes through shaping institutions (Doepke and Zilibotti, 2008; Tabellini, 2010; Spolaore and Wacziarg, 2013; Michalopoulos and Papaioannou, 2013; Giavazzi et al., 2013; Padró i Miquel et al., 2015; Currie et al., 2016). Since family firms perform and behave differently compared to non-family firms in many respects, we contribute to this literature by providing evidence that family culture could affect economic outcomes by influencing the establishment of family firms and further the industry structure as measured by the share of family firms in a region.

[15] Anderson and Reeb (2003); Bennedsen et al. (2007); Lins et al. (2013); Sraer and Thesmar (2007).

et al., 2013; Pérez-González, 2006; Bertrand et al., 2008), whether Confucianism influences the selection of successors (Chen et al., 2021), and the industries the family firms are in (Bennedsen et al., 2007). Amit and Villalonga (2014) have made a detailed summary of this literature.[16] We contribute to this literature first by investigating the role of family culture in corporate governance in terms of shareholder structure, hiring decisions, and financing. Secondly, we provide evidence that historical family culture could partly explain firm performance.

The rest of the paper is organized as follows. In Section 2, we introduce the background of historical family culture and family firms in China. In Section 3, we describe the data we use in this project and the validation of our genealogy measure. In Section 4, we present the main results, including OLS, IV and the Matching estimation. In Section 5, we discuss the channels through which family ties predict the presence and prevalence of family firms. In Section 6, we summarize the robustness checks we have conducted. In Section 7, we provide the firm performance implications of historical family culture. Section 8 is the conclusion.

2. Background

2.1. Historical family culture

Alesina and Giuliano (2014) define family culture as "cultural patterns of family loyalties, allegiances and authorities".[17] A strong family culture indicates that more individuals rely on the family as a provider of services, insurance, and other resources (Alesina and Giuliano, 2011).[18] The traditional family in China is composed of parents, children, siblings, and other extended family members. Chinese families in history have several distinguishing features compared to western nuclear families. First, they are larger in scale, and co-residence of several generations is preferred. The ideal historical Chinese family consists of five generations living together as one unit (Baker, 1979), forming the foundation of family connections. Second, historical Chinese family culture advocates collectivism and individuals are dominated by families. Third, under the influence of Confucianism, parents and children, males and females, and elder and younger siblings adhere to different roles, under the principle of respecting elders and maintaining a harmonious relationship. Historical Chinese family culture particularly emphasizes the importance of the family, filial piety, and mutual support among family members. This value system was primarily constructed

[16] It has also been discovered that family firms are less tax aggressive (Chen et al., 2010), experience higher abnormal short sales prior to adverse earnings shocks (Anderson et al., 2012), invest less when inheritance law is stricter (Ellul et al., 2010; Tsoutsoura, 2015), and invest more in employee relations (Kang and Kim, 2020). In addition, founders play more important roles in family firms than professional managers in non-family firms (Becker and Hvide, 2019).

[17] Family culture and family ties are highly similar. They are interchangeable in our paper.

[18] Alesina and Giuliano (2010) measure family ties using individual responses from the World Value Survey regarding the importance of the family and the love and respect that children need to have for their parentsDesmet et al. (2017) also use questions in World Value Survey to measure cultural values.

by Confucius and his disciples.

The Song dynasty (960AD-1279AD) is regarded as the most crucial period in Chinese history in establishing a new type of family clan system, which was largely promoted by Song Confucians (Feng and Chang, 2008; Biao, 2007). After the collapse of the aristocracies of the Tang dynasty (618AD-907AD), Song Confucians needed to develop a new local system to adapt to social changes. First, they reinterpreted the literature and compiled books to solve the theoretical problems of the new type of family clan organization and family structure. They also provided guidelines on how to behave and organize activities within extended families. For example, *Family Rituals*,[19] which was written by Zhu Xi in the spirit of maintaining family hierarchy, love and respect among family members, contains detailed discussions on building ancestral halls, and organizing ceremonies like marriages, funerals and family sacrifices.[20] Second, Song Confucians developed various practical and standard methods and models to construct the family clan system at the local and grassroot levels.[21] Furthermore, some Confucian scholars began to build academies to educate their descendants.[22] These academies helped them compete for educational resources and bureaucratic positions, which in turn fed back into the development of family clans.

Confucian scholars were highly involved with the teaching activities of academies during the Song Dynasty, which facilitated the dissemination of ideas on the family clan system and laid a solid foundation for the development of family culture in later periods. In Chinese history, kinship groups have played a leading role in political and economic affairs (Watson, 1982). Even in contemporary periods, kin-based organizations or networks still serve their purposes of supporting cooperation, for example, in finding jobs and entrepreneurship (Bian, 1994; Peng, 2004).

2.2. Genealogy

We use genealogies as a measure of historical family culture, as illustrated in Shiue (2016). Genealogies (*Jia Pu* (家谱) or *Zu Pu* (族谱)) are the records of family history and the relationship between lineage members within and across generations. In China, the significant parts of genealogies are genealogical tables or charts demonstrating the pedigrees of lineage members organized by the birth order of each generation. They usually contain the basic information about each lineage member, such as the name, birth date, death date, marriage,

[19] *Jia Li* (家礼)

[20] Zhang Zai's chapter, *The Law of the Clan* (*Zong Fa* (宗法) , 1020AD-1077AD), suggests that the construction of the family clan system facilitates the connection among family members and helps maintain stability in local societies, which is imperative for the emperor's political control of the country. Other famous scholars during this period, such as Lv Zuqian, Lu Jiuyuan, Ye Mengde, and Zhao Din, also compiled influential family clan regulations or guidelines to encourage filial piety and unite family members.

[21] For instance, Ouyang Xiu and Su Xun created genealogy templates for commoners to trace family origins and record lineage members (Lv, 2018).

[22] Famous examples include Dongjia academy in Dean County, Hualin academy in Fengxin County and Leitang academy in Jianchang county (Wang, 2013).

and important honors (Hu, 2023).[23] The vast majority of Chinese genealogies are recorded in a patrilineal way.[24] Usually, genealogies are updated and held by the eldest member of the extended family, who passes on the responsibility to the next generation.[25]

The history of written genealogies could be traced back to the Han Dynasty (206 BC-220 AD) (Feng and Chang, 2008). However, ordinary households began to compile their own genealogies without government intervention only since the Song Dynasty.[26] By the Ming and Qing Dynasties (1368AD-1911AD), the genealogical tradition had spread widely among civilians (Harrell et al., 1995). A significant spatial variation could be observed in Figure 1, which is consistent with the common understanding that southeastern parts of China have denser genealogies.

Genealogies reflect historical family culture. The popular convention of compiling genealogies shows ordinary people's interests in family origins, lineage continuity, and intra-lineage member connections (Shiue, 2016). They serve the purposes of worshipping ancestors and uniting family members (Feng and Chang, 2008). Compiling genealogies is a joint work involving an extended family. Some genealogies also have practical functions like land allocation, providing public goods, organizing defenses, and other joint activities. It emphasizes the importance of the family, increases family interactions, and strengthens family identities. It indicates that individuals within the extended family rely on the family as a provider of services, insurance, and other resources, which is a form of strong family culture (Alesina and Giuliano, 2011). [27]

Some of the existing studies use genealogy as a measure of the strength of clan culture.[28] Conceptually, clans are typically regarded as a form of kinship organizations or kinship groups.[29] Clans and extended families differ in generational coverage and scope but are similar in other aspects. Moreover, their boundary is ambiguous and debatable.[30] We use it to measure the historical family culture as it generally reflects people's interests in family origins in both clans and extended families. Our interpretation does not contradict the usage of genealogy in the literature; instead, we complete and extend the implications of

[23] Many genealogies also include sections on lineage history and migration, ancestral halls, graves of prominent (or all) ancestors, and biographies of key lineage members(Harrell, 1987).

[24] There are no universal rules on how to record wives and sisters. In addition, they vary in formats and details regarding whether or how to record adopted sons and lineage members who did not reach adulthood.

[25] Liu et al. (1959), Telford (1986) and Shiue (2016) provide more extensive discussions on the contents of genealogies.

[26] Until the Song Dynasty (960AD-1279AD), preserved genealogies were mainly compiled by royal houses and political elites (Shiue, 2016).

[27] As noted in *The Law of the Clan* (*Zong Fa* (宗法), by Zhang Zai, 1020AD-1077AD), without genealogies, the extended families do not know their origins and cannot be kept together very long.

[28] See Chen et al. (2020); Greif and Tabellini (2017); Zhang (2020).

[29] The discussion can be found in Fei and Liu (1982); Watson (1982); Ebrey et al. (1986); Greif and Tabellini (2010)

[30] For example, Greif and Tabellini (2010) define a clan as a kinship-based community its members identify with and are loyal to. Fei and Liu (1982) regards Chinese clans as family clans. Enke (2019) claims that living in a clan is an index of kinship tightness. Genealogy records not only big kinship-based clans but also smaller extended families.

genealogy in a broader perspective.

As discussed by Greif and Tabellini (2017), after China's economic reform, households, and not clans, were given land-use rights in the former collective farms, and privately-owned businesses were permitted. Some large clans were gradually decomposed, and families and households became the norm. Using a national representative individual-level survey in China, we demonstrate that genealogy has strong predictive power for the concepts and behaviors related to family ties. The validation test is provided in Section 3.5.

2.3. Family firms in China

Typically, family firms are characterized by a concentration of ownership, control and often key management positions among family members (Bertrand and Schoar, 2006; Miller et al., 2007), although the definitions of family firms vary across studies. Based on existing theoretical frameworks, family firms have advantages in several respects. First, they alleviate the conflicts of interest between principals and agents because there is less separation between the two roles (Jensen and Meckling, 1976). Second, family firms have more incentives to embrace a long-term perspective in management due to reputation concerns (Bertrand and Schoar, 2006). Third, there is less information asymmetry between the firm and its large shareholders (Bøhren et al., 2019). Further, family ownership, which is relatively concentrated, may lead to closer monitoring of managers (Bloom and Van Reenen, 2007). However, the demerits of family firms are also evident. For example, nepotism in family firms may lead to low-skilled and undiversified labor forces provided by narrow kinship networks. In addition, adhering to rigid inheritance rules, such as primogeniture or equal sharing rules, could be devastating to family firms when such rules conflict with the development of firms.[31] Because of family firms' coexisting merits and demerits, empirical studies of their performance demonstrate differential results.[32]

Claessens et al. (2000) think that the family firm is a norm in East Asia, where family culture and Confucianism are deeply rooted in the social environment. Prior to 1978, private sectors comprised a tiny fraction of the Chinese economy, since they were basically eliminated during the socialist transformation and cultural revolution. Private firms have developed rapidly since the economic reform and opening up, especially after Deng Xiaoping's famous southern tour in 1992. The growth of family-owned businesses aligns with the expansion of the private sector economy. According to the Chinese Private Enterprise Survey, there were approximately 3.63 million private family firms in China in 2006. PwC's Global Family Business Survey in 2018 reports that there were 27 million private firms by the end of 2017 in China, with family businesses forming the majority. As for publicly listed firms in China, according to the CSMAR dataset, more than half of them are family

[31] Bloom and Van Reenen (2007) find that primogeniture leads to inferior performance for family firms.

[32] Amit and Villalonga (2014) summarize 45 related papers in 20 countries and find that the performance of family firms compared to their non-family counterparts is contingent on family business definition, geographic location, industry affiliation, and intertemporal variation in economic conditions.

firms.[33] This proportion is high compared to listed firms in other countries.[34] Based on these statistics, we can find that family firms are an inseparable and pivotal part of the Chinese economy.

3. Data and variables

3.1. Genealogy data

Our genealogy data is collected from Shanghai library Chinese Genealogy Knowledge Service Platform.[35] This online platform was built for researchers and the general public who are interested in genealogies and their family origins, and based on *The General Catalog of Chinese Genealogy*.[36] It slightly extends the number of genealogies and complements the information on the same genealogies, such as their locations and time. This online platform, in total, includes more than 54,000[37] Chinese genealogies originating from more than 1,600 places in China, covering 608 surnames, and it is the largest Chinese genealogy collection to date.[38] The online platform contains digitized genealogy data, and these genealogy books are mainly preserved in libraries, local archives and households. We use the number of genealogies compiled during 1368[39] to 1949 within a region normalized by the 1953 population (genealogy density thereafter) as a measure of historical family ties. We only consider genealogies prior to 1949 to avoid reverse causality. Unlike previous studies at the provincial or city level, we geo-localize genealogy books by county, ensuring that this variable varies at a granular level. This helps us to control for all observable county level characteristics that may bias our estimation. The geographic distribution of genealogy is displayed in Figure 1.

3.2. Chinese Private Enterprise Survey

We focus on private firms because most family firms in the world are private (Bøhren et al., 2019). The Chinese Private Enterprise Survey (CPES) is a large-scale nationwide survey of privately owned enterprises in China. It was conducted jointly by (i) the United Front Work Department of the CPC Central Committee, (ii) the All-China Federation of Industry and Commerce, (iii) the State Administration for Industry and Commerce of the People's

[33] Renowned examples include Wanda Group, Hengda Group and Midea Group. The fraction of family firms will change under different definitions, but it shows a considerably high fraction of firms are family firms.

[34] According to La Porta et al. (1999), 45% of publicly listed international firms are family-owned. At least one-third of S&P 500 firms are family firms (Anderson and Reeb, 2003).

[35] See `https://jiapu.library.sh.cn/#/`

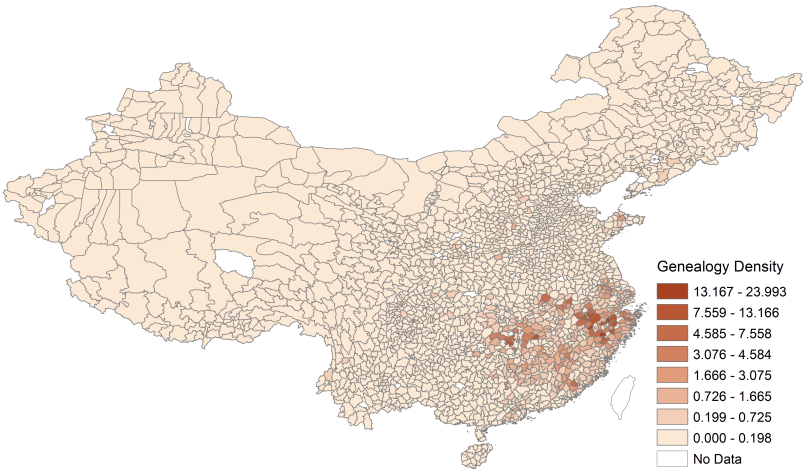
[36] *Zhong Guo Jia Pu Zong Mu*, published in 2008 by Shanghai Library, which has been used in several studies by Chen et al. (2020); Greif and Tabellini (2017); Zhang (2020); Chen et al. (2022).

[37] See `https://jpv1.library.sh.cn/jp/page/about`

[38] The original *The General Catalog of Chinese Genealogy* covers 52401 genealogies and was the largest collection (Zhang, 2020).

[39] Very few genealogies existed prior to 1368, the beginning of Ming Dynasty.

Figure 1: Genealogy Density at the County Level



Notes. This figure shows the geographic distribution of genealogy density as measured by the number of genealogy books normalized by the 1953 population.

Republic of China, and (iv) the Private Economy Research Institute of China. It includes 14 waves (up to 2021) and each wave drew 0.03% to 0.05% of all the private firms in China. The survey used the multi-stage stratified random sampling method to ensure that all of the 31 provinces are well represented (Lu and Tao, 2009). It has been used extensively in papers published in economics, management and finance journals, such as Li et al. (2008), Lu and Tao (2009), Jia (2014), Shen and Su (2017).

CPES in 2010 is perfectly suited for studying private family firms in China because it contains intensive information on family engagement in firm ownership and management, which was not surveyed in other waves. It carefully distinguishes the founder's ownership and his/her family members' ownership of equity shares. Many other datasets mix them together and cannot distinguish firms owned by an individual (without any family engagement) and firms owned by a family. It also provides detailed information on family members' occupations in the firms, as well as financing resources, inheritance decisions and firm performance. Consequently, for our research purpose, we only consider the survey conducted in 2010, which covers around 3000 firms with complete information in key variables.[40]

In the literature, the definition of family firms varies across studies (see Miller et al. (2007) for a summary). We adopt a relatively strict and conservative definition. A private firm is defined as a family firm if the founding family owns at least 50%[41] of equity shares and at least another family member (other than the founder) engages in the firm (either holds positive equity shares or a board member position or the CEO position). Family members include individuals who are related to the founder by blood or marriage, such as

[40] Genealogy density is not correlated with missing values in these key variables, as can be seen in Table A.4.

[41] The threshold could be significantly lower for listed firms (Anderson and Reeb, 2003; Anderson et al., 2012; Ellul et al., 2010).

their parents, spouses, children, and siblings. This definition follows Bøhren et al. (2019) and Amore et al. (2014) to ensure that the controlling family owns the majority of the equity share in private firms. In addition, in the spirit of Claessens et al. (2002), Gómez-Mejía et al. (2007) and Cai et al. (2013), the engagement of family members can enhance the influence of the controlling family in the firm's management, and it conforms to a typical understanding that family firms are characterized as an organization controlled and managed by multiple family members (Bertrand and Schoar, 2006; Miller et al., 2007). Firms controlled by an individual without other family members' engagement are not family firms by our definition, discussing family ties or nepotism in such firms is not meaningful.[42] We adopt alternative definitions of family firms in robustness checks and show that the main results still hold.

Based on the baseline definition of family firms, we generate the dependent variable, the fraction of family firms in each county (the number of family firms surveyed over the number of firms surveyed). In 2010 CPES data, some counties contain a limited number of surveyed firms, which lead to a noisy calculation of family firm share. Consequently, we only consider counties with at least five firms surveyed in our baseline regression (We also change the threshold of the number of firms in our robustness check.). In the end, there are 204 counties in our sample. We find that the mean of genealogy density in the sample counties is not significantly different from that of other counties in China, which partly suggests the representativeness of our sample. We show the geographic distribution of genealogy density and share of family firms for counties with at least five firms surveyed in Figure 2 and detailed summary statistics in the Appendix Table A.3.

3.3. China Family Panel Survey

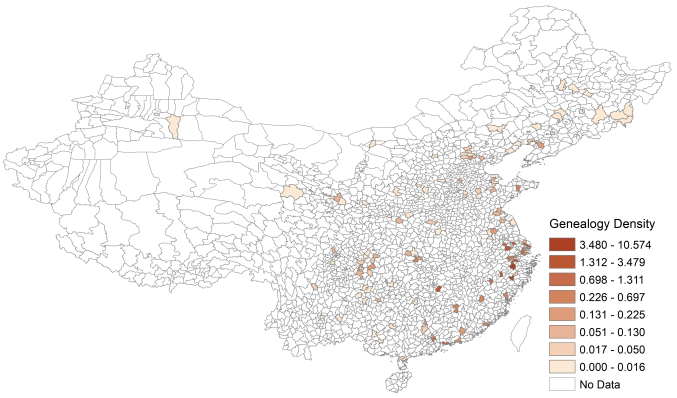
To validate our proxy measure for historical family culture - genealogy density, we use detailed individual level information from the China Family Panel Survey (CFPS). Though genealogy is not a direct measure of historical family culture, we could validate it by demonstrating that it has a large predictive power for people's attitudes toward family. The China Family Panel Survey is a nationally representative longitudinal survey project designed to document changes in Chinese society, economy, population, education, and health etc. The survey was conducted at the individual, household and community/village level. We use the survey wave in 2010, as this is the only wave containing the genealogy information. This nationally representative survey covers 33,600 adults and 14,960 households in 25 provinces,[43] representing 95% of the population in China. Furthermore, to rule out the effect of trust from the validation test, we merge the data on questions related to trust in the national wave of 2012.[44]

[42] Shen and Su (2017) believe all the firms in CPES are family firms because the largest shareholders for all these firms are individuals or families. However, they do not distinguish individual firms and family firms.

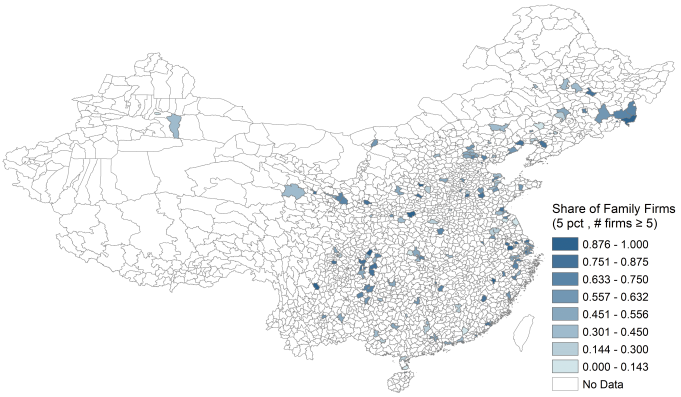
[43] The sample of provinces excludes Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia and Hainan.

[44] There are no questions related to trust in the survey in 2010.

Figure 2: Genealogy Density and Share of Family Firms for Counties with at least 5 Surveyed Firms



(a) Genealogy Density



(b) Share of Family Firms (5% threshold)

Notes. Panel (a) in this figure shows the geographic distribution of genealogy density for counties with at least 5 surveyed firms in CPES, and panel (b) displays the share of family firms for counties with at least 5 surveyed firms in CPES.

The adult and family survey contains extensive questions on family and individual demographics, economic situations, beliefs and attitudes, etc. Specifically, the household level survey in 2010 records whether the household had a genealogy or not. We construct a proxy variable of historical family culture based on this question. The existing literature usually uses subjective questions[45] from the World Value Survey (WVS) to measure the strength of family ties across countries (Bertrand and Schoar, 2006; Alesina and Giuliano, 2010). Another way to measure family ties is to assess the frequency of contact between family members or how close to the parents children live after they leave their parental house

[45]These three questions are: 1. How important family is in their life; 2. Choose which of the following two statements they agree: (1) Regardless of what the qualities and faults of one's parents are, one must always love and respect them, or (2) One does not have the duty to respect and love parents who have not earned it; 3. Choose with which of these two views they most tend to agree: (1) It is the parents' duty to do their best for their children even at the expense of their well-being, or (2) Parents have a life of their own and should not be asked to sacrifice their well-being for the sake of their children.

(Alesina et al., 2015). We use both subjective and objective questions in CFPS closest to these in WVS and estimate their correlations with genealogy. The five questions are (1) How important is happy family, (2) How important is children having achievements, (3) How close are the relationships among family members, (4) Did you visit the graves of dead family members, (5) The number of relatives that visited you during the last spring festival. Furthermore, we create the first component of the principal component of the above five questions to assess the overall family value of an individual. We are not able to directly aggregate the individual responses as the measure of family ties at the county level in our later estimation, because the counties surveyed in CFPS are very different from counties surveyed in CPES, which might end up a with very small sample size in our baseline estimation.

3.4. Other datasets

Our county-level controls include a variety of geographical, historical and socioeconomic variables. Geographical controls are distance to coast, altitude, ruggedness and calorie suitability. We compute the distance from the centroid of a county to the coast in GIS software. The altitude data is from CGIAR SRTM.[46] The ruggedness variable is obtained from Nunn and Puga (2012). Distance to coast, altitude, and ruggedness might affect historical transportation, crucial for business development and interaction with strangers. Convenient transportation reduces the cost of communicating with strangers, which might indirectly lead to weaker historical family culture (Fincher et al., 2008) and scarcity of family firms. We also control for the calorie suitability before 1500 from Galor and Özak (2016) to account for the possibility that overall ancient agriculture potential might shape the cultural traits in China. As discussed in Gowdy and Krall (2016), more advanced production modes in agriculture are correlated with stronger family ties because they require cooperation among family members. We average over raster points in a county for all variables originally available in raster format, such as ruggedness.

For historical variables, we control for biography density, population density in 1880 and tax per capita in 1820. Biography density is the total number of biographies in each county from 1368 to 1949 normalized by the 1953 population. It is used to measure historical human capital since the vast majority of the individuals with a biography during that period are Keju degree holders (selected from civil examinations). Further, controlling for biography density could also address the concern that some genealogies are complied for the purpose of finding connections to renowned kin or ancestors. We obtain the number of biographies from 1368 to 1949 (the same period as our genealogy data) in each county from the China Biographical Database Project (CBDB).[47] We normalize it using the 1953 population. Additionally, we add population density in 1880 and tax per capita in 1820 to measure historical prosperity. The data is from Bai and Jia (2016). However, these two

[46] See https://bigdata.cgiar.org/srtm-90m-digital-elevation-database/
[47] See https://projects.iq.harvard.edu/cbdb/home

variables are at the prefecture level, and they do not cover the entire territory of China. Referring to Altonji and Pierret (2001), we code the missing values as zeros and add another dummy indicating missing values of these two variables [48] to avoid an unnecessary reduction in our sample size. [49]

Our socioeconomic controls are mainly collected from the 2010 Chinese Census conducted by the National Bureau of Statistics of China. It covers the entire population, and provides the most comprehensive demographic information. We collect urbanization, sex ratio and minority rate at the county level since they could be related to natural conditions and cultural factors that might influence our outcomes. We control for a provincial capital dummy. In addition, in light of the possibility that the constraints of entrepreneurs' access to formal institutes could force them to rely on family networks (Robb and Robinson, 2014), we control for bank density (average number of banks established during 1990 to 2010 in each county, normalized by 2000 population). The data is obtained from the China Banking and Insurance Regulatory Commission.[50]

For our IV, we collect information about Confucian academies during the Song dynasty (960AD-1279AD) from Bai (2012):[51] the original data is gathered from extensive gazetteers and historical documents. Finally, for our placebo, we use the number of Buddhist sites in each county in 1820, normalized by the 1953 population. The data on Buddhist sites originates from the National Gazetteer of the Qing Dynasty (1820).[52] We collect it from the website of WorldMap.[53]

3.5. Validation of genealogy measure

To document that genealogy is a plausible proxy of historical family culture in China, we conduct a validation test of genealogy by relating it to the subjective and objective questions capturing the strength of contemporary family ties in CFPS. Specifically, we estimate the following ordinary least squares model:[54]

$$Y_{ij} = \alpha + \beta\, Genealogy\ Density_j + X_{ij}^{'}\gamma + Z_j^{'}\delta + \varepsilon_{ij} \quad (1)$$

where Y_{ij} represents the responses of individual i in county j to five questions re-

[48] The underlying assumption of this method is that the missing values of population density in 1880 (and tax per capita in 1820) are similar. In light of the fact that these missing values only exist outside of the China Proper Area, this is not a strong assumption.

[49] Our sample size will decrease from 204 to 177 if we drop these missing values; however, the estimates remain highly similar.

[50] See http://xkz.cbirc.gov.cn/jr/

[51] 493 Confucian academies during the Song dynasty with clear information.

[52] It contains 2407 Buddhist sites with precise locations.

[53] See http://worldmap.harvard.edu/data/geonode:buddhist_sites_1820

[54] The results are robust when we adopt ordered probit model regressions for the first four questions related family ties and negative binomial regressions for the last question, that is the number of relatives visit them durin the spring festive, which is in the count data format. The detailed results are shown in Table A.5.

lated to family ties in CFPS dataset and the first component of the principal component analysis of these five questions. These questions are described in detail in Section 3.3. $Genealogy\ Density_j$ is historical genealogy density (number of genealogies compiled from 1368 to 1949, normalized by population in 1953) in county j. X_{ij} is a set of individual characteristic variables including age, gender, household registration status, ethnic group, education level and marital status. Z_j denotes county level geographic, historical, and socioeconomic controls. Geographic variables include ruggedness, altitude, distance to the coast and calorie suitability. Historical variables include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Contemporary socioeconomic variables are urbanization rate, sex ratio, minority rate, bank density, and provincial capital dummy. Standard errors are clustered at the county level.

Table 1 displays the results of the validation. In odd columns, we perform unconditional regressions. In even columns, we control for the individual characteristics and community/village fixed effects. We find that genealogy density is significantly positively correlated with all questions relevant to family ties. The coefficients of genealogy density are relatively stable for both magnitude and significance. We can observe that individuals in counties with higher genealogy density attach more importance to the happiness of their families, care more about the children's achievements, have closer family relationships, are more likely to have visited dead family members' grave in the past year, and have more relatives visiting them in the spring festival. Finally, the overall strength of family ties measured by the principal components of previous outcomes is significantly stronger for individuals in counties with higher genealogy density. Regarding the magnitude, one standard deviation increase in genealogy density leads to the principal components increase by 0.046 standard deviation. The results imply that historical genealogy density has strong predictive power for the strength of family ties. Alternatively, we use another measure of genealogy - a dummy variable equal to 1 if the household of individual i has a genealogy book - as the independent variable. We include community or village fixed effects in Specification 1 instead of controlling for county level variables. The results in Table A.6 display a similar positive correlation between holding a genealogy book and contemporary measures of family ties.

Another potential issue for the measure of genealogy is that family ties might pick up the difference in other cultural norms such as trust, clan culture, and collectivist culture across individuals or regions. For instance, the existing studies find that there is a positive correlation between family ties and trust across countries by using the indicators from WVS (Alesina and Giuliano, 2014; Buggle and Durante, 2017), whereas the correlation is negative within half of the countries (Bertrand and Schoar, 2006). As discussed in Section 2.2, genealogy may also reflect clan culture. Similarly, both family culture and collectivist culture stress cooperation among family members or groups. To address this issue, we first test the correlation between genealogy density and measures of other three cultural norms:

Table 1: Validation of Genealogy: CFPS Data

Dependent variable	Happy family important		Child. having achieve. important		Family relation. are close	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.034***	0.032***	0.033***	0.026**	0.058***	0.037**
	(0.009)	(0.009)	(0.010)	(0.011)	(0.015)	(0.017)
Controls	No	Yes	No	Yes	No	Yes
Observations	29514	28608	29514	28608	27073	26255
R^2	0.001	0.046	0.001	0.051	0.003	0.088
Dependent variable	Visit grave last year		# Relatives visit spring festival		Principal component of family ties	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.067***	0.048***	0.089***	0.084***	0.059***	0.046***
	(0.018)	(0.016)	(0.023)	(0.022)	(0.011)	(0.014)
Controls	No	Yes	No	Yes	No	Yes
Observations	29504	28598	29348	28443	29518	28612
R^2	0.004	0.064	0.007	0.087	0.003	0.068

Notes. OLS regressions. All continous variables are standardized. Controls include individual level age, gender, household registration status FEs, ethnic group FEs, education FEs, marriage status FEs, and county level geographic controls, historical controls, and socioeconomic controls. Standard errors clustered at the county level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

trust, clan culture, and collectivist culture. We construct a binary measure of general trust based on the question in the CFPS dataset: do you in general trust or suspect others? Similarly, contemporary clan culture is defined by a question in CFPS dataset: are there any ancestral halls in your village or community? We follow Talhelm et al. (2014) and use rice cultivation to measure collectivist culture. Since rice area is not available at the county level, we use rice suitability to proxy for the collectivist culture. The estimation follows Specification 1. The results are presented in Table A.7. We can observe that the coefficients of genealogy density are close to zero and insignificant for trust and ancestral hall, though there seems to be a positive correlation between genealogy density and rice suitability. Hence we further perform the robustness of the validation test by including these three cultural norms and Confucian academies in the columns of principal components of family ties in Table 1. Again, we find that the positive correlation between genealogy density and family ties is very robust, as shown in Table A.8. The above tests indicate that genealogy density is a plausible measure or proxy variable of the strength of historical family culture in China, and it is less likely to be affected by other cultural norms.

4. Main results

In this section, we present our main empirical findings. We estimate the impact of genealogy density on the share of family firms. To alleviate potential endogeneity concerns, we adopt an instrumental variable approach and a matching approach to examine how historical family culture drives the emergence of family firms.

4.1. Historical family culture and fraction of family firms

In this section, we investigate the relationship between historical family culture and family firm share in different counties. As discussed in Section 3, our measure of historical family culture is county-level genealogy density (number of genealogies compiled from 1368 to 1949, normalized by population in 1953.). We only consider genealogies compiled prior to 1949 to avoid reverse causality. The estimation is based on the following equation:

$$Y_j = \alpha + \beta \, Genealogy\ Density_j + X_j^{'}\gamma + \mu_p + \varepsilon_j \tag{2}$$

where Y_j is the fraction of family firms[55] surveyed in county j; $Genealogy\ Density_j$ is genealogy density in county j; the coefficient of main interest is β, which captures the effect of historical family culture on the prevalence of family firms. The vector $X_j^{'}$ denotes geographic, historical, and socioeconomic controls that could be related to the outcome variable and genealogy density simultaneously. Geographic variables include ruggedness, altitude, distance to the coast and calorie suitability. Historical variables include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Contemporary socioeconomic variables are urbanization rate, sex ratio, minority rate, bank density, and a provincial capital dummy. μ_p are province fixed effects; ε_j is the error term. We use robust standard errors in our baseline estimation.[56]

The results are reported in Table 2. Genealogy density is standardized. Province fixed effects partially account for some geographic and institutional variations, and variations in climate and policies. For instance, the implementation of one-child policy varies at the province level (Ebenstein, 2010). The coefficients of genealogy density remain stable as we gradually add more controls, suggesting that additional socioeconomic variables explain limited variations of the outcome variable. Focusing on column (4), a one standard deviation increase in genealogy density is associated with a 5.2% increase in family firm share with full controls. In light of the fact that the average fraction of family firms under our baseline definition is 49.7%, this magnitude is non-negligible. These private firms were established after 1978. The result confirms our hypothesis that historical family culture still

[55] See a detailed discussion of the definition of family firms in Section 3.2.

[56] Our observations are unweighted, but our results remain robust when we use different weighting schemes, such as the number of firms surveyed in each county, GDP per capita, or local population.

Table 2: Genealogy density and fraction of family firms (OLS)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.053***	0.051***	0.050**	0.052**
	(0.017)	(0.018)	(0.019)	(0.021)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.250	0.254	0.262	0.268
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is the fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is the number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

persistently predicts the distribution of family firms after the socialist transformation and cultural revolution periods (1949-1976), which is consistent with the studies emphasizing the strength of culture (Giuliano, 2007; Fernández and Fogli, 2009; Algan and Cahuc, 2010). Kinship-based cooperation resumed and adapted well to the new environment when China opened up and was determined to develop its private sectors (Peng, 2004). For suggestive evidence, we extend the analysis to the listed firms in Section A.1 in the Appendix.

Kelly (2019) points out that papers on long-term persistence might have inflated t statistics due to severe spatial autocorrelation in residuals. We test this possibility employing Moran's I statistic (based on the same spatial weights matrix as Kelly (2019)). The statistic is only -0.022 (Z score is -0.429 and p-value is 0.334) when using residuals of the first regression in Table 2,[57] which suggests very low spatial autocorrelation in our regression residuals.

4.2. Instrumental variable results

To alleviate the concern about the potential endogeneity, we conduct empirical analysis using an instrumental variable, the influence of Confucian academies during the Song dynasty (960AD-1279AD), which occurred prior to our sampling period of genealogies. OLS estimation may be biased due to omitted variables. For instance, historical economic development or institutions could potentially promote the genealogy compilation and contemporary family firm development. The relevance of the instrument is based on the fact that during the Song dynasty, Confucian scholars greatly promoted the establishment of a new type of fam-

[57]The results are similar when we use residuals of other regressions in Table 2.

ily clan system (Feng and Chang, 2008; Biao, 2007), and one channel is giving lectures at their academies, as discussed in Section 2. These Confucian scholars were active around one thousand years ago, making it difficult to measure their promotion of family culture using the copies of their books or the presence of ancestral halls, since very few are preserved. However, it is plausible that their ideologies and values regarding family would be spread through academies, where they taught their disciples. For example, Zhu Xi had been directly involved in teaching activities for more than 40 years. According to Ji (1996), he was passionate about constructing Confucian academies and giving lectures every time he was appointed to a new place as a government official. He gave lectures in more than 10 academies dispersed in at least 7 counties,[58] many of which were constructed under his direction. Scholars such as Zhang Zai, Cheng Yi, Fan Zhongyan, Lv Zuqian, and Lu Jiuyuan are also recorded in Ji (1996) for their involvement in teaching or constructing academies. Furthermore, Confucian academies during that period also interacted with each other by inviting scholars to give lectures or debates and mirrored each other in terms of the contents and ideologies they delivered. As a result, to create our instrumental variable, it is reasonable to consider the influence of all the Confucian academies during the Song dynasty, which spread the idea of constructing the family clan system and promoted family values.[59]

The instrumental variable is the influence of Song academies within 200 km of each county, taking into account both the number and the distance of academies to counties. The formula of the instrument is displayed in Equation 3, where n is the number of Confucian academies within 200 km, d_{ij} is the distance of academy i to the centroid of the county j. $(200 - d_{ij})$ is interpreted as the influence of academy i, assuming that a county would be influenced more by closer Confucian academies. The threshold distance of 200 km is set based on the fact that the vast majority of disciples of Zhu Xi, one of the most important Confucian scholars in the Song Dynasty, were born in a county within 200 km of the nearest academies of Zhu Xi, as shown in Figure A.1. We use different thresholds of distance as our robustness checks, as reported in Table A.33 and Table A.34. Furthermore, to address the concern that our results could be driven by the particular functional form, we also employ the count of Confucian academies within 200 km as an additional robustness check in Table A.35.

[58] Information summarized from Ji (1996) p.393-394, it can also be verified in Bai (2012) p.29

[59] Chen et al. (2022) share a similar spirit and use distance to only the Zhu Xi academies as the instrumental variable for genealogy density. We differ by considering all the Confucian academies during the Song dynasty.

Table 3: Genealogy density and fraction of family firms (IV)

Second stage results				
Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.230**	0.238**	0.222**	0.237**
	(0.114)	(0.117)	(0.113)	(0.115)
Mean dependent vars.	0.497	0.497	0.497	0.497
First stage results				
Dependent variable	Genealogy density			
	(1)	(2)	(3)	(4)
Academy(within 200 km)	0.374***	0.391***	0.414***	0.417***
	(0.106)	(0.117)	(0.119)	(0.123)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
KP F-statistic	12.56	11.20	12.16	11.55
R^2	0.416	0.426	0.432	0.449
Mean dependent vars.	0.460	0.460	0.460	0.460

Notes. County-level IV regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. We employ the influence of Confucian academies within 200 km during Song dynasty as the IV for genealogy density. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

$$Academy_j = \sum_{i=1}^{n}(200 - d_{ij}) \tag{3}$$

The IV estimation results are presented in Table 3. Column (1) controls for the province fixed effects only. It suggests that the instrument is relevant to the proxy of historical family culture and is strongly positively correlated with genealogy density in the first stage. Columns (2) to (4) gradually include an extensive set of control variables in the baseline estimation, and the coefficients of academies remain stable and sizable: a one standard deviation increase in the influence of Confucian academies leads to 0.417 higher genealogy density as displayed in column (4). The second stage results document that historical family

culture instrumented by Song academies is positively correlated with the fraction of family firms in counties. The results are robust across all specifications in terms of significance and magnitude of coefficients.

The magnitude of the effects of historical family culture is larger compared to our OLS estimations. The first explanation for the inflated IV coefficient is the discrepancy between the Local Average Treatment Effect (LATE) estimated by IV and the Average Treatment Effect (ATE) estimated by OLS. There are heterogeneous treatment effects across counties with different characteristics, as shown in Table 6. If the compliers-counties with higher genealogy density due to their proximity to Song academies-are associated with characteristics that strengthen the effects of genealogy density, this association could result in a larger IV coefficient compared to the OLS coefficient. The second explanation for the inflated IV coefficient could be because our instrument is not particularly strong, as the KP-F statistics of our instrument are only slightly above 10. This could result in a biased IV estimator in finite samples.

To further investigate the reasons behind the larger IV coefficients compared to the OLS coefficients, we employ a new method proposed by Ishimaru (2024) to decompose the IV-OLS gap. Specifically, this gap will be decomposed into three components: (1) covariate weight difference: this refers to the difference in how the IV and OLS coefficients assign weights to the covariates, including fixed effects; (2) treatment-level weight difference: this pertains to the difference in how the IV and OLS methods place weights on different treatment levels; (3) marginal effect difference: this represents the difference between the IV- and OLS-identified marginal effects, which typically originates from endogeneity bias.

Table A.10 presents the decomposition of the IV-OLS gap from our baseline estimation, which includes a full set of controls. Columns (1) to (3) report the OLS estimate, the IV estimate, and their gap, respectively. Columns (4) to (6) detail the contributions to the gap from the covariate weight difference, the treatment-level weight difference, and the marginal effect difference, respectively. The sum of columns (4) through (6) equals the total IV-OLS gap shown in column (3). As we observe, the IV-OLS gap is primarily due to the covariate weight difference. This suggests that the IV estimation assigns a different importance to covariates (including province fixed effects) compared to OLS, reflecting their relationship with the instrument rather than solely with genealogy density and the fraction of family firms. The treatment-level weight difference and endogeneity bias contribute minimally to the IV-OLS gap.

There are two potential threats to the identification. First, Song academies may have a long term influence on contemporary economic development, which are relevant to the share of family firms. To exclude this possibility, we regress contemporary socioeconomic outcomes (including urbanization rate, average education, sex ratio and bank density) on the instrumental variable under the baseline Specification 2. As reported in Table A.11, the effects of Song academies on other contemporary outcome variables are very small

and not significant, which aligns with the findings in Chen et al. (2022). The results suggest that Song academies are unlikely to affect the contemporary fraction of family firms through channels other than historical family culture. Cultural aspects endure far longer than economic prosperity. Over the course of a thousand years, the logic of economic development has undergone significant transformation. China has transitioned from an agrarian economy to a major industrial one, with economic hubs shifting to coastal cities. Despite these changes, cultural traits and kinship networks continue to be deeply ingrained in the collective consciousness.

Second, this instrument might reflect the overall Confucianism during the Song dynasty instead of family culture only. For instance, Confucianism contains other aspects of culture and norms such as humaneness, rite and centring, and loyalty to the emperor. Bol (2008) thinks that most Confucians and Neo-Confucians shared concerns about the government, family, economy, culture, and nature that followed from them. To rule out the possibility that other aspects of Confucianism could be directly related to the prevalence of family firms, we estimate the impact of contemporary Confucianism on the fraction of family firms. We measure the overall contemporary Confucianism based on the distribution of 372 modern existing Confucian temples[60] with the same method as Equation 3, since they are not directly used by famous scholars to promote family culture. As shown in Table A.12, Song academies still strongly predict the distribution of Confucian temples in 2010. However, contemporary Confucian temples are not related to the prevalence of family firms anymore. It indicates the important roles played by those scholars who facilitated family clan culture in Song academies.

4.3. A matching approach

To address the potential endogeneity, we also adopt the method of matching neighbor counties, which was widely used in the literature, such as Banerjee and Iyer (2005), Huillery (2009), Voigtländer and Voth (2012), Deryugina et al. (2020). Though our OLS analysis controls for precise and demanding variables, it remains plausible that not all factors that correlate with both family firm share and genealogy density are included. To overcome the omitted variable problem, we match counties by geographic location and directly compare the outcomes of counties close to each other. This strategy exploits the spatial discontinuity of genealogy densities across counties. The underlying assumption of the method is that geographic neighbors share similar unobservable characteristics. Hence the difference in the outcome variable is unlikely to be driven by the difference in the omitted variables. For instance, if the natural conditions in one county are especially suitable for crops which involve more family cooperation, it is reasonable to assume that its nearby county is also suitable. If government capacity is weak in one county because of the far distance from the national and provincial capitals, then its neighboring counties should have a similar

[60]Data from Fan (2004) and Kong and Kong (2011), manually verified in Google and Baidu.

Table 4: Genealogy density and fraction of family firms (Matching)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.062***	0.077***	0.063**	0.057**
	(0.017)	(0.013)	(0.024)	(0.028)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Pair FEs	Yes	Yes	Yes	Yes
Observations	120	120	120	120
R^2	0.582	0.622	0.649	0.663
Mean dependent vars.	0.462	0.462	0.462	0.462

Notes. County-level matching. For each county, we find its mutually nearest county and match them to be a pair (as a result, some observations would be dropped). Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects and pair (pairs of mutually nearest counties) fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

issue. Therefore, directly comparing the outcome of matched counties could improve the estimation since it eliminates these shared unobservable factors.[61]

There are good reasons to think that neighboring counties are similar in the context of China. The counties were usually formed around rivers, and county capitals were generally located alongside rivers. Second, county boundaries exist before the emergence of genealogy, and most are natural borders such as rivers and mountains that display strong historical persistence to the present day.[62] Third, the county areas are relatively small, and the mean area is 2675 square kilometers in our sample, confirming the similar geographic and social conditions among neighboring counties. County borders are thus somewhat arbitrary. This leads us to assume that neighboring counties share similar unobserved characteristics. To perform our matching estimation, we have the following equation:

$$Y_j = \alpha + \beta \, Genealogy\ Density_j + X_j^{'}\gamma + \mu_p + \theta_n + \varepsilon_j \tag{4}$$

The only difference between Equation 2 and Equation 4 is that Equation 4 has a neighborhood (or pair) fixed effect θ_n. Unlike Huillery (2009) where counties are connected

[61] The argument is also demonstrated in Card and Krueger (2015).

[62] The borders of counties may be adjusted due to dynasty change or population change over time, but are not likely to be due to family culture.

together, some of our counties are dispersed and do not share borders with any of the other 203 counties in our sample. Consequently, our priority is to match mutually nearest counties, either connected or not, to obtain disjoint pairs. We assume counties are similar within each pair but different across pairs. Therefore, pairs must be disjoint; otherwise, by transmission, we are assuming many counties are similar. Since two paired counties are mutually nearest to each other, some counties are dropped because their closest county finds another county closer. In total, we have 60 pairs of mutually nearest counties. Among these 60 pairs of counties which can be reasonably assumed as neighbors, the average distance is 33.7 km and the maximum distance is 118.7 km.[63]

The results of the matching estimation are displayed in Table 4. There are statistically positive effects of genealogy density on the share of family firms across specifications. The results are robust after we include the geographic controls in column (2), historical controls in column (3) and socioeconomic controls in column (4). The matching estimates confirm the results in the OLS estimation and have similar magnitudes of effects of genealogy density, even though the number of observations is reduced by around 40%. Specifically, a one standard deviation increase in genealogy density is associated with a 5.7 percentage points increase in the fraction of family firms. The results strongly suggest that our findings are unlikely to be driven by the unobserved or omitted characteristics at the county level.[64]

Additionally, we match each county's nearest neighbor to form a pair (not necessarily mutually nearest) for another robustness check. No counties will be dropped, but a county could appear several times if it is closest to several counties. However, the results are similar to our baseline matching method.

4.4. Genealogy density for three historical periods

Do genealogy density variables for different historical periods show different levels of influence on the fraction of family firms? Older genealogies may have a stronger influence due to a deeper root of family culture, but the opposite could be true if family culture does not persist over time. To explore this question, we calculate the genealogy density for three historical periods: the Republic of China (1911-1949), the Qing Dynasty (1644-1911), and the Ming Dynasty (1368-1644). We standardised all three genealogy density variables to ensure comparability across different historical periods.

Table 5 displays the results of our analysis. The first two columns show the effect of genealogy density during the Republic of China (1911-1949) as the independent variable. Columns (3) and (4) display the effect of genealogy density during the Qing Dynasty (1644-1911), while columns (5) and (6) show the effect of genealogy density during the Ming

[63] Our results are robust when we only consider pairs with a distance less than 60 km, 80 km, or 100km.

[64] Fan et al. (2022) show that neighboring counties along the Qinling-Huaihe Line (on the rice-wheat border) are different with respect to the collectivist culture, which might impact the family control over firms. For robustness, in Section A.7.10 we find that our results still hold after excluding pairs near or across the Qinling-Huaihe Line as shown in Table A.32.

Table 5: Genealogy density for three historical periods

Dependent variable	Fraction of family firms							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Genealogy density (1911-1949)	0.019**	0.019**					-0.023	-0.022
	(0.007)	(0.008)					(0.015)	(0.017)
Genealogy density (1644-1911)			0.021**	0.020*			0.029***	0.026**
			(0.009)	(0.010)			(0.011)	(0.013)
Genealogy density (1368-1644)					0.025***	0.027***	0.033***	0.036***
					(0.008)	(0.009)	(0.010)	(0.012)
Geographic controls	No	Yes	No	Yes	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204	204	204	204
R^2	0.238	0.259	0.241	0.260	0.241	0.263	0.251	0.270
Mean dependent vars.	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. We calculate genealogy density variables for three historical periods: the Republic of China (1911-1949), the Qing Dynasty (1644-1911), and the Ming Dynasty (1368-1644). Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Dynasty (1368-1644). All three genealogy density variables exhibit a significant positive effect on the prevalence of family firms, with the coefficients of genealogy density during the Ming Dynasty slightly larger than others. In columns (7) and (8), we conduct a horserace regression with all three genealogy density variables for the three historical periods. This analysis reveals that older genealogies display a stronger effect on the fraction of family firms. Regions with a longer history of compiling genealogies have more time to establish and reinforce their cultural identity, thus leading to a more robust family culture.

4.5. Heterogeneous effects

4.5.1. Heterogeneous effects by human capital accumulation

The effects of family culture may vary in regions with different levels of human capital accumulation. In regions with lower levels of human capital accumulation, family networks may help compensate for the lack of talent, and serve as an important source of business support for entrepreneurs, which may foster the creation and survival of family firms. However, in regions with higher levels of human capital accumulation, the formal labor market may be more attractive for firm recruitment, which could weaken the role of family networks.

We measure human capital accumulation using the average level of education – years of schooling in 2010 – in each county. We group all counties into two categories based on whether their average education level in 2010 was below or above the median, and conduct separate regressions for each group. Panel A of Table 6 displays the results, with the first

two columns showing counties with below-median average education levels and the last two columns showing counties with above-median levels.

The coefficients in column (2) indicate that family culture strongly predicts the fraction of family firms in regions with lower human capital accumulation. In contrast, the coefficients in column (4) suggest that family culture has a much weaker effect in regions with higher human capital accumulation, where the formal labor market is more attractive for firm recruitment. These findings indicate that family networks may be particularly valuable in areas with limited access to talented workforce.

Table 6: Genealogy density and fraction of family firms: heterogeneity

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Panel A: human capital	Education below median		Education above median	
Genealogy density	0.058**	0.075***	0.027	0.008
	(0.022)	(0.024)	(0.058)	(0.071)
Observations	103	103	101	101
R^2	0.319	0.439	0.504	0.558
Panel B: corruption	Less corruption		More corruption	
Genealogy density	0.037***	0.034**	0.290***	0.278**
	(0.011)	(0.014)	(0.106)	(0.133)
Observations	104	104	100	100
R^2	0.217	0.267	0.316	0.426
Panel C: public service	Less public service		More public service	
Genealogy density	0.404***	0.483***	0.042***	0.044**
	(0.133)	(0.137)	(0.012)	(0.017)
Observations	101	101	103	103
R^2	0.380	0.509	0.170	0.262
Panel D: marketization	Lower marketization		Higher marketization	
Genealogy density	0.018	0.034	0.167***	0.239***
	(0.014)	(0.034)	(0.046)	(0.063)
Observations	101	101	103	103
R^2	0.316	0.392	0.358	0.485
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes

Notes. County-level OLS regressions. The heterogeneous results are analyzed separately by average educational level, corruption, public service, and marketization. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

4.5.2. Heterogeneous effects by the quality of public administration

We conduct heterogeneous analysis based on the quality of public administration. We use two variables to measure this quality. The first variable is the average corruption level from 1998 to 2010, measured by the number of corruption cases per 10,000 government officials at the provincial level. The second variable is the public service expenditure at the provincial level in 2010. The results are displayed in Panel B and Panel C of Table 6, respectively.

In the first two columns of Panel B of Table 6, we present the results for regions where the corruption level is below the median. Conversely, the results for regions where the corruption level exceeds the median are displayed in the last two columns of the same panel. As observed, in regions with higher corruption levels, family culture exerts a more pronounced influence on the prevalence of family firms. While significant coefficients are noted in both column (2) and column (4), the coefficient in column (4) is statistically larger than that in column (2) (confirmed by a t-test).

Similarly, we present the results for regions with public service expenditure below the median in the first two columns of Panel C of Table 6, and the results for regions with public service expenditure above the median in the last two columns of the same panel. The findings indicate that genealogy density predicts the fraction of family firms across different regions; however, the effect is stronger in regions with lower public service expenditure. The coefficient in column (2) is statistically larger than that in column (4) (confirmed by a t-test). The results in Panel B and Panel C of Table 6 confirm that in regions with lower quality of public administration, family culture is more effective in protecting and fostering the prevalence of family firms. These results provide evidence that family culture can serve as a substitute for well-functioning local political institutions and protect entrepreneurs from political abuse by local officials.

4.5.3. Heterogeneous effects by marketization

We conduct heterogeneous analysis by marketization using the marketization index, which is a continuous variable measuring the relative marketization process in each prefecture in 2010. It is composed of five indexed, each of which reflects a specific aspect of marketization. They are the relationship between the government and the market, the development of the non-state-owned economy, the development of the product market, the development of the factor market, and the development of the market intermediary organization and legal system environment. In the first two columns of Panel D of Table 6, we show how genealogy density impacts the fraction of family firms when the marketization index is below the median. Columns (3) and (4) of Panel D of Table 6 illustrate the impact of genealogy density when the marketization index is above the median. The coefficient in column (4) is significantly larger than the coefficient in column (2) (confirmed by a t-test).

This suggests that the role of genealogy density is stronger in regions with a higher

marketization index. This might be because, in regions with less government intervention and a more developed non-state economy, family networks complement market forces, driving the establishment of more family firms. The establishment of family firms relies not only on family culture but also on a healthy market where many production factors are well supplied, and government intervention is minimal. Without these conditions, family culture alone does not have sufficient influence on the establishment of firms.

4.5.4. Heterogeneous effects by financial development

We conduct heterogeneous analysis based on financial development, measured by bank density. However, we do not find strong evidence of heterogeneous effects across regions with varying levels of financial development. First, this may be attributed to the strong correlation between financial development and family culture, as demonstrated by Chen et al. (2022).[65] Regions with different levels of financial development have varying genealogy densities, and the marginal effect of increasing genealogy density may not be constant. Second, our data on bank density may measure formal financial development, while informal financial development is difficult to capture. In addition to relying on banks or family networks for finance, some entrepreneurs might also utilize other informal markets. This could influence the results of our heterogeneous analysis.

5. Mechanisms

In this section, we investigate the channels through which historical family culture might impact the prevalence of family firms. First, family firms' common characteristics are family members' ownership of the firm and their participation in firm management (Bertrand and Schoar, 2006; Miller et al., 2007). As pointed out by Cai et al. (2013), job assignments in firms are affected by family ties with firm heads. Historical family culture could enhance family members' engagement in the firm, leading firms to evolve to be, or persist as, family firms. Second, the source of financing at the initial stage could be correlated with firm types and their later development (Robb and Robinson, 2014). More family support of funds might alter the startup's capital structure and increase its likelihood of becoming a family firm. We test whether firms in regions with a stronger family culture are more inclined to raise capital from their family members instead of formal financial institutes, such as banks. Third, family succession is an important issue for some of the firms and has been discussed widely by scholars (Ayres, 1990; Burkart et al., 2003; Shen and Su, 2017). In some literature, inherited control is a feature of (or closely related to) family business (Pérez-González, 2006; Bennedsen et al., 2007). It is highly possible that entrepreneurs attaching great importance to family ties tend to pass the firms on to the next generation.

[65] The table for the heterogeneous analysis by bank density is available upon request.

5.1. Family engagement

The first potential mechanism is that historical family culture could facilitate family engagement in firm ownership and management, resulting in stronger family control over the firm. We examine this hypothesis at the firm level (including all the firms in 2010 CPES), taking advantage of the detailed information on family members' participation in the firm. The estimation is based on the following equation:

$$Y_{ij} = \alpha + \beta \, Genealogy\ Density_j + X_{ij}^{'}\gamma + \mu_p + \varepsilon_{ij} \tag{5}$$

where Y_{ij} are family engagement variables in firm i county j, including the fraction of equity shares held by family members, the fraction of board members that are family members, whether the CEO is a family member and the fraction of managers who are family members. $Genealogy\ Density_j$ is the number of genealogies from 1368 to 1949 per 10,000 local population in county j in 1953; the vector $X_{ij}^{'}$ denotes control variables at the individual level (founder's characteristics), firm level and county level. In particular, we add the education fixed effects, founder's age and gender as individual controls. Firm-level controls include log sales measuring the size of the firm, firm age, a dummy indicating whether the firm was privatized from a state-owned enterprise and industry fixed effects. County-level controls are the same as our baseline specification in Equation 2. μ_p are province fixed effects; ε_{ij} is the error term. Standard errors are clustered at the county level.

The results are presented in Table 7. We report estimates without historical and socioeconomic controls in odd columns, and with those controls in even columns for each dependent variable. In columns (5) and (6), we adopt Probit estimation since the outcome is an indicator variable. We find that historical family culture is significantly and positively correlated with all four measures of family member engagement in firms. The magnitudes of the coefficients are non-negligible compared to the mean value of the dependent variables. For instance, in column (2), a one standard deviation increase in genealogy density leads to a 2.1 percentage points increase in family ownership of equity shares, given that the mean value is only 16.8% (12.5% increase in equity share ownership compared with the mean value). Also, the fraction of managers who are family members increases by 1.7 percentage points with a one standard deviation increase in genealogy density, as displayed in column (8). Overall, the results confirm our hypothesis that historical family culture promotes family engagement in the firms.

It could be further investigated why family members are more likely to be involved in owning and managing the firms. This could be traced back to the time when the founder raised initial capital.

5.2. Initial capital

The second channel we test is whether founders emphasizing family values are more likely to raise the initial capital from their family members rather than from formal institutions when establishing the enterprise. Family culture could serve as a substitute for formal institutions due to higher trust, reciprocity, obligation or altruism among family members. Hence, it is easier for founders to borrow from their family members than from banks which usually require collateral and are not always accessible in certain regions. More initial capital from family members ensures stronger family control over the firm and a potentially higher level of family engagement. The CPES data contains questions regarding the source of the initial capital of the firm. Specifically, they ask the entrepreneurs whether their initial capital is from several sources (multiple choices), including family members and bank loans. We test our hypothesis using Equation 5 but replace the dependent variables with the two dummies indicating whether the source of initial capital is from family members or banks. Since our outcome variable is for the year of establishment, we control for the log of initial assets and founding year fixed effects to account for the firm's size and year-specific variations when they were established.

Table 7: Mechanism: family engagement in the firm

Dependent variable	Family engagement in the firm							
	Equity Share		Board Members		CEO		Managers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Genealogy density	0.018***	0.021***	0.028***	0.026***	0.012***	0.010***	0.019**	0.017*
	(0.004)	(0.004)	(0.009)	(0.009)	(0.003)	(0.003)	(0.009)	(0.010)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2874	2874	1655	1655	3207	3207	2517	2517
R^2/Pesudo R^2	0.045	0.050	0.113	0.115	0.059	0.070	0.160	0.163
Mean dependent vars.	0.168	0.168	0.376	0.376	0.073	0.073	0.197	0.197

Notes. Firm-level regressions, probit (marginal effects reported) in column (5) and (6) and OLS in other columns. We employ four variables to measure family engagement: the fraction of equity share held by family members, the fraction of board members who are family members, a dummy indicating that the CEO is a family member and the fraction of managers who are family members. Genealogy density is the number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log sales, firm age, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

The results are reported in Table 8. In the columns of an odd number, we exclude historical and socioeconomic controls, and in the even-number columns, we have full controls. The estimates in column (2) show that the probability of raising initial capital from family members is 1.7% higher when historical culture, as measured by genealogy density, is increased by one standard deviation. In contrast, the probability of raising capital from banks decreases by 2.3% as displayed in column (4). The results are robust to excluding historical and socioeconomic controls. In sum, higher historical genealogy density increases the likelihood of raising initial capital from family members and reduces the possibility of raising capital from banks. It suggests that family members are more likely to offer financial support to founders at the beginning of the entrepreneurship in counties with a stronger historical family culture, thus possibly becoming shareholders or participating in the firm management later.

To examine the possibility that entrepreneurs relying more on family for initial capital is due to fewer banks in regions with high genealogy density (Chen et al., 2022), we perform a mediation analysis to explore whether bank density serves as a channel in Table A.9. We can see that bank density is less likely to work as a channel as its coefficients are small and insignificant in columns (3) and (4), and these of genealogy density are stable in both cases with or without bank density as a control. This implies that family culture may not necessarily suppress the total demand for finance but affect the structure of demand for finance. Overall, the above analysis suggests that historical family culture could serve as a substitute for formal institutions in financing early-stage firms and possibly an alternative channel to create firms in the absence of a modern financial system.

5.3. Inheritance

Another potential channel for the main effects relates to firm succession. Family culture embodies more incentives to build a family legacy. Thus, in regions with a stronger historical family culture, both founders and their children probably have stronger intentions to keep the firms under family control over generations. CPES asked entrepreneurs whether they planned to pass on the firms to their children and asked their children whether they would like to inherit firms from parents. We can construct two binary variables based on their answer (1 indicates Yes and 0 indicates No). For comparison purpose, we restrict firms to ones that have answers to both questions.[66] In our estimation, we employ Equation 5 but replace the dependent variables with two indicator variables on bequest and inheritance and perform Probit regressions.

As can be seen in Table 9, historical family culture is positively related to the willingness to inherit by children and the tendency to pass on the firms to offspring by founders. Specifically, in column (2) with full controls, a one standard deviation increase in genealogy density is associated with a 2% increase in the willingness to inherit firms from parents.

[66]The results still hold when we include all firms that have responses to any of these two questions.

Table 8: Mechanism: the initial capital (Probit, marginal effects)

Dependent variable	Whether the initial capital is from:			
	Family members		Bank loan	
	(1)	(2)	(3)	(4)
Genealogy density	0.013*	0.017***	-0.015*	-0.023***
	(0.007)	(0.006)	(0.008)	(0.008)
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	3009	3009	3036	3036
Pseudo R^2	0.093	0.102	0.087	0.109
Mean dependent vars.	0.084	0.084	0.239	0.239

Notes. Firm-level probit regressions (marginal effects reported). The first dependent variable is a dummy indicating the initial capital was raised from family members. The second dependent variable is a dummy indicating the initial capital was raised from the bank. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log initial assets, founding year fixed effects, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2000, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

Column (4) suggests that founders' tendency to pass on the firms to their children increases by 3% when genealogy density increases by one standard deviation. The results confirm our hypothesis that people in regions with stronger historical family firms have a higher desire to build a family legacy by keeping the firm under family control over generations.

6. Robustness checks

We conduct extensive examinations to address the concerns of potential confounding factors and other channels and check the robustness of the results. Details are provided in Section A.7. We start by documenting that catastrophic political movements such as the Cultural Revolution, which might prevent the genealogy compilation, do not correlate with the genealogy density, and the main results still hold with controlling for it. In addition, genealogy density after 1976 (after the Cultural Revolution) displays a similar positive correlation with the family firm distribution. By controlling for Jinshi (presented scholar) density, Juren (recommended man) density, contemporary Confucian temples or Confucian academies during the Song Dynasty, we show that traditional civil examination (Keju)[67] or,

[67] Bai and Jia (2016) and Chen et al. (2020) discussed Keju exams in detail.

Table 9: Mechanism: inheritance (Probit, marginal effects)

Dependent variable	Willing to inherit		Plan to pass on	
	(1)	(2)	(3)	(4)
Genealogy density	0.020***	0.020**	0.027*	0.031*
	(0.008)	(0.009)	(0.016)	(0.016)
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	1109	1109	1109	1109
Pseudo R^2	0.163	0.166	0.158	0.161
Mean dependent vars.	0.321	0.321	0.405	0.405

Notes. Firm-level probit regressions (marginal effects reported). "Willing to inherit" is a dummy indicating the entrepreneur's offspring are willing to inherit the firm. "Plan to pass on" is a dummy indicating the entrepreneur plans to pass on the firm to his/her offspring. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log sales, firm age, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

more broadly, Confucianism does not bias our results. To further eliminate the possibility that genealogy may affect the family firm prevalence through contemporary economic development or average level of education, we control for GDP per capita, average nighttime light, average years of schooling, and illiteracy rate, respectively. And we obtain similar results. The results also are robust to controlling for rice and wheat suitability which may affect the family culture. To mitigate the concern of migration, we demonstrate that the results are robust with the inclusion of migration rate and immigration rate as controls, or excluding counties in provincial capitals and sub-provincial cities which normally attract migrants. We also show that marketization and government-business relations, which might correlate with genealogy density and affect the fraction of family firms, do not bias our estimation.

Furthermore, we adopt alternative definitions of family firms, alternative samples, and alternative methods to check the robustness. In addition, we employ various methods to construct our instrument and use different approaches to match counties. Finally, we conduct a placebo test by looking at the effects of the historical presence of Buddhist temples and find no influence on contemporary family firms, implying that unobserved economic and institutional factors driving the compilation of genealogy books and construction of historical Buddhist temples do not bias our results.

7. Historical family culture and firm performance

In this section, we demonstrate the private firm performance implications of historical family culture in China. First, we conduct a firm level analysis of private firms to show that historical family culture leads to better firm performance. Second, we provide an explanation for this result by documenting that firms in regions with a stronger family culture display a lower leverage ratio, which may improve firm performance.

To explore the economic implications of family culture, we extend the analysis to the aggregate impact of family culture on firm performance instead of distinguishing between family and non-family firms like existing studies. The empirical framework we adopt is the same as the Specification 5. In particular, as a measure of performance, we use return on asset (ROA), that is, the ratio of the net income to the total asset of each firm in 2010.

Columns (1) and (2) in Table 10 present the results for the genealogy density for different specifications of the empirical model without controlling for historical and social economics variables, and with the full set of control variables. There is consistent evidence of genealogy density's positive impact on firm performance. Notably, a one standard deviation increase in genealogy density is associated with an increase of ROA of about 8.7 percentage points.

Table 10: Firm performance and the underlying mechanism

Dependent variable	ROA		Leverage		ROA	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.073***	0.087***	-0.075**	-0.079***	0.062***	0.075***
	(0.018)	(0.020)	(0.031)	(0.020)	(0.015)	(0.017)
Leverage					-0.210***	-0.207***
					(0.038)	(0.035)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2227	2227	2769	2769	2227	2227
R^2	0.039	0.046	0.183	0.194	0.060	0.066
Mean dependent vars.	0.297	0.297	20.876	20.876	0.297	0.297

Notes. Firm-level OLS regressions. We use ROA to measure firm performance. In column (1) and (2), we do not control for leverage. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log sales, firm age, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

How does historical family culture affect the firm performance? Schmid (2013) finds

that family firms rely less heavily on debt compared to non-family firms in Germany.[68] One mechanism we exploit here is that firms in regions with a stronger family culture accumulate less debt, which may be beneficial for the firm performance. Family culture also indicates a will to build a family legacy, leading families to be more risk averse. Thus firms have fewer incentives to take risky strategies such as debt financing, which may limit their performance (Lang et al., 1996; Opler and Titman, 1994; Aivazian et al., 2005).[69] In addition, as discussed in Section 5.2, firms in regions with a stronger family culture raise a higher share of initial capital from family members and a lower share from banks. We would, therefore, expect that family culture leads to a lower level of debt. We first examine this conjecture by testing the correlation between the debt level and historical family culture. In the following analysis, we adopt the same specification as Equation 5 and use *Leverage*, which is the ratio of debt to asset as the main dependent variable to measure a firm's debt level. We regress *Leverage* on genealogy density separately with and without controlling for historical and socioeconomic variables separately.

The results are displayed in columns (3) and (4) in Table 10. We can find a robustly negative impact of genealogy density on *Leverage* across specifications. That is, firms in regions with a stronger family culture have a lower level of debt. In terms of size, a one standard deviation increase in genealogy density is associated with about 0.079 fewer percentage points in the leverage ratio. To further assess whether leverage is indeed the channel through which historical family culture predicts the firm performance, we conduct a mediation analysis by adding *Leverage* into the estimations in columns (1) and (2) in Table 10. The results are exhibited in columns (5) and (6) in the same table. It is shown that *Leverage* is negatively associated with the performance in all specifications. The size of the coefficient of genealogy density is still positive and drops by around 16%, suggesting that leverage partly accounts for the effect of historical family culture on performance. However, the positive relationship could also be a consequence of reduced conflicts between principals and agents. Family managers could devote more to the firm than professional managers because their interests align more with firm heads (Cai et al., 2013).

We further examine the heterogeneity of the above analysis among family and non-family firms as shown in Table A.36. The results in columns (1) and (2) exhibit that family culture affects the performance of both family and non-family firms. We propose several suggestive explanations for the findings as a systematic and rigorous examination of their relationships is beyond the scope of this paper. One explanation is that genealogy density may improve the overall institutional environment with better contract implementation and fewer business disputes (Yu et al., 2023), which benefits both types of firms. Another potential explanation is the spillover of knowledge and technology from family firms to non-

[68] It also depends on the level of creditor monitoring.

[69] Though existing literature reveals both positive and negative relationship between indebtedness and firm performance, firms in our sample are all unlisted private firms and are relatively young and small. This is more consistent with the scenario in which the negative correlation appears.

family firms along vertical supply chains. The results could also be due to the exit of less productive non-family firms in regions with a larger genealogy density (Aga and Francis, 2017). It is also worth noting that the genealogy density only exerts pronounced relations with the leverage ratio for family firms and accounts for around 18% of the impact of genealogy density on their performance. This is consistent with the previous main findings and implies that family culture may, at least partly, affect aggregate performance via family firms.

8. Conclusion

Family firms play an important role in the economy, but little is known about the determinants of the prevalence of family firms. This paper provides systematically plausible evidence that historical family culture affects the spatial distribution of modern family firms. To identify these effects, we exploit the fact that historical genealogy in China greatly reflects people's beliefs and behaviors related to the family. Drawing on detailed data on private firms, we show that there is a large fraction of family firms in counties with a stronger historical family culture. The instrumental variable approach and nearest matching method improve the likelihood that the relationship we find is causal. We have also conducted extensive robustness checks to mitigate the concern that omitted variables drive our results. We observe that family members engage more in firms in terms of owning equity shares, holding board positions and working as CEOs and managers in these regions. Historical family culture serves as a substitute for formal institutions in financing. Firms in regions with a stronger family culture are more likely to raise initial capital from family members. Family culture also embodies stronger incentives to build family legacies. As shown by the results, entrepreneurs have stronger intentions to pass on firms to their children, and their children are more willing to inherit the firms. Further, historical family culture promotes firm performance partly by decreasing the leverage ratio.

Our study speaks to topics of general interest, even though it is based on data from China. First, a strong family culture can still promote certain types of entrepreneurship, though it might impede overall economic development. Second, there are deep historical roots for contemporary industry structure in terms of the fraction of family firms. The heterogeneity of the prevalence of family firms across countries should be correlated with their family values. Immigrants to the same country from various cultural backgrounds will have different probabilities of establishing family firms. Third, family culture influences the management of firms in various respects due to entrepreneurs' cultural values, indicating that culture not only shapes the outcomes of countries and individuals, but also deeply affects that of firms.

Last but not least, it is worth noticing that role the influence of family culture might be diminished by other cultural traits and norms, though its role in shaping local industrial

structure has been propounding in some regions in East Asia and Europe. One counter-example is the Middle East before the 19th century, during which family ties were very strong, but family firms were rare (Kuran, 2012). This historical fact could be explained by the specific natural and social environment in the Middle East, where families sought to diversify risks and raise capital by building up cross-family partnerships, which was enhanced by the trust among Muslims and Islamic partnership law by then. Furthermore, the Islamic polygyny and the persistence of egalitarian inheritance rules led to a high probability of premature dissolution of enterprises (Ergene and Berker, 2009; Kuran, 2018), resulting in the missing of family firms. This indicates the power of family culture might be attenuated when faced with other strong norms and institutions, highlighting the need for exploring the interactions between family ties and other cultural traits in future research.

References

Aga, G., Francis, D., 2017. As the market churns: productivity and firm exit in developing countries. Small Business Economics 49, 379–403.

Aivazian, V.A., Ge, Y., Qiu, J., 2005. The impact of leverage on firm investment: Canadian evidence. Journal of Corporate Finance 11, 277–291.

Alesina, A., Algan, Y., Cahuc, P., Giuliano, P., 2015. Family values and the regulation of labor. Journal of the European Economic Association 13, 599–630.

Alesina, A., Giuliano, P., 2010. The power of the family. Journal of Economic Growth 15, 93–125.

Alesina, A., Giuliano, P., 2011. Family ties and political participation. Journal of the European Economic Association 9, 817–839.

Alesina, A., Giuliano, P., 2014. Family ties, in: Handbook of Economic Growth. Elsevier. volume 2, pp. 177–215.

Alfaro-Urena, A., Manelici, I., Vasquez, J.P., 2020. The effects of joining multinational supply chains: New evidence from firm-to-firm linkages. Available at SSRN 3376129 .

Algan, Y., Cahuc, P., 2007. The roots of low european employment: family culture? NBER Macroeconomics Annual .

Algan, Y., Cahuc, P., 2010. Inherited trust and growth. American Economic Review 100, 2060–92.

Altonji, J.G., Pierret, C.R., 2001. Employer learning and statistical discrimination. The Quarterly Journal of Economics 116, 313–350.

Amit, R., Villalonga, B., 2014. Financial performance of family firms. The SAGE Handbook of Family Business , 157–178.

Amore, M.D., 2017. Social capital and family control. Explorations in Economic History 65, 106–114.

Amore, M.D., Garofalo, O., Minichilli, A., 2014. Gender interactions within the family firm. Management Science 60, 1083–1097.

Anderson, R.C., Duru, A., Reeb, D.M., 2009. Founders, heirs, and corporate opacity in the united states. Journal of Financial Economics 92, 205–222.

Anderson, R.C., Reeb, D.M., 2003. Founding-family ownership and firm performance: evidence from the s&p 500. The Journal of Finance 58, 1301–1328.

Anderson, R.C., Reeb, D.M., Zhao, W., 2012. Family-controlled firms and informed trading: Evidence from short sales. The Journal of Finance 67, 351–385.

Ayres, G.R., 1990. Rough family justice: Equity in family business succession planning. Family Business Review 3, 3–22.

Bai, L., Wu, L., 2018. Political conflict and development dynamics: Economic legacies of the cultural revolution. Unpublished manuscript, School of Economics, University of Edinburgh .

Bai, L., Wu, L., 2020. Political movement and trust formation: Evidence from the cultural revolution (1966–76). European Economic Review 122, 103331.

Bai, X., 2012. The study of academies in the Ming and Qing dynasties (Mingqing Shuyuan Yanjiu). The Forbidden City Press (Beijing).

Bai, Y., 2019. Farewell to confucianism: The modernizing effect of dismantling china's imperial examination system. Journal of Development Economics 141, 102382.

Bai, Y., Jia, R., 2016. Elite recruitment and political stability: the impact of the abolition of china's civil service exam. Econometrica 84, 677–733.

Baker, H.D., 1979. Chinese family and kinship. Columbia University Press.

Banerjee, A., Iyer, L., 2005. History, institutions, and economic performance: The legacy of colonial land tenure systems in india. American Economic Review 95, 1190–1213.

Banfield, E.C., 1958. The Moral Basis of a Backward Society. New York: Free Press.

Becker, S.O., Hvide, H.K., 2019. Entrepreneur death and startup performance. Available at SSRN 2199645 .

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

Bertrand, M., Johnson, S., Samphantharak, K., Schoar, A., 2008. Mixing family with business: A study of thai business groups and the families behind them. Journal of Financial Economics 88, 466–498.

Bertrand, M., Schoar, A., 2006. The role of family in family firms. Journal of Economic Perspectives 20, 73–96.

Bian, Y., 1994. Guanxi and the allocation of urban jobs in china. The China Quarterly 140, 971–999.

Biao, Y., 2007. Clan, ancestral hall, and sacrifice in the song dynasty. Frontiers of History in China 2, 166–180.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. The Quarterly Journal of Economics 122, 1351–1408.

Bøhren, Ø., Stacescu, B., Almli, L.F., Søndergaard, K.L., 2019. When does the family govern the family firm? Journal of Financial and Quantitative Analysis 54, 2085–2117.

Bol, P.K., 2008. Neo-Confucianism in history. Harvard University Press.

Buggle, J., Durante, R., 2017. Climate risk, cooperation, and the co-evolution of culture and institutions .

Burkart, M., Panunzi, F., Shleifer, A., 2003. Family firms. The Journal of Finance 58, 2167–2201.

Cai, H., Li, H., Park, A., Zhou, L.A., 2013. Family ties and organizational design: Evidence from chinese private firms. Review of Economics and Statistics 95, 850–867.

Campello, M., 2006. Debt financing: Does it boost or hurt firm performance in product markets? Journal of Financial Economics 82, 135–172.

Card, D., Krueger, A.B., 2015. Myth and Measurement: The New Economics of the Minimum Wage-Twentieth-Anniversary Edition. Princeton University Press.

Cetorelli, N., Strahan, P.E., 2006. Finance as a barrier to entry: Bank competition and industry structure in local us markets. The Journal of Finance 61, 437–461.

Chen, M., Xiao, J.Z., Zhao, Y., 2021. Confucianism, successor choice, and firm performance in family firms: Evidence from china. Journal of Corporate Finance 69, 102023.

Chen, S., Chen, X., Cheng, Q., Shevlin, T., 2010. Are family firms more tax aggressive than non-family firms? Journal of Financial Economics 95, 41–61.

Chen, T., Kung, J.K.s., Ma, C., 2020. Long live keju! the persistent effects of china's civil examination system. The Economic Journal 130, 2030–2064.

Chen, Z., Ma, C., Sinclair, A.J., 2022. Banking on the confucian clan: Why china developed financial markets so late. The Economic Journal 132, 1378–1413.

Claessens, S., Djankov, S., Fan, J.P., Lang, L.H., 2002. Disentangling the incentive and entrenchment effects of large shareholdings. The Journal of Finance 57, 2741–2771.

Claessens, S., Djankov, S., Lang, L.H., 2000. The separation of ownership and control in east asian corporations. Journal of Financial Economics 58, 81–112.

Coleman, J.S., 1990. Foundations of social theory .

Conley, T.G., 1999. Gmm estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Currie, T., Turchin, P., Bednar, J., Richerson, P.J., Schwesinger, G., Steinmo, S., Wacziarg, R., Wallis, J., 2016. Evolution of institutions and organizations, MIT Press.

Dalci, I., 2018. Impact of financial leverage on profitability of listed manufacturing firms in china. Pacific Accounting Review .

Davis, J.A., Hampton, M.M., Lansberg, I., et al., 1997. Generation to generation: Life cycles of the family business. Harvard Business Press.

Deryugina, T., MacKay, A., Reif, J., 2020. The long-run dynamics of electricity demand: Evidence from municipal aggregation. American Economic Journal: Applied Economics 12, 86–114.

Desmet, K., Ortuño-Ortín, I., Wacziarg, R., 2017. Culture, ethnicity, and diversity. American Economic Review 107, 2479–2513.

Doepke, M., Zilibotti, F., 2008. Occupational choice and the spirit of capitalism. The Quarterly Journal of Economics 123, 747–793.

Ebenstein, A., 2010. The "missing girls" of china and the unintended consequences of the one child policy. Journal of Human Resources 45, 87–115.

Ebrey, P.B., Watson, J.L., et al., 1986. Kinship Organization in Late Imperial China, 1000-1940. volume 5. Univ of California Press.

Ellul, A., Pagano, M., Panunzi, F., 2010. Inheritance law and investment in family firms. American Economic Review 100, 2414–50.

Enke, B., 2019. Kinship, cooperation, and the evolution of moral systems. The Quarterly Journal of Economics 134, 953–1019.

Ergene, B.A., Berker, A., 2009. Inheritance and intergenerational wealth transmission in eighteenth-century ottoman kastamonu: an empirical investigation. Journal of Family History 34, 25–47.

Fan, J.P., Gu, Q., Yu, X., 2022. Collectivist cultures and the emergence of family firms. The Journal of Law and Economics 65, S293–S325.

Fan, X., 2004. Confucian temples in China (Zhongguo Kongmiao). Sichuang Wenyi Chubanshe.

Fei, J.C., Liu, T.J., 1982. The growth and decline of chinese family clans. Journal of Interdisciplinary History , 375–408.

Feng, E., Chang, J., 2008. The history of Chinese Clans (Zhongguo Zongzu Shi). Shanghai Renmin Chubanshe (Beijing).

Fernández, R., Fogli, A., 2009. Culture: An empirical investigation of beliefs, work, and fertility. American Economic Journal: Macroeconomics 1, 146–77.

Fincher, C.L., Thornhill, R., Murray, D.R., Schaller, M., 2008. Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. Proceedings of the Royal Society B: Biological Sciences 275, 1279–1285.

Galasso, V., Profeta, P., 2018. When the state mirrors the family: the design of pension systems. Journal of the European Economic Association 16, 1712–1763.

Galor, O., Özak, Ö., 2016. The agricultural origins of time preference. American Economic Review 106, 3064–3103.

Gernet, J., 1995. Buddhism in Chinese society: an economic history from the fifth to the tenth centuries. Columbia University Press.

Gethin, R., et al., 1998. The foundations of Buddhism. Oxford University Press.

Giavazzi, F., Schiantarelli, F., Serafinelli, M., 2013. Attitudes, policies, and work. Journal of the European Economic Association 11, 1256–1289.

Giuliano, P., 2007. Living arrangements in western europe: Does cultural origin matter? Journal of the European Economic Association 5, 927–952.

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, 106–137.

Gowdy, J., Krall, L., 2016. The economic origins of ultrasociality. Behavioral and Brain Sciences 39.

Greif, A., 2006. Family structure, institutions, and growth: the origins and implications of western corporations. American Economic Review 96, 308–312.

Greif, A., Tabellini, G., 2010. Cultural and institutional bifurcation: China and europe compared. American Economic Review 100, 135–40.

Greif, A., Tabellini, G., 2017. The clan and the corporation: Sustaining cooperation in china and europe. Journal of Comparative Economics 45, 1–35.

Harrell, S., 1987. On the holes in chinese genealogies. Late Imperial China 8, 53–79.

Harrell, S., et al., 1995. Chinese historical microdemography. volume 20. Univ of California Press.

Hu, S., 2023. Survival of the literati: Social status and reproduction in ming–qing china. Journal of Population Economics , 1–46.

Huillery, E., 2009. History matters: The long-term impact of colonial public investments in french west africa. American Economic Journal: Applied Economics 1, 176–215.

Ishimaru, S., 2024. Empirical decomposition of the iv-ols gap with heterogeneous and nonlinear effects. Review of Economics and Statistics 106, 505–520.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3, 305–360.

Ji, X., 1996. A distionary of academies in China (Zhongguo Shuyuan Cidian). Zhejiang Educational Press.

Jia, N., 2014. Are collective political actions and private political actions substitutes or complements? empirical evidence from china's private sector. Strategic Management Journal 35, 292–315.

Kang, J.K., Kim, J., 2020. Do family firms invest more than nonfamily firms in employee-friendly policies? Management Science 66, 1300–1324.

Kelly, M., 2019. The standard errors of persistence .

Kong, X., Kong, Z., 2011. The study of Confucian temples in the world (Shijie Kongzimiao Yanjiu). Zhongyang Bianyi Chubanshe.

Kuran, T., 2012. The Long Divergence: How Islamic Law Held Back the Middle East. Princeton University Press.

Kuran, T., 2018. Islam and economic performance: Historical and contemporary links. Journal of Economic Literature 56, 1292–1359.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., 1999. Corporate ownership around the world. The Journal of Finance 54, 471–517.

Lang, L., Ofek, E., Stulz, R., 1996. Leverage, investment, and firm growth. Journal of Financial Economics 40, 3–29.

Li, H., Meng, L., Wang, Q., Zhou, L.A., 2008. Political connections, financing and firm performance: Evidence from chinese private firms. Journal of Development Economics 87, 283–299.

Lins, K.V., Volpin, P., Wagner, H.F., 2013. Does family control matter? international evidence from the 2008–2009 financial crisis. The Review of Financial Studies 26, 2583–2619.

Liu, H.c.W., Liu, W., et al., 1959. The traditional Chinese clan rules. volume 7. JJ Augustin.

Lu, Y., Tao, Z., 2009. Contract enforcement and family control of business: Evidence from china. Journal of Comparative Economics 37, 597–609.

Lv, S., 2018. A brief of Chinese clan institutions (Zhongguo Zongzu Zhidu Xiaoshi). Zhishi Chanquan Chubanshe (Beijing).

Mehrotra, V., Morck, R., Shim, J., Wiwattanakantang, Y., 2013. Adoptive expectations: Rising sons in japanese family firms. Journal of Financial Economics 108, 840–854.

Michalopoulos, S., Papaioannou, E., 2013. Pre-colonial ethnic institutions and contemporary african development. Econometrica 81, 113–152.

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, 829–858.

Padró i Miquel, G., Qian, N., Xu, Y., Yao, Y., 2015. Making democracy work: Culture, social capital and elections in china. Social Capital and Elections in China (April 2015). NBER Working Paper .

Mueller, H.M., Philippon, T., 2011. Family firms and labor relations. American Economic Journal: Macroeconomics 3, 218–45.

Nunn, N., Puga, D., 2012. Ruggedness: The blessing of bad geography in africa. Review of Economics and Statistics 94, 20–36.

Opler, T.C., Titman, S., 1994. Financial distress and corporate performance. The Journal of Finance 49, 1015–1040.

Peng, Y., 2004. Kinship networks and entrepreneurs in china's transitional economy. American Journal of Sociology 109, 1045–1074.

Pérez-González, F., 2006. Inherited control and firm performance. American Economic Review 96, 1559–1588.

Pierce, L., Snyder, J.A., 2020. Historical origins of firm ownership structure: The persistent effects of the african slave trade. Academy of Management Journal 63, 1687–1713.

Robb, A.M., Robinson, D.T., 2014. The capital structure decisions of new firms. The Review of Financial Studies 27, 153–179.

Schmid, T., 2013. Control considerations, creditor monitoring, and the capital structure of family firms. Journal of Banking and Finance 37, 257–272.

Shen, N., Su, J., 2017. Religion and succession intention-evidence from chinese family firms. Journal of Corporate Finance 45, 150–161.

Shiue, C.H., 2016. A culture of kinship: Chinese genealogies as a source for research in demographic economics. Journal of Demographic Economics 82, 459–482.

Spolaore, E., Wacziarg, R., 2013. How deep are the roots of economic development? Journal of Economic Literature 51, 325–69.

Sraer, D., Thesmar, D., 2007. Performance and behavior of family firms: Evidence from the french stock market. Journal of the European Economic Association 5, 709–751.

Tabellini, G., 2010. Culture and institutions: economic development in the regions of europe. Journal of the European Economic Association 8, 677–716.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., Li, M., Hong, Y.y., 2014. Large-scale psychological differences within china explained by rice versus wheat agriculture. Science 344, 603–608.

Telford, T.A., 1986. Survey of social demographic data in chinese genealogies. Late Imperial China 7, 118–148.

Tsoutsoura, M., 2015. The effect of succession taxes on family firm investment: Evidence from a natural experiment. The Journal of Finance 70, 649–688.

Villalonga, B., Amit, R., 2006. How do family ownership, control and management affect firm value? Journal of Financial Economics 80, 385–417.

Voigtländer, N., Voth, H.J., 2012. Persecution perpetuated: the medieval origins of anti-semitic violence in nazi germany. The Quarterly Journal of Economics 127, 1339–1392.

Wang, S., 2013. The spatial variation of song clans and its reasons. Anhui Shixue 1.

Wang, Y., 2021. The political legacy of violence during china's cultural revolution. British Journal of Political Science 51, 463–487.

Watson, J.L., 1982. Chinese kinship reconsidered: Anthropological perspectives on historical research. The China Quarterly 92, 589–622.

Weber, M., 1904. The protestant ethic and the spirit of capitalism (charles scribner's sons, new york, 1958). The Sociology of Religion (Beacon Press, Boston 1963) .

Yu, W., Zhu, L., Liang, P., 2023. Traditional clan culture and business dispute management. Economic Science (Chinese) , 220–240.

Zhang, C., 2020. Clans, entrepreneurship, and development of the private sector in china. Journal of Comparative Economics 48, 100–123.

Appendices

A. Appendix

A.1. Suggestive evidence on listed firms

For additional insights regarding the effects of historical family culture on the structure of firms, we extend the analysis to listed firms. Public companies are different from private companies in many aspects such as size, access to the financial market, and regulations, implying a possibly different role of culture in industry structure and firm management. We make use of a panel dataset of all listed firms in China from 2003 to 2019 and conduct a repeated cross section estimation as our main explanatory variable, genealogy density, is time-invariant. We first replicate our analysis from Section 4.1 for the share of family firms among listed firms at the county level.[70] In addition to the Specification 2, we control for the year fixed effects and update the relevant social economic variables to be time-variant in our estimation.

Table A.1 presents the results for the full sample and a sample that excludes the provincial capital cities and sub-provincial cities. As shown in Panel A, there is no similar significantly positive effect of historical family culture in the share of family firms among listed firms in all columns. To address the concern that people may migrant to especially big cities or register firms in big cities for broader market and better access to financial sources, we exclude provincial capital cities and sub-provincial cities from the sample.[71] The subsample results are in Panel B, and we do not observe a consistent positive effect of genealogy density with the sample size being only 1/3 of the full sample. Instead, the coefficients of genealogy density even become negative.

The insignificant impact of historical family culture on the share of family firms may be due to the fact that it is much easier for outside investors to obtain the equity of listed firms through the stock market; hence the firms are less likely to be controlled by a single natural person or family. It could also be possible that firms in regions with stronger family cultures have fewer incentives to go public with hoping to keep the family control over the firm. We do not directly test these potential explanations here; instead, we examine the mechanism of family engagement among listed firms following Section 5.1. The results show that there is no consistently positive impact of genealogy density on all four measures of family engagement in listed family firms.[72]

[70] The family firms in the listed firm analysis are defined by the data source CSMAR. A listed firm is defined as a family firm if the actual control person is a natural person, as shown in the ownership structure of the firm.

[71] Notice that the genealogy density is for the county where a firm is registered, but not the home county of the person who controls the firm.

[72] See Table A.2.

Table A.1: Genealogy density and fraction of listed family firms

Dependent variable	Fraction of Family Firms			
	(1)	(2)	(3)	(4)
Panel A: Full sample				
Genealogy density	0.015	0.018	0.015	-0.009
	(0.012)	(0.011)	(0.011)	(0.010)
Observations	1575	1575	1575	1554
R^2	0.698	0.715	0.717	0.773
Mean dependent vars.	0.500	0.500	0.500	0.500
Panel B: Excluding provincial capitals and sub-provincial cities				
Genealogy density	-0.004	-0.007	-0.012	-0.020**
	(0.009)	(0.009)	(0.010)	(0.008)
Observations	567	567	567	567
R^2	0.707	0.727	0.732	0.826
Mean dependent vars.	0.690	0.690	0.690	0.690
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Year FEs	Yes	Yes	Yes	Yes
Province FEs	Yes	Yes	Yes	Yes

Notes. County-level Fixed Effects regressions. Dependent variable is fraction of family firms in each county in each year. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010 in 2010 , minority rate in 2010 in 2010, , bank density (1990-2010), GDP per capita, health, and a provincial capital dummy. Province and Year fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

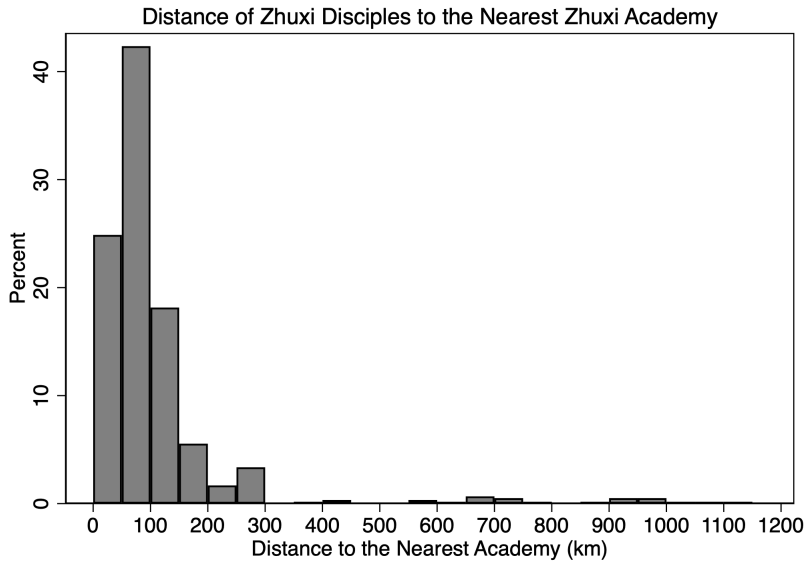
Table A.2: Mechanisms: family engagement in listed firms

Dependent variable	Family engagement in the firm							
	Director Share		Supervisor Share		Manager Share		Salary Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Genealogy	-0.004	-0.006	-0.002	-0.001	-0.007	-0.009**	-0.004	-0.007
	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)	(0.006)	(0.006)
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes	No	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4929	4687	4931	4689	4928	4686	4694	4473
R^2	0.173	0.179	0.087	0.087	0.195	0.205	0.171	0.175
Mean dependent vars.	0.222	0.225	0.020	0.020	0.159	0.161	0.238	0.241

Notes. Firm-level Fixed Effects Estimates. Dependent variables are fraction of directors, supervisors, and managers who are family members. The last dependent variable is the ratio of salary of family members to the total salary. Genealogy density is the number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate 2010 in 2010, bank density (1990-2010), GDP per capita, health and a provincial capital dummy. Province and Year fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.2. Figures

Figure A.1: Distance of Zhuxi Disciples to the Nearest Zhuxi Academy



Notes. This figure shows the distribution of distances between the disciples of Zhu Xi and the nearest academies of Zhu Xi in the Song Dynasty.

A.3. Summary statistics

Table A.3: Summary statistics

County-level variables	Mean	S.D.	Min	Medium	Max	Count
Fraction of family firms	0.497	0.211	0.000	0.500	1.000	204
Genealogy density	0.460	1.974	0.000	0.000	19.409	204
Academy (200 km)	1061.6	1704.1	0.0	268.9	10329.2	204
Altitude	413.9	602.4	0.1	186.0	3336.7	204
Ruggedness	1.739	1.712	0.017	1.322	9.189	204
Distance to coast	408.5	459.2	0.2	208.4	2551.2	204
Calorie suitability	1940.4	517.5	0.0	1977.0	2730.9	204
Biography density	12.7	32.9	0.0	1.1	272.4	204
Population density in 1880	1.612	1.564	0.000	1.459	6.248	204
Tax per capita in 1820	0.072	0.075	0.000	0.049	0.641	204
Bank density (1990-2010)	1.2	1.5	0.2	0.9	18.3	204
Sex ratio in 2010	104.6	6.2	82.9	103.6	141.1	204
Minority rate in 2010	7.3	15.6	0.0	1.3	93.3	204
Urbanization rate in 2010	43.0	24.9	3.8	40.1	94.4	204
Provincial capital dummy	0.353	0.479	0.000	0.000	1.000	204
Firm-level variables	**Mean**	**S.D.**	**Min**	**Medium**	**Max**	**Count**
Equity shares held by family members	0.168	0.225	0.000	0.020	1.000	2874
Family board members	0.376	0.363	0.000	0.333	1.000	1655
Family CEO	0.073	0.260	0.000	0.000	1.000	3365
Family managers	0.197	0.285	0.000	0.000	1.000	2517
Initial capital from family members	0.084	0.277	0.000	0.000	1.000	3374
Initial capital from bank loans	0.239	0.427	0.000	0.000	1.000	3374
Willing to inherit	0.321	0.467	0.000	0.000	1.000	1677
Plan to pass on	0.405	0.491	0.000	0.000	1.000	1828
Return on Asset	0.297	1.128	-13.300	0.082	26.260	2227
Leverage	20.876	28.886	0.000	0.000	360.000	2769
Entrepreneur's age	45.801	8.727	19.000	46.000	93.000	3487
Entrepreneur's education	3.868	1.091	1.000	4.000	6.000	3487
Female	0.153	0.360	0.000	0.000	1.000	3487
log sales	6.472	2.470	0.000	6.661	15.187	3487
Firm age	8.758	4.667	0.000	8.000	21.000	3487
Whether it was privatized	1.840	0.367	1.000	2.000	2.000	3487
log initial asset	5.148	1.736	0.000	4.796	10.915	3004

Table A.4: Genealogy Density and Missing Values in Firm Variables

Dependent variable	Missing Values in:							
	Equity Share		Board Members		CEO		Managers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Genealogy density	-0.010	-0.006	0.006	0.012	0.003	0.004	-0.009	-0.007
	(0.010)	(0.012)	(0.012)	(0.012)	(0.002)	(0.003)	(0.007)	(0.007)
Geographic controls	No	Yes	No	Yes	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4382	4382	4382	4382	4382	4382	4382	4382
R^2	0.057	0.086	0.029	0.038	0.029	0.038	0.053	0.064

Notes. Firm-level OLS regressions. Outcome variables are dummies indicating whether this is a missing value in the fraction of equity share held by family members, the fraction of board members that are family members, whether CEO is a family member and the fraction of managers who are family members. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

A.4. Validation of genealogy

Table A.5: Validation of Genealogy: Other Estimation Methods

Dependent variable	Happy family important		Child. having achieve. important		Family relation. are close	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.047***	0.050***	0.048***	0.046***	0.055***	0.036*
	(0.018)	(0.015)	(0.018)	(0.016)	(0.016)	(0.020)
Controls	No	Yes	No	Yes	No	Yes
Observations	29514	28525	29514	28525	27073	26175
Pseudo R^2	0.001	0.029	0.001	0.029	0.001	0.030
Dependent variable	Visit grave last year		# Relatives visit spring festival			
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy densit	0.106***	0.073***	0.081***	0.071***		
	(0.034)	(0.028)	(0.020)	(0.019)		
Controls	No	Yes	No	Yes		
Observations	29504	28515	29348	28363		
Pseudo R^2	0.004	0.051	0.001	0.019		

Notes. Ordered probit regressions are adopted for the first four dependent variables, and negative binomial regressions are used for "# Relatives visit in spring festival". All continuous independent variables are standardized. Controls include individual level age, gender, household registration status FEs, ethnic group FEs, education FEs, marriage status FEs, and county level geographic controls, historical controls, and socioeconomic controls. Standard errors clustered at the county level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Validation of Genealogy: Using Genealogy Book in CFPS Data

Dependent variable	Happy family important		Child. having achieve. important		Family relation. are close	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy book	0.013	0.025	0.034**	0.029*	0.059**	0.042*
	(0.015)	(0.017)	(0.015)	(0.016)	(0.023)	(0.023)
Controls	No	Yes	No	Yes	No	Yes
Observations	33289	29421	33289	29421	30632	27038
R^2	0.000	0.157	0.000	0.154	0.001	0.428
Dependent variable	Visit grave last year		# Relatives visit #spring festival		Principal component of family ties	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy book	0.144***	0.150***	0.253***	0.132***	0.071***	0.059***
	(0.021)	(0.023)	(0.023)	(0.024)	(0.017)	(0.018)
Controls	No	Yes	No	Yes	No	Yes
Observations	33281	29414	33099	29254	33293	29425
R^2	0.004	0.352	0.011	0.312	0.001	0.228

Notes. OLS regressions. All continuous variables are standardized. Controls include age, gender, household registration status FEs, ethnic group FEs, education FEs, marriage status FEs and community or village FEs. Standard errors clustered at the household level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Genealogy and Other Cultural Norms

Dependent variable	Trust		Ancestral hall		Rice suitability	
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.002	0.007	-0.000	-0.052	0.344***	0.238***
	(0.007)	(0.007)	(0.019)	(0.059)	(0.045)	(0.043)
Controls	No	Yes	No	Yes	No	Yes
Observations	22196	21566	29055	28209	28956	28491
R^2	0.000	0.038	0.000	0.152	0.122	0.546

Notes. OLS regressions. All continuous variables are standardized. Controls include individual level age, gender, household registration status FEs, ethnic group FEs, education FEs, marriage status FEs, and county level geographic controls, historical controls, and socioeconomic controls. Standard errors clustered at the county level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Validation of Genealogy: Controlling for Other Cultural Norms

Dependent variable	Principal component of family ties					
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.045***	0.041***	0.047***	0.054***	0.049***	0.044**
	(0.014)	(0.015)	(0.015)	(0.016)	(0.016)	(0.017)
Trust others		0.020				0.022
		(0.015)				(0.015)
Ancestral hall			0.016			0.025
			(0.020)			(0.018)
Rice suitability				-0.148		-0.055
				(0.116)		(0.046)
Academy(200 km)					0.012	0.025
					(0.027)	(0.029)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28612	21566	28209	28491	28612	21165
R^2	0.058	0.074	0.069	0.068	0.068	0.077

Notes. OLS regressions. All continuous variables are standardized. Controls include individual level age, gender, household registration status FEs, ethnic group FEs, education FEs, marriage status FEs, and county level geographic controls, historical controls, and socioeconomic controls. Standard errors clustered at the county level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A.5. Mechanisms

Table A.9: Genealogy density and fraction of family firms: the role of bank density

Dependent Variable:	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.051***	0.052***	0.051***	0.052***
	(0.018)	(0.020)	(0.018)	(0.021)
Bank density			0.002	0.001
			(0.012)	(0.013)
Geographic controls	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.254	0.268	0.254	0.268

Notes. County-level OLS regressions. Dependent variable is the fraction of family firms in each county. Genealogy density is the number of genealogies from 1368 to 1949, normalized by 1953 population. Bank density is the average number of banks established during 1990 to 2010 in each county, normalized by 2000 population Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010 and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.6. IV: Decomposition of the IV-OLS gap and exclusion restrictions

Table A.10: Decomposition of the IV-OLS Gap

Coefficients			Decomposition		
OLS	IV	IV-OLS Gap	Covariate Weight	Treat Weight	Endogeneity
(1)	(2)	(3)	(4)	(5)	(6)
0.052	0.237	0.185	0.161	0.024	-0.000
(0.021)	(0.115)	(0.105)	(0.243)	(0.255)	(0.000)

Notes: Column (1) to column (3) report the OLS estimates, the IV estimates, and their gaps. Column (4) to column (6) report the estimates of the covariate weight difference, the treatment-level weight difference, and the marginal effect difference components. Standard errors are in parentheses.

Table A.11: IV: exclusion restrictions

Dependent variable	Urbanization in 2010		Average edu in 2010		Sex ratio in 2010		Bank density (1990-2010)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Academy(200 km)	0.168	-2.434	0.207	0.072	1.759*	1.395	0.597	0.481
	(3.449)	(2.970)	(0.187)	(0.075)	(1.020)	(1.117)	(0.507)	(0.442)
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204	204	204	204
R^2	0.280	0.557	0.320	0.897	0.318	0.419	0.145	0.260
Mean dependent vars.	43.0	43.0	9.8	9.8	104.6	104.6	1.2	1.2

Notes. County-level OLS regressions. Academy (200 km) measures the influence of Confucian academies within 200 km during Song dynasty. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy (We exclude the corresponding variable if it is the dependent variable). Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.12: IV: exclusion restrictions

Dependent variable	Confucianism in 2010		Fraction of family firms	
	(1)	(2)	(3)	(4)
Academy (200 km)	317.7***	215.6**		
	(111.741)	(96.746)		
Confucianism in 2010			0.031	0.024
			(0.025)	(0.031)
Geographic controls	Yes	Yes	Yes	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.708	0.783	0.243	0.253
Mean dependent vars.	1017.2	1017.2	0.497	0.497

Notes. County-level OLS regressions. Academy (200 km) measures the influence of Confucian academies within 200 km during Song dynasty. Confucianism in 2010 measures the influence of contemporary Confucian temples within 200 km using the same method as Academy (200 km). Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7. Robustness checks

A.7.1. The Cultural revolution and political influence

Historical examples of attempts at eliminating the family as an institution, such as the Cultural Revolution, have been catastrophic failures (Alesina and Giuliano, 2014). Zhang (2020) finds that the number of genealogies compiled before 1950 in each prefecture is strongly correlated with that after 1980,[73] which suggests the persistence of family culture. However, on the other hand, Greif and Tabellini (2017) point out that the Cultural Revolution impacted family clan culture and destroyed some genealogies. To mitigate the concern that the Cultural Revolution might bias our estimation, we first examine the correlation between genealogy density and the intensity of the Cultural Revolution. Following Bai and Wu (2018) and Bai and Wu (2020), Cultural Revolution is measured by the number of revolution-related deaths per 1,000,000 local population during the Cultural Revolution, and the logarithm transformation of the death number plus 0.01. Table A.13 presents the results. We gradually include baseline controls in regressions, and column (3) shows that the Cultural Revolution has positive but insignificant correlation with the genealogy density. In contrast, the instrumental variable, Song Confucian academies, still displays strong predictive power for genealogy. In addition, we control for the intensity of the Cultural

[73] We also find this result.

Revolution and political influence, which may affect both family firm share and genealogy density in the baseline specification. The political influence is proxied by the distance to Beijing of each county (Wang, 2021). As one can see in Table A.14, the coefficients remain positively significant after controlling for the intensity of the Cultural Revolution in the first four columns or controlling for political influence in the last two columns. The magnitude is close to our baseline estimation, suggesting that the Cultural Revolution or political influence does not drive our results. The impact of Cultural Revolution on family firms is, however, small and insignificant.

Further, we calculate genealogy density after 1976 using the number of genealogies compiled after the Cultural Revolution. Despite much fewer genealogies compiled during the contemporary period, genealogy density after 1976 is still positively correlated with the prevalence of family firms, as one can see in Table A.15.

A.7.2. Keju and Confucianism

Chen et al. (2022) indicate that genealogy density could also be related to Keju (the civil examination from the Sui Dynasty until the Qing Dynasty) and Confucianism because successful candidates in Keju exams were likely to compile genealogies. Confucian literature dominates Keju's content. Candidates who passed the provincial-level tests were Juren (recommended man, provincial graduate), and those who passed national-level tests were Jinshi (presented scholar, palace examination graduate). Bai and Jia (2016) and Chen et al. (2020) discussed Keju exams in more detail. Confucianism is measured by Jinshi density in Chen et al. (2020) and Juren density in Bai (2019). In our baseline estimation, we control for the county-level historical biography density, which is mainly the biographies of successful candidates in Keju exams. To measure the influence of Keju and Confucianism in alternative ways, we control for Jinshi density in column (1), and Juren density[74] in column (2) of Table A.16. In column (3), we employ another measure of Confucianism based on the distances of contemporary Confucian temples within 200 km of each county, using the same formula as Equation 3.[75] In column (4), we control for our instrument, in case one might concern that it is a measure of Confucianism. Across columns, we find our results are highly robust when we control for various measures of Confucianism. It demonstrates that family culture, instead of Confucianism, predicts the ratio of modern family firms.

A.7.3. Economic prosperity and average level of education

Family culture is correlated with economic development levels (Bertrand and Schoar, 2006). In our baseline estimation, we control for variables on historical economic prosperity, but these variables could contain measurement errors. In the matching approach, we have

[74]Both Jinshi density and Juren density data are from Bai and Jia (2016), only prefecture-level data is available and the data is missing in some prefectures.

[75]The result is similar if we use each county's shortest distance to a contemporary Confucian temple.

shown that family culture still has a strong predictive power when we compare counties with similar economic development levels. Furthermore, in this part, we control for the GDP per capita in 2010 in the first two columns and average nighttime light in 2010 in the last two columns of Table A.17. The results suggest that genealogy density still has a positively significant effect on the fraction of family firms.

We account for the potential impact of the average level of education by controlling for county-level average years of schooling and illiteracy rate in 2010 in Table A.18. The inclusion of these controls has little effect on our results.

A.7.4. Rice suitability and wheat suitability

The degree of interdependency and independency, which can shape family culture, varies greatly between northern and southern China due to differences in the suitability of rice and wheat agriculture (Talhelm et al., 2014). To address this concern, we control for rice suitability and wheat suitability, respectively, and present the results in Table A.19. The results indicate that family culture continues to significantly predict the fraction of family firms with a similar magnitude.

A.7.5. Migration

Entrepreneurs could migrate to relatively developed cities due to better business environments or enlarged markets. They are not necessarily influenced by local culture. To allay this concern, we exclude counties belonging to a provincial capital or a sub-provincial city, leaving only 130 observations in our sample. Unsurprisingly, Table A.20 demonstrates that the effects of genealogy density are even more pronounced in these relatively smaller cities with few migrants. Further, we compute each county's migration rate[76] and immigration rate[77] using the 2010 population census and add them as a control separately in our baseline estimation. Table A.21 shows that our results are robust when controlling for migration rate or immigration rate.

A.7.6. Marketization and government-business relations

Marketization and government-business relations could also impact the prevalence of family firms. Regions with a lower marketization level or less supportive government might rely more on family ties to organize production. To account for these factors, we control for a marketization index and a government-business relation index. The marketization index is a continuous variable measuring the relative marketization process in each prefecture in 2010. It is composed of five indexed, each of which reflects a specific aspect of marketization. They are the relationship between the government and the market, the development of

[76]Ratio of migrants to the registered population.
[77]Ratio of immigrants to the registered population.

the non-state-owned economy, the development of the product market, the development of the factor market, and the development of the market intermediary organization and legal system environment. The index is obtained from China Market Index Database. Government-business relation index is a continuous indicator measuring the relationship between government and business in promoting business development. It is a weighted average index that comprises five aspects: how much attention governments give to firm development, the public service provided by governments, the tax burden of firms, the integrity of governments, and the transparency of governments. The index is from China government-business relations ranking 2018.[78]

Table A.22 shows that the results are very close to our baseline estimation. Conditional on marketization and government-business relations, family culture still positively predicts the fraction of family firms.

A.7.7. Alternative definitions of family firms

As discussed in Section 4.1, we define a firm to be a family firm if the founding family owns at least 50% of equity shares and multiple family members engage in the firm (either holds positive equity shares, or board member position or CEO position). This definition could differ from some of the literature regarding the threshold of family ownership and whether to ensure multiple family members' engagement. Among private firms, it is sensible to set the threshold to be 50% to guarantee the control power of the family (Bøhren et al., 2019; Amore et al., 2014). Referring to Villalonga and Amit (2006) and Ellul et al. (2010), we also set the threshold to be 10% and 20%, although they study publicly listed firms. Furthermore, we increase the threshold to 33.3% and 66.7%, following Bøhren et al. (2019), which represents a family's different control power over the firm. Appendix Table A.23 demonstrates that our results are insensitive to the shift of the thresholds with family engagement being guaranteed.

Table A.23 also displays results without ensuring other family members' engagement. The coefficients are still positive and significant, despite a decrease in the magnitude. This suggests a strong link between genealogy density and multiple family members' involvement in the firm. Family culture will increase the concentration of control powers in a family, but might not in a single individual.[79]

A.7.8. Alternative samples of counties

We select counties with at least a certain number of firms surveyed. This threshold might impact our results. There is a tradeoff when setting the threshold. A low threshold will

[78] Too many missing values in previous years.

[79] The results in Table 9 further emphasize the importance of family engagement, which is a plausible manifestation of willingness of inheritance, in displaying the key feature of family firms and in keeping the controlling power over the firm for the long run.

lead to more noisy observations since it would be questionable if we only use one or two firms to compute family firm share within a county, whilst a high threshold will lead to a small and unrepresentative sample. We choose counties with at least 5 firms surveyed, taking into consideration both the accuracy of family firm share and sample size. To ensure our results are robust under different thresholds, we report the results of selecting counties with at least 3 or at least 10 firms surveyed in the Appendix (Table A.24 and Table A.25). In addition, we also analyze the results at the firm level, with the outcome variable to be a dummy indicating whether a firm is family firm in Table A.26.

As we can see, when the threshold is 3, the sample size increases from 204 to 276. The results remain significant with similar coefficients to our baseline. When we set the threshold at 10, the accuracy of the measure of family firm share has been improved, but there is also a reduction in sample size from 204 to 114. The coefficients become even more significant with a smaller sample size, which results from decreased standard errors. Moreover, at the firm level, we consistently find the strong positive correlation.

Furthermore, there are counties without any genealogies recorded. We are concerned that there could be systematic differences between counties with genealogies and counties without genealogies, which cannot be addressed by adding controls. We examine whether our conclusions hold if we only consider counties with a positive number of genealogies. As reported in Table A.27 in the Appendix, the estimates become even more pronounced after excluding counties without any genealogies.

A.7.9. Alternative methods

First, considering that our outcome variables essentially are fractions and are left censored at 0 and right censored at 1, we adopt Tobit estimation and report the coefficients in Table A.28. Second, our baseline specification always includes province fixed effects. Using this approach, we focus on variation within provinces, which has its advantages since unobservable characteristics could differ systematically across provinces. However, we also report results without province fixed effects in Table A.29. Third, to alleviate the concern of the spatial correlation of the observations, we calculate Conley spatial standard errors of baseline estimates (Conley, 1999) with a cutoff distance of 100km and 500km in Table A.30. We also report standard errors clustered at the prefecture level and province level in the same table. Finally, we employ genealogy density in logarithm (plus one) to mitigate the bias from extreme values, as in Table A.31. In brief, all these alternative methods confirm our baseline results.

A.7.10. Instrumental variable approach and matching estimation

In Equation 3, we calculate the influence of Song Confucian academies within 200 km. This might overestimate or underestimate the actual effects of an academy. Consequently, using the same equation, we amend the threshold to be 100 km and 300 km as robustness checks.

As one can see in Table A.33 and A.34, the estimates are robust. Additionally, we adopt the count of Confucian academies within 200 km as another robustness check in Table A.35. It suggests that our results are not driven by the particular formula which computes the influence of academies.

Although geographically nearby counties share various unobservable characteristics, one might still be concerned that the matching approach based on geographical distance cannot capture all of the unobservables. For instance, nearby counties could have differential economic development levels. To mitigate this concern, we employ our baseline matching method separately based on GDP per capita and average night light in 2010. Specifically, for each county, we find its "mutually nearest county" in terms of GDP per capita or average night light (pairs are disjoint). There are some papers using a matching approach based on characteristics other than geographical distance, such as Deryugina et al. (2020) and Alfaro-Urena et al. (2020). The results are displayed in the Panel A and B of Table A.32, respectively.

As one can see, for a pair of counties with similar economic prosperity, the county with stronger family ties is more likely to have a higher share of family firms. It provides another dimension of evidence and demonstrates that our findings are not driven by economic prosperity. Combining the evidence from matching based on geographical distance and matching based on economic prosperity, we bolster our confidence that it should be genealogy density, other than unobservable variables, that explains the difference in the prevalence of family firms.

Another concern for the matching approach is that neighboring counties along the Qinling-Huaihe Line (on the rice-wheat border) are different with respect to the collectivist culture, which might impact the prevalence of family firms Fan et al. (2022). Qinling-Huaihe Line is near the latitude of 33 degrees north, and we only keep pairs of counties whose latitudes are both above 34 degrees north or both below 32 degrees north. Consequently, pairs near or across the Qinling-Huaihe Line will be excluded. The results are displayed in Panel C of Table A.32. The estimates are close to our baseline matching approach since only several pairs are dropped.

Our baseline matching approach ensures each pair contains mutually nearest counties. Consequently, an observation would be dropped if its nearest county finds another county to be closer. To allay the concern of small sample size, we find each county's nearest neighbor to form a pair (not necessarily mutually nearest), which guarantees that no observation would be dropped. However, on the other hand, the same county could appear several times if it is the nearest county to multiple counties. Panel D of Table A.32 demonstrates that the effects of historical family culture on the distribution of family firms are still very robust, even though the magnitude decreases a little compared to Table 4.

A.7.11. A placebo test

To further address the concern that genealogies might reflect historical economic prosperity, we employ Buddhist site density in 1820[80] as a placebo test. Although Buddhism originates in India, it has become an inseparable part of Chinese society, both historically and currently. According to Gernet (1995), as early as 1021 CE, there were 458,855 recorded Buddhist monks and nuns actively living in monasteries. However, orthogonal to family culture, Buddhism aims to overcome suffering caused by personal desire. Buddhist traditions usually emphasize transcending the individual self, reaching the realm of no desire and non-existence of the self (Gethin et al., 1998). It does not emphasize the importance of family nor mutual support among family members. However, historically, building a Buddhist temple involved high costs - much higher than compiling genealogies. Consequently, we regard it as a reasonable placebo to examine whether our results are driven by historical economic prosperity.

As can be observed in Table A.37, historical Buddhism basically exerted no influence on contemporary family firms. Overall, the coefficients are very close to zero, which indicates that regions capable of constructing more Buddhist temples in history do not differ from those with worse economic conditions in terms of the contemporary fraction of family firms.

This placebo test also mitigates the concern that our findings could be driven by other cultural factors. In historically more developed regions or regions with stronger traditional culture, people might be reluctant to embrace modern rules or cooperate with strangers to establish a firm. However, as a representative traditional culture, Buddhism displays no impact on the distribution of family firms. Our previous conclusions are more likely to be driven by family culture, not other traditional cultures that do not emphasize family values.

A.7.12. Robustness: other possible confounding factors

[80] Number of Buddhist temples and monasteries in 1820 in each county, normalized by 1953 population.

Table A.13: Sources of Variation of Genealogy Density

Dependent Variable:	Genealogy Density		
	(1)	(2)	(3)
Altitude	-0.134	-0.158	-0.091
	(0.113)	(0.154)	(0.151)
Ruggedness	0.130*	0.144*	0.098
	(0.071)	(0.082)	(0.070)
Distance to coast	-0.084	-0.071	-0.049
	(0.078)	(0.082)	(0.081)
Calorie index pre1500	-0.050	-0.081	0.007
	(0.055)	(0.088)	(0.090)
Celebrity density		-0.017	-0.046*
		(0.016)	(0.024)
Pop density in 1880		0.064	0.008
		(0.205)	(0.191)
Tax per capita 1820		-0.024	-0.057
		(0.058)	(0.061)
Academy (200 km)			0.415***
			(0.119)
Cultural revolution			0.003
			(0.023)
Province FEs	Yes	Yes	Yes
Observations	204	204	204
R^2	0.372	0.374	0.432

Notes. OLS regressions. All the variables are standardized here. Standard errors clustered at the county level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14: Cultural Revolution and Political Influence

Dependent variable	Fraction of family firms					
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.054***	0.052**	0.055***	0.054***	0.053***	0.052**
	(0.018)	(0.021)	(0.017)	(0.021)	(0.017)	(0.021)
Cultural revolution mortality	0.002	0.001				
	(0.003)	(0.003)				
ln(Culture revolution mortality + 0.01)			0.010	0.011		
			(0.007)	(0.008)		
ln (Distance to Beijing)					0.018	0.006
					(0.031)	(0.035)
Geographic controls	No	Yes	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204	204
R^2	0.252	0.270	0.259	0.279	0.250	0.268
Mean dependent vars.	0.497	0.497	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. In the first two columns, we control for the number of revolution-related deaths per 1,000,000 people during the cultural revolution. In columns 3 and 4, we control for the ln (Cultural revolution mortality + 0.01). In the last two columns, we control for the distance to Beijing. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.15: Genealogy Density after 1976

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density after 1976	0.009**	0.008	0.010*	0.010*
	(0.005)	(0.005)	(0.006)	(0.006)
Distance to Beijing	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.233	0.238	0.248	0.253
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. We only consider counties with at least 5 firms surveyed. Genealogy density after 1976 is number of genealogies after 1976, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.16: Keju and Confucianism

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.060***	0.060***	0.054***	0.033*
	(0.019)	(0.015)	(0.020)	(0.019)
Jinshi density	Yes	No	No	No
Juren density	No	Yes	No	No
Confucianism 2010	No	No	Yes	No
Academy (200 km)	No	No	No	Yes
Geographic controls	Yes	Yes	Yes	Yes
Historical controls	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	176	176	204	204
R^2	0.252	0.267	0.272	0.291
Mean dependent vars.	0.503	0.503	0.497	0.497

Notes. County-level OLS regressions. In column (1), we control for Jinshi density. In column (2), we control for Juren density. In column (3), we control for a measure of Confucianism based on distances to contemporary Confucian temples within 200 km. In column (4), we control for our instrument. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.17: Economic Prosperity

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.058***	0.072***	0.055***	0.057**
	(0.018)	(0.020)	(0.018)	(0.022)
GDP per capita	Yes	Yes	No	No
Night light density	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	140	140	198	198
R^2	0.364	0.445	0.257	0.278
Mean dependent vars.	0.499	0.499	0.494	0.494

Notes. County-level OLS regressions. In the first two columns, we control for GDP per capita in 2010. In the last two columns, we control for average nighttime light in 2010. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.18: Average Level of Education

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.053***	0.051**	0.053***	0.052**
	(0.017)	(0.021)	(0.017)	(0.021)
Years of schooling	Yes	Yes	No	No
Illiteracy rate	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.250	0.269	0.250	0.268
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. In the first two columns, we control for average years of schooling 2010. In the last two columns, we control for county-level illiteracy rate in 2010. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.19: Rice Suitability and Wheat Suitability

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.051***	0.049**	0.054***	0.058***
	(0.017)	(0.019)	(0.018)	(0.020)
Rice suitability	Yes	Yes	No	No
Wheat suitability	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.270	0.289	0.259	0.287
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. In the first two columns, we control for rice suitability. In the last two columns, we control for wheat suitability. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.20: Genealogy density and fraction of family firms (exclude provincial capital and sub-provincial cities)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.058***	0.059***	0.061***	0.066***
	(0.019)	(0.020)	(0.022)	(0.024)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	130	130	130	130
R^2	0.228	0.234	0.277	0.286
Mean dependent vars.	0.491	0.491	0.491	0.491

Notes. County-level OLS regressions. Counties belonging to a provincial capital or a sub-provincial city are removed. Dependent variable is fraction of family firms in each county. We only consider counties with positive genealogies and at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010 and bank density (1990-2010). Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.21: Migration

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.054***	0.056***	0.053***	0.050**
	(0.017)	(0.021)	(0.018)	(0.022)
Migration rate	Yes	Yes	No	No
Immigration rate	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.251	0.271	0.250	0.269
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. In the first two columns, we control for the migration rate in 2010. In the last two columns, we control for the immigration rate in 2010. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.22: Marketization and Government-business Relations

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.054***	0.052**	0.053***	0.055**
	(0.018)	(0.021)	(0.018)	(0.022)
Marketization	Yes	Yes	No	No
Gov-business	No	No	Yes	Yes
Geographic controls	No	Yes	No	Yes
Historical controls	No	Yes	No	Yes
Socioeconomic controls	No	Yes	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	201	201	192	192
R^2	0.246	0.265	0.237	0.262
Mean dependent vars.	0.494	0.494	0.488	0.488

Notes. County-level OLS regressions. In the first two columns, we control for the marketization index. In the last two columns, we control for the government-business relations index. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7.13. Robustness: alternative definitions of family firms

Table A.23: Genealogy density and fraction of family firms (OLS)

Dependent variable	Fraction of family firms (10% share + family engage.)			
	(1)	(2)	(3)	(4)
Genealogy density	0.049***	0.047***	0.046**	0.048**
	(0.018)	(0.018)	(0.019)	(0.021)
Dependent variable	Fraction of family firms (20% share + family engage.)			
Genealogy density	0.050***	0.048**	0.047**	0.049**
	(0.018)	(0.019)	(0.020)	(0.021)
Dependent variable	Fraction of family firms (33.3% share + family engage.)			
Genealogy density	0.051***	0.049***	0.048**	0.050**
	(0.018)	(0.018)	(0.020)	(0.021)
Dependent variable	Fraction of family firms (66.7% share + family engage.)			
Genealogy density	0.052***	0.050***	0.050***	0.052***
	(0.015)	(0.016)	(0.016)	(0.018)
Dependent variable	Fraction of family firms (50% share)			
Genealogy density	0.018**	0.017*	0.016*	0.015
	(0.009)	(0.010)	(0.009)	(0.010)
Dependent variable	Fraction of family firms (66.7% share)			
Genealogy density	0.027**	0.026**	0.026**	0.024*
	(0.012)	(0.012)	(0.012)	(0.012)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county under different definitions of being a family firm. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7.14. Robustness: alternative samples

Table A.24: Genealogy density and fraction of family firms (counties with at least 3 firms surveyed)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.049**	0.049**	0.056**	0.055**
	(0.023)	(0.024)	(0.023)	(0.023)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	276	276	276	276
R^2	0.220	0.232	0.247	0.256
Mean dependent vars.	0.515	0.515	0.515	0.515

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 3 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.25: Genealogy density and fraction of family firms (counties with at least 10 firms surveyed)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.042***	0.040***	0.042***	0.049***
	(0.013)	(0.014)	(0.014)	(0.015)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	114	114	114	114
R^2	0.405	0.414	0.433	0.454
Mean dependent vars.	0.492	0.492	0.492	0.492

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 10 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.26: Genealogy density and the distribution of family firms (Probit, marginal effects)

Dependent variable	Family firm dummy			
	(1)	(2)	(3)	(4)
Genealogy density	0.039***	0.039***	0.040***	0.044***
	(0.010)	(0.010)	(0.010)	(0.010)
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	3077	3077	3077	3077
Pseudo R^2	0.057	0.057	0.059	0.060
Mean dependent vars.	0.506	0.506	0.506	0.506

Notes. Firm-level probit regressions (marginal effects reported). Dependent variable is a dummy indicating a family firm. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log sales, firm age, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.27: Genealogy density and fraction of family firms (counties with positive genealogies)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.057***	0.063***	0.061***	0.071***
	(0.020)	(0.021)	(0.019)	(0.021)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	75	75	75	75
R^2	0.398	0.428	0.504	0.523
Mean dependent vars.	0.525	0.525	0.525	0.525

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with positive genealogies and at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7.15. Robustness: alternative methods

Table A.28: Genealogy density and fraction of family firms (Tobit)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.054**	0.052**	0.051**	0.053**
	(0.024)	(0.024)	(0.024)	(0.024)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
Pseudo R^2	-2.713	-2.779	-2.880	-2.964
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level Tobit regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. * p<0.1; ** p<0.05; *** p<0.01.

Table A.29: Genealogy density and fraction of family firms (droping province FEs)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.043***	0.038***	0.039***	0.039***
	(0.011)	(0.012)	(0.012)	(0.013)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	No	No	No	No
Observations	204	204	204	204
R^2	0.020	0.029	0.041	0.047
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are not included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.30: Genealogy density and fraction of family firms (different standard errors)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.053	0.051	0.050	0.052
Clustered SE (city)	(0.017)***	(0.018)***	(0.019)**	(0.021)**
Clustered SE (province)	(0.019)***	(0.019)**	(0.019)**	(0.020)**
Conley SE (100km)	(0.015)***	(0.015)***	(0.017)***	(0.017)***
Conley SE (500km)	(0.010)***	(0.010)***	(0.010)***	(0.011)***
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.250	0.254	0.262	0.268
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors clustered at the city level, standard errors clustered at the province level, Conley standard errors with a cutoff of 100km and 500km are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.31: Genealogy density and fraction of family firms (take log)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
log (Genealogy density+1)	0.045***	0.044***	0.044***	0.049***
	(0.014)	(0.014)	(0.015)	(0.016)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.261	0.265	0.274	0.283
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. * p<0.1; ** p<0.05; *** p<0.01.

A.7.16. Robustness: alternative matching estimates

Table A.32: Robustness: alternative matching estimates

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Panel A: matching based on GDP per capita				
Genealogy density	0.122***	0.116***	0.147***	0.213**
	(0.033)	(0.032)	(0.036)	(0.071)
Panel B: matching based on night light density				
Genealogy density	0.190**	0.185***	0.176**	0.202**
	(0.078)	(0.066)	(0.073)	(0.092)
Panel C: matching without Qinling-Huaihe counties				
Genealogy density	0.060***	0.073***	0.061**	0.053*
	(0.017)	(0.013)	(0.024)	(0.029)
Panel D: matching with replacement				
Genealogy density	0.046***	0.046***	0.041***	0.038***
	(0.015)	(0.016)	(0.015)	(0.013)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Pair FEs	Yes	Yes	Yes	Yes

Notes. Alternative county-level matching approaches based on GDP per capita, night light density, without qinling-huaihe counties, and with replacement separately. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects and pair fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7.17. Robustness: IV and matching

Table A.33: Genealogy density and fraction of family firms (IV – 100 km)

Second stage results				
Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.243*	0.252**	0.233*	0.252**
	(0.126)	(0.125)	(0.120)	(0.120)
Mean dependent vars.	0.497	0.497	0.497	0.497
First stage results				
Dependent variable	Genealogy density			
	(1)	(2)	(3)	(4)
Academy(within 100km)	0.268***	0.290***	0.310***	0.315***
	(0.077)	(0.087)	(0.089)	(0.089)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
KP F-statistic	12.09	11.21	12.22	12.45
R^2	0.399	0.412	0.417	0.435
Mean dependent vars.	0.460	0.460	0.460	0.460

Notes. County-level IV regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. We employ the influence of Confucian academies within 100 km during Song dynasty as the IV for genealogy density. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.34: Genealogy density and fraction of family firms (IV– 300 km)

Second stage results				
Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.228*	0.240*	0.224*	0.233*
	(0.120)	(0.124)	(0.120)	(0.120)
Mean dependent vars.	0.497	0.497	0.497	0.497
First stage results				
Dependent variable	Genealogy density			
	(1)	(2)	(3)	(4)
Academy(within 300 km)	0.346***	0.364***	0.386***	0.390***
	(0.094)	(0.099)	(0.104)	(0.106)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
KP F-statistic	13.54	13.53	13.69	13.57
R^2	0.403	0.412	0.417	0.436
Mean dependent vars.	0.460	0.460	0.460	0.460

Notes. County-level IV regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. We employ the influence of Confucian academies within 300 km during Song dynasty as the IV for genealogy density. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.35: Genealogy density and fraction of family firms (IV – number of academies)

Second stage results				
Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Genealogy density	0.230*	0.242*	0.225	0.233*
	(0.140)	(0.145)	(0.141)	(0.140)
Mean dependent vars.	0.497	0.497	0.497	0.497
First stage results				
Dependent variable	Genealogy density			
	(1)	(2)	(3)	(4)
No. of Academy(within 200 km)	0.312***	0.325***	0.346***	0.352***
	(0.088)	(0.092)	(0.095)	(0.093)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
KP F-statistic	12.68	12.44	13.28	14.36
R^2	0.032	0.040	0.054	0.053
Mean dependent vars.	0.460	0.460	0.460	0.460

Notes. County-level IV regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. We employ number of Confucian academies within 200 km during Song dynasty as the IV for genealogy density. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

A.7.18. Robustness: results for firms

Table A.36: Firm performance and the underlying mechanism by firm type

Dependent variable	ROA		Leverage		ROA	
Sample	Family firm	Non-family	Family firm	Non-family	Family firm	Non-family
	(1)	(2)	(3)	(4)	(5)	(6)
Genealogy density	0.101***	0.118*	-0.101***	0.006	0.083***	0.120**
	(0.024)	(0.062)	(0.028)	(0.032)	(0.021)	(0.059)
Leverage					-0.182***	-0.228***
					(0.038)	(0.059)
Observations	1177	1050	1177	1050	1177	1050
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
P-value difference	0.798		0.007		0.559	
Observations	1177	1050	1177	1050	1177	1050
R^2	0.075	0.083	0.192	0.302	0.088	0.114

Notes. Firm-level OLS regressions. We use ROA to measure firm performance by family and non-family firms. In column (1) and (2), we do not control for leverage. Genealogy density is number of genealogies from 1368 to 1949, normalized by 1953 population. Individual controls include age, education fixed effects and gender of the entrepreneur. Firm controls include log sales, firm age, a dummy indicating whether the firm was privatized and industry fixed effects. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Standard errors in parentheses are clustered at the county level. * p<0.1; ** p<0.05; *** p<0.01.

A.7.19. Robustness: a placebo test

Table A.37: Buddhist temples and fraction of family firms (OLS)

Dependent variable	Fraction of family firms			
	(1)	(2)	(3)	(4)
Buddhist temple density	-0.012	-0.013	-0.004	-0.002
	(0.009)	(0.010)	(0.016)	(0.015)
Geographic controls	No	Yes	Yes	Yes
Historical controls	No	No	Yes	Yes
Socioeconomic controls	No	No	No	Yes
Province FEs	Yes	Yes	Yes	Yes
Observations	204	204	204	204
R^2	0.232	0.238	0.245	0.250
Mean dependent vars.	0.497	0.497	0.497	0.497

Notes. County-level OLS regressions. Dependent variable is fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Buddhist temple density is number of Buddhist temples in 1820, normalized by 1953 population. Geographic controls include altitude, ruggedness, distance to coast and calorie suitability. Historical controls include biography density (1368-1949), population density in 1880 and tax per capita in 1820. Socioeconomic controls include sex ratio in 2010, minority rate in 2010, urbanization rate in 2010, bank density (1990-2010) and a provincial capital dummy. Province fixed effects are included. Genealogy density is standardized. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.