

The Cultural Origins of Family Firms*

Jian Xie[†]

University of Warwick

Song Yuan[‡]

University of Warwick

December, 2021

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 variable approach and the nearest neighbor matching method. Examining the 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 due to a lower leverage ratio.

Keywords: Family Culture; Family Firms; Genealogy; Cultural Origins; Firm Performance

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

*We are grateful to James Fenske and Daniel SgROI for their guidance and support. We thank Robert Akerlof, Jitendra Aswani, Manuel Bagues, Sascha Becker, Gustavo Cortes, Jana Fidrmuc, Bishnupriya Gupta, Sharun Mukand, Debin Ma, Carol Shiue, John Thanassoulis, and participants in the CAGE AMES seminar, Warwick Business School, LSE Economic History Graduate Seminar, OWL workshop and NEUDC conference for their helpful comments. All errors are our own.

[†]Department of Economics, University of Warwick. Email: j.xie.7@warwick.ac.uk

[‡]Department of Economics, University of Warwick. Email: s.yuan.2@warwick.ac.uk

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 2000 largest industrial firms retain founding family ownership (Anderson, Duru and Reeb, 2009). Among publicly listed international firms, 45% of them are family-controlled (La Porta, Lopez-de Silanes and Shleifer, 1999; Ellul, Pagano and Panunzi, 2010). A growing body of research has studied family firms on various topics such as their performance and behavior, and compared this to their non-family counterparts. Many implicitly assume the existence of family firms is exogenous. Despite the importance of family firms in economic development and in understanding the national economy, surprisingly, very few empirical works in economics have investigated the origin and spatial distribution of family firms.¹

We explore the role of culture, particularly historical family culture, in the prevalence of modern family firms in the context of China. Family culture, or family ties, is usually defined as the individual values that stress the link between family members and loyalty to the family (Alesina and Giuliano, 2014), which is documented to influence economic development and individual behaviors widely. However, family culture may also affect the distribution and behaviors of family firms. For instance, founders might derive utility from seeing relatives involved in the business and having their family members on the management board. Families with strong family ties are also more likely to have a will to build a family legacy and a desire to ensure survival and family control (Bertrand and Schoar, 2006).² Therefore, regions with a stronger family culture are more likely to have a larger share of family firms.

China is an ideal setting for studying this topic. First, due to socialist transformation, almost all current private firms were established after the reform and opening policy in 1978. This offers a good opportunity to understand the roots of variation of family firm prevalence in China since these firms have a relatively shorter history and family culture is predetermined. 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

¹Mueller and Philippon (2011) find that family firms are more prevalent in countries in which labor relations are hostile, based on around 30 observations.

²Bertrand and Schoar (2006) finds that family ties are positively correlated with establishment size and the fraction of self-employed. However, the fraction of family firms is regarded as the ideal measure of industry structure outcome, which does not exist across countries.

of family firm prevalence – the fraction of family firms – does not exist (or is not comparable) in many countries. Third, family firms are highly prevalent and play an essential role in the Chinese economy. For example, in China around 50% of private firms are family firms.³

To investigate the historical, cultural roots of the prevalence of family firms, we gather detailed data on firms from the Chinese Private Enterprise Survey (CPES) and historical genealogy at the county level, the smallest level possible in China. The analysis focuses on unlisted private firms, which make up the majority of the Chinese economy and face fewer strict regulations and have fewer outside investors than listed firms. In addition to the standard balance sheets, the firm data contains rich information on the resumes of the entrepreneurs and the family members’ engagement in firms in 2010. This allows us to plausibly distinguish between family⁴ 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.⁵ Specifically, we use historical genealogy density by normalizing the number of genealogy books using the population in 1953 to measure the historical family culture, which helps us address the reverse causality issue in the identification. Genealogy books describe family history and relationships among lineage members across and within generations. Genealogy reflects family culture because it involves living in an extended family, shows that people are interested in their family origins, and has practical functions such as providing public goods and organizing defence, which further strengthens intra-lineage member connections (Shiue, 2016). It serves the purposes of worshipping ancestors and uniting family members (Feng and Chang, 2008). 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 China Family Panel Survey (CFPS) in 2010.

³The number is calculated from Chinese Private Enterprise Survey in 2010.

⁴Our baseline definition for the family firm is that a family owns more than 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).

⁵The 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, Ma and Sinclair (2020) 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 the nuclear families dominate modern China.

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 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, which is based on the influence of Confucian academies during the Song dynasty (960AD-1279AD) within 200 km. These academies largely 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 the nearest neighbor matching (NNM) under the assumption 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 NNM estimation finds similar effects of historical culture as the baseline results. In addition, we have done extensive robustness checks using alternative definitions of family firms, alternative samples and alternative methods. For example, we exclude counties in provincial capitals and sub-provincial cities to address the concern of migration. The positive impact of historical family culture still holds. Furthermore, to document that other cultural factors do not drive our results, we conduct a placebo test by looking at the effects of historical presence of Buddhist temples and find no influence on contemporary family firms.

We test three mechanisms through which family ties might drive the presence and prevalence of family firms. In families with strong family 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%. 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, 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 also more willing to inherit the business.

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 firms in counties with a stronger historical family culture on average perform better. The mediation analysis indicates that a lower leverage ratio, which may improve the firm performance⁶ serves as one mechanism since firms in these regions are generally more risk averse and have less debt financing.

The analysis is extended to the listed firms by utilizing a panel dataset of all listed firms 2003-2019. We perform similar estimations as for the private firms. We find no significant impact of historical family culture on the share of family firms among listed firms. This is possible because outside investors can quickly obtain equity in or control over the firms through the stock market. Demand for professional managers might increase, and financing from family members becomes less important. In addition, public firms could be self-selected since firms in regions with a strong historical family culture may have fewer incentives to go public. Family owners could refuse to sell equity stakes to outsiders because they want to retain family management (Bloom and Van Reenen, 2007).

This research project is related to several strands of literature. The first 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 non-family counterparts (Anderson and Reeb, 2003; Bennedsen et al., 2007; Lins, Volpin and Wagner, 2013; Sraer and Thesmar, 2007). The results seem to be conditional on whether the founders are CEOs (Villalonga and Amit, 2006), whether current CEOs are related to founders (Mehrotra et al., 2013; Pérez-González, 2006; Bertrand et al., 2008) and the industries the family firms are in (Bennedsen et al., 2007). Amit and Villalonga (2014) have made a detailed summary of this literature. It has also been discovered that family firms are less tax aggressive (Chen et al., 2010), experience higher abnormal short

⁶See studies by Campello (2006); Dalci (2018); Lang, Ofek and Stulz (1996); Opler and Titman (1994)

sales prior to negative earnings shocks (Anderson, Reeb and Zhao, 2012), and invest less when inheritance law is stricter (Ellul, Pagano and Panunzi, 2010; Tsoutsoura, 2015). In addition, founders play more important roles in family firms than professional managers in non-family firms (Becker and Hvide, 2019). We contribute to this literature, first by investigating why some firms are initially established as family firms and the effect of family ties on the firm management in terms of shareholder structure and hiring decisions, and second, we provide evidence that historical family culture could partly explain firm performance.

The second strand of literature to which we contribute is on the impact of family culture. Alesina and Giuliano (2010) show that strong family ties are positively correlated with home production and lower labor force participation of women and young adults. There is also evidence that individuals with strong family ties are less likely to participate in politics (Alesina and Giuliano, 2011) and are less mobile and prefer more regulated labor markets (Alesina et al., 2015). Galasso and Profeta (2018) document that weak family culture facilitates the emergence of a pension system that acts as a flat safety net. Enke (2019) illustrates that kinship tightness affects cooperative behaviors, public goods provision and bilateral trade. Furthermore, family ties are shown to be negatively correlated with economic development (Bertrand and Schoar, 2006). The paper that is most closely related to ours is by Bertrand and Schoar (2006). They show that family ties are positively correlated with establishment size and the fraction of self-employed across countries. We contribute to this literature by systematically analysing the effect of family ties on family firms' presence by using the share of family firms as the measure of industry structure at a more disaggregated level.

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; Michalopoulos and Papaioannou, 2013; Padró i Miquel et al., 2015; Giavazzi, Schiantarelli and Serafinelli, 2013; Fan, Gu and Yu, 2021). Family firms perform and behave differently compared to non-family firms in many respects. We contribute to this literature by providing evidence that culture, specifically family ties, could affect economic outcomes by influencing the composition of firms and even directly affecting firm performance.

Our research also contributes more broadly to the literature on the understanding of the role of kin and kin-like relations in entrepreneurship, the development of the private economy and the success of market reforms in China. Existing studies mainly demonstrate how clan or kinship networks affect finance, private enterprise and mar-

ket transition (Peng, 2004; Nee, 1992; Zhang, 2020; Chen, Ma and Sinclair, 2020) or qualitatively document the role of family (Whyte, 1995, 1996) in economic development. We add to the literature by providing quantitative evidence that historical family culture may affect economic development through its influence on industry structure and firm management.

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 NNM estimation. In Section 5, we explore further robustness of our results. In Section 6, we discuss the channels through which family ties predict the presence and prevalence of family firms. In Section 7, we provide the firm performance implications of historical family culture and analysis on listed firms. Section 8 is the conclusion.

2 Background

2.1 Historical Family Culture

The traditional family in China is composed of parents, children, siblings, and other extended family members. It has been a fundamental component of the social structure since the Confucian period. Compared to western nuclear families, Chinese families in history have several distinguishing features. First, they are larger in scale, and co-residence of several generations is preferred. The typical historical Chinese family consists of some five-generations living together as one unit (Baker, 1979), which forms the foundation of family connections. As pointed out by Lv (2018), Song Confucians were particularly enthusiastic about promoting the co-residence of several generations. 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, 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 largely constructed 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, which emphasized and strengthened kinship bonds. For example, Zhang Zai's chapter, *The Law of the Clan* (1020AD-1077AD),⁷ 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. *Family Rituals* (1169AD),⁸ 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. Other famous scholars during this period, such as Lv Zuqian, Lu Jiuyuan, Ye Mengde, and Zhao Din, also compiled influential family clan regulations or guidelines⁹ to encourage filial piety and unite family members.

Second, Song Confucians developed various practical and standard methods and models to construct the family clan system at the local and grassroot levels. For instance, Ouyang Xiu and Su Xun created genealogy templates for commoners to trace family origins and record lineage members (Lv, 2018), which were widely spread in later dynasties. Zhang Zai and Cheng Yi initiated the construction of family temples and ancestral halls to maintain a harmonious family system and worship ancestors (Feng and Chang, 2008). Clansmen held important ceremonies, rites and sacrifices, and also discussed crucial issues there. In addition, Fan Zhongyan started a practice of preparing charitable clan estates to help family members who were in trouble (Twitchett, 1959).¹⁰ Some family clans also created benevolence halls, which functioned in similar ways. Furthermore, some Confucian scholars began to build academies to educate their descendants.¹¹ These academies helped them compete for educational resources and bureaucratic positions, which in turn fed back into the development of family clans.

⁷*Zong Fa*

⁸*Jia Li*

⁹Lv Zuqian's book: *Jia Fan*; Lu Jiuyuan's book: *Lu Shi Jia Gui*; Ye Mengde's book: *Shi Lin Jia Xun*; Zhao Din's book: *Jia Xun Bi Lu*

¹⁰According to Wang (2013), inspired by Fan, there were 68 charitable clan estates with clear information established by Confucian scholars during the Song dynasty. The majority of them were located in southeastern China.

¹¹Famous examples include Dongjia academy in Dean County, Hualin academy in Fengxin County and Leitang academy in Jianchang county (Wang, 2013)

Altogether, Song Confucians' theoretical guidance improved people's awareness of supporting family members, and these practices and methods played a role in building the family clan system. 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. Genealogies (*Jia Pu* or *Zu Pu*) are the records of family history and relationship between lineage members within and across generations. In China, the significant parts of genealogies are genealogical tables or charts demonstrating the relationship and pedigrees of lineage members organized by the birth order of each generation. They usually contain the basic information of each lineage member, such as the name, birth date, death date, marriage, and important honors. Many genealogies also include sections on lineage history and migration, ancestral halls, graves of prominent (or all) ancestors, biographies of key lineage members, and sometimes corporate landholdings (Harrell, 1987). The vast majority of Chinese genealogies are recorded in a patrilineal way.¹² Usually, genealogies are updated and held by the eldest member of the extended family, who passes on the responsibility to the next generation. Liu, Liu et al. (1959), Telford (1986) and Shiue (2016) provide more extensive discussions on the contents of genealogies.

The history of written genealogies could be traced back to the Han Dynasty (206 BC-220 AD) (Feng and Chang, 2008). Until the Song Dynasty (960AD-1279AD), preserved genealogies were mainly compiled by royal houses and political elites (Shiue, 2016), because genealogies were under the custody of the government and ordinary citizens were largely restricted in this regard. However, this regulation was relaxed during the Song Dynasty, and ordinary households began to compile their own genealogies without government intervention. By the Ming and Qing Dynas-

¹²There are no universal rules on how to record wives and sisters. In addition, they vary on formats and details regarding whether or how to record adopted sons and lineage members who did not reach adulthood.

ties (1368AD-1911AD), the genealogical tradition had spread widely among civilians (Harrell et al., 1995). A significant spatial variation could be observed in Figure A.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 defences, and other joint activities. It emphasizes the importance of the family, increases family interactions and strengthens family identities. As noted in *The Law of the Clan* (1020AD-1077AD),¹³ without genealogies, the extended families do not know their origins and cannot be kept together very long.

Some of the existing studies use genealogy as a measure of the strength of clan culture (Greif and Tabellini, 2017; Chen, Kung and Ma, 2017; Zhang, 2020). Conceptually, clans are typically regarded as a form of kinship organizations or kinship groups (Fei and Liu, 1982; Watson, 1982; Ebrey, Watson et al., 1986; Greif and Tabellini, 2010). Clans and extended families differ in generational coverage and scope but are similar in other aspects. Moreover, their boundary is ambiguous and debatable. 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 small extended families. 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 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 historical family culture exerts persistent influence on the contemporary family culture. The validation

¹³*Zong Fa* , by Zhang Zai

test is provided in Section 3.5.

2.3 Family Firms and Private Economy 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 (Böhren 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. Nepotism, for example, is a widely recognized feature of family firms that could impede their growth because the narrow kinship networks might only provide low-skilled and undiversified labor forces. 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.¹⁴ Because of family firms' coexisting merits and demerits, empirical studies of their performance demonstrate differential results. 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.

The development of family firms in China is strongly associated with the development of the private economy. 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. It is estimated that the average annual growth rate of the Chinese private sector was around 14.3% between 1996 to 2002 (Allen, Qian and Qian, 2005). In 2008, non-public sectors in China were worth 19 trillion RMB (approximately USD2.6 trillion) in total, contributing 65% of the GDP and accounting for 75% of the national employment and around half the tax revenue (Cao, Cumming

¹⁴Bloom and Van Reenen (2007) find that primogeniture leads to inferior performance for family firms.

and Wang, 2015). In 2019, the total employment of private sectors was around 230 million people.¹⁵

Claessens, Djankov and Lang (2000) think that the family firm is a norm in East Asia, where family culture and Confucianism are deeply rooted in the social environment. According to the Chinese Private Enterprise Survey, there were approximately 4.98 million private firms in China in 2006, around 73% of which were family firms.¹⁶ It is close to the estimation in Nicholson (2008), which pointed out that about 75% of all firms in most developed economies are family firms. 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 family firm dataset, approximately two-thirds are family firms.¹⁷ Renowned examples include Wanda Group, Hengda Group and Midea Group. This proportion is high compared to listed firms in other countries. According to La Porta, Lopez-de Silanes and Shleifer (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). Considering that family firms are usually small and medium-sized and there are plenty of state-owned listed firms, the proportion of family firms should only be higher among private sectors. 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.¹⁸ 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*.¹⁹ It slightly extends the number of genealogies and complements the information on the same genealogies, such as their locations and

¹⁵Data from <http://data.stats.gov>

¹⁶The fraction of family firms will change under different definitions, but it shows a considerably high fraction of firms are family firms.

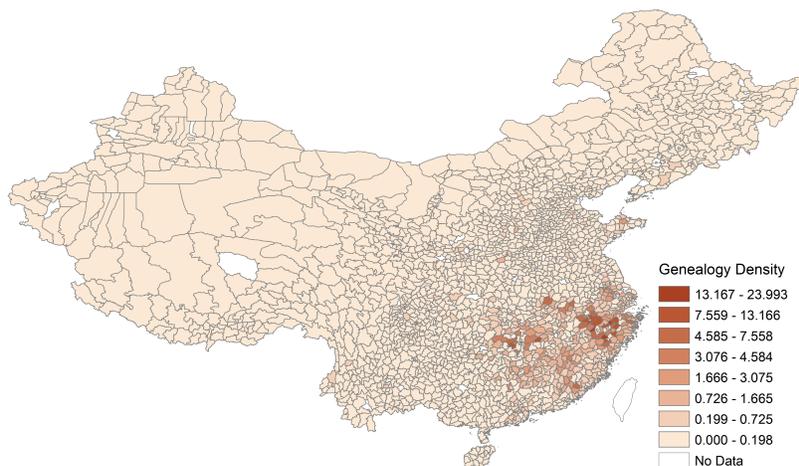
¹⁷The same as the above note.

¹⁸See <https://jiapu.library.sh.cn/#/>

¹⁹*Zhong Guo Jia Pu Zong Mu*, published in 2008 by Shanghai Library, which has been used in several studies by Chen, Kung and Ma (2017); Greif and Tabellini (2017); Zhang (2020); Chen, Ma and Sinclair (2020).

time. This online platform, in total, includes more than 54,000²⁰ Chinese genealogies originating from more than 1,600 places in China, covering 608 surnames, and it is the largest Chinese genealogy collection to date.²¹ 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 1368²² 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, we geo-localize genealogy books in counties, the smallest level possible. 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 A.1.

Figure 1: Genealogy Density in China



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 in China. It includes 13 waves and usually takes place every two years.²³ It is 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 Republic of China, and

²⁰See <https://jpv1.library.sh.cn/jp/page/about>

²¹The original *The General Catalog of Chinese Genealogy* covers 52401 genealogies and was the largest collection (Zhang, 2020).

²²Very few genealogies existed prior to 1368, the beginning of Ming Dynasty.

²³With one exception from 1997 to 2000.

(iv) the Private Economy Research Institute of China. With the backing of powerful government authorities, the survey’s response rate was consistently higher than 80%. Each wave drew 0.03% to 0.05% of all the private firms in China, covering all 31 provinces.

The survey involves intensive interviews with firm owners with questions about their personal experiences, attitudes, family backgrounds, and basic financial backgrounds of their firms (Li et al., 2008). Although questions regarding the personal experiences and financial background of the firms are consistent across waves, other questions could vary. Specifically, only the wave in 2010 contains detailed information on family engagement in firm ownership and management. For other waves, this sort of information is rather limited. As a result, for our research purpose, we only consider the survey conducted in 2010, which covers around 3000 firms with complete information in key variables.

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%²⁴ 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). This definition follows Bøhren et al. (2019) and Amore, Garofalo and Minichilli (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 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). 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.²⁵ In the end, there are 204 counties in our sample. We show

²⁴The threshold could be significantly lower for listed firms (Anderson and Reeb, 2003; Anderson, Reeb and Zhao, 2012; Ellul, Pagano and Panunzi, 2010).

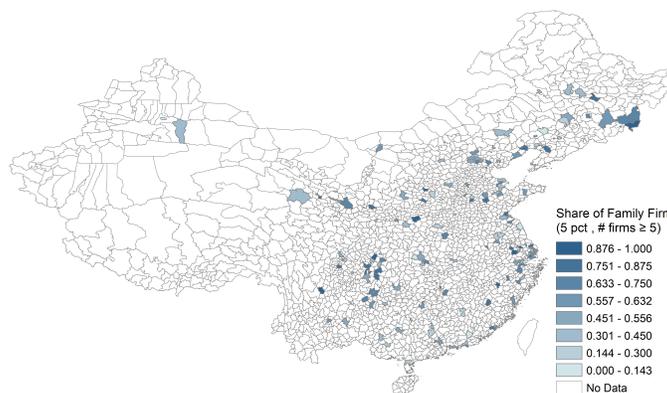
²⁵We also change the threshold of the number of firms in our robustness check.

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.

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)

3.3 China Family Panel Survey

To validate our proxy measure for historical family culture - genealogy density, we use the 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 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,²⁶ 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 2012.²⁷

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 questions²⁸ 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 (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 with very small sample size in our baseline estimation.

We also exclude the confounding effect of trust by directly testing the correlation between genealogy and trust and controlling for the measures of trust in our major validation test. We use two questions related to trust from the CFPS data in 2012: (1) How much do you trust strangers, (2) Do you in general trust or suspect others.

²⁶The sample of provinces excludes Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia and Hainan.

²⁷There are no questions related to trust in the survey in 2010.

²⁸These 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.

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.²⁹ 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 compiled 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).³⁰ 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 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 variables³¹ to avoid an unnecessary

²⁹See <https://bigdata.cgiar.org/srtm-90m-digital-elevation-database/>

³⁰See <https://projects.iq.harvard.edu/cbdb/home>

³¹The 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.

reduction in our sample size.³²

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.³³

For our IV, we collect information about Confucian academies during the Song dynasty (960AD-1279AD) from Bai (2012):³⁴ 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).³⁵ We collect it from the website of WorldMap.³⁶

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:

$$Y_{ij} = \alpha + \beta \text{Genealogy Density}_j + X'_{ij}\gamma + Z'_j\delta + \varepsilon_{ij} \quad (1)$$

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

³²Our sample size will decrease from 204 to 177 if we drop these missing values; however, the estimates remain highly similar.

³³See <http://xkz.cbirc.gov.cn/jr/>

³⁴493 Confucian academies during the Song dynasty with clear information.

³⁵It contains 2407 Buddhist sites with precise locations.

³⁶See http://worldmap.harvard.edu/data/geonode:buddhist_sites_1820

section 3.3. *Genealogy Density_j* is historical genealogy density (number of genealogies compiled during 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. 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.2 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 only pick up the difference in trust across individuals or regions. 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; Bugge and Durante, 2017), whereas the correlation is negative within half of the countries (Bertrand and Schoar, 2006). To address this issue, we first test the correlation between genealogy and the binary measure of general trust in the CFPS dataset: do you in general trust or suspect others? The estimation follows specification 1. The results are presented

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 | 0.016*** (0.005) | 0.015*** (0.004) | 0.018*** (0.005) | 0.014** (0.006) | 0.046*** (0.011) | 0.029** (0.013) |
| 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 | 0.021*** (0.006) | 0.015*** (0.005) | 0.404*** (0.103) | 0.382*** (0.101) | 0.059*** (0.011) | 0.046*** (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. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in Table A.3. We can observe that the coefficient of genealogy density is nearly zero and not significant, implying that there is little correlation between genealogy density and trust at the individual level. We further perform the robustness of the validation test by including the trust measure in all regressions in Table 1. Again, we find that the positive correlations between genealogy and all subjective and objective measures of family ties are very robust, as shown in Table A.4. The above tests indicate that genealogy density is a plausible measure or proxy variable of the strength of historical family culture in China.

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 nearest neighbor matching method 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 during 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:

$$Y_j = \alpha + \beta \text{Genealogy Density}_j + X_j' \gamma + \mu_p + \varepsilon_j \quad (2)$$

where Y_j is the fraction of family firms surveyed in county j ; $\text{Genealogy Density}_j$ is genealogy density in county j ; the coefficients 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.

The results are reported in Table 2. Genealogy density is standardized. Province fixed effects might partially account for some geographic and institutional variations. 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 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

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

| Dependent variable | Fraction of family firms | | | |
|------------------------|--------------------------|---------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.053*** (0.017) | 0.051*** (0.018) | 0.050** (0.019) | 0.052** (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 fraction of family firms in each county. We only consider counties with at least 5 firms surveyed. Genealogy density is number of genealogies during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

private sectors (Peng, 2004).

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 2,³⁷ which suggests very low spatial autocorrelation in our regression residuals.

4.2 Instrumental Variable Results

To alleviate the concern on 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. 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 family 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

³⁷The results are similar when we use residuals of other regressions in Table 2.

using the copies of their books or 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,³⁸ 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.

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.18 and Table A.19. 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.20.

³⁸Information summarized from Ji (1996) p.393-394, it can also be verified in Bai (2012) p.29

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

| <i>Second stage results</i> | | | | |
|-----------------------------|--------------------------|---------------------|---------------------|---------------------|
| Dependent variable | Fraction of family firms | | | |
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.230** (0.114) | 0.238** (0.117) | 0.222** (0.113) | 0.237** (0.115) |
| Mean dependent vars. | 0.497 | 0.497 | 0.497 | 0.497 |
| <i>First stage results</i> | | | | |
| Dependent variable | Genealogy density | | | |
| | (1) | (2) | (3) | (4) |
| Academy(within 200 km) | 0.374*** (0.106) | 0.391*** (0.117) | 0.414*** (0.119) | 0.417*** (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 during 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.

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

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 effects of historical family culture is larger compared to our OLS estimations. On the one hand, it is possible that the OLS estimations underestimate the effect of historical culture due to omitted variables or measurement errors; on the other hand, the IV might estimate the local average treatment effect of counties for which the family culture was affected by Song academies. However, in the context of accounting for measurement errors and omitted variable problems in genealogy density, the IV estimates deliver new insights regarding the effect size and direction.

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.5, the effects of Song academies on other contemporary outcome variables are very small and not significant. The results suggest that Song academies are unlikely to affect the contemporary fraction of family firms through channels other than historical family culture.

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 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 temples³⁹ 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.6, Song academies still strongly predict the distribution of Confucian temples in 2010. However, contempo-

³⁹Data from Fan (2004) and Kong and Kong (2011), manually verified in Google and Baidu.

rary 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 Nearest Neighbor Matching

To address the potential endogeneity, we also adopt the method of nearest neighbor matching.⁴⁰ 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 issue. Therefore, directly comparing the outcome of matched counties could improve the estimation since it eliminates these shared unobservable factors.⁴¹

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.⁴² 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:

⁴⁰The method is also used in studies such as [Huillery \(2009\)](#), [Voigtländer and Voth \(2012\)](#) and [Banerjee and Iyer \(2005\)](#).

⁴¹The argument is also demonstrated in [Card and Krueger \(2015\)](#).

⁴²The 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.

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

| Dependent variable | Fraction of family firms | | | |
|------------------------|--------------------------|---------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.062*** (0.017) | 0.077*** (0.013) | 0.063** (0.024) | 0.057** (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 nearest neighbor 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 during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

$$Y_j = \alpha + \beta \text{Genealogy Density}_j + X_j' \gamma + \mu_p + \theta_n + \varepsilon_j \quad (4)$$

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 together, some of our counties are dispersed and do not share borders with any of 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 mutual 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.

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.

5 Robustness Check

5.1 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, Garofalo and Minichilli, 2014). Referring to Villalonga and Amit (2006) and Ellul, Pagano and Panunzi (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.7 demonstrates that our results are insensitive to the shift of the thresholds with family engagement being guaranteed.

Table A.7 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.

5.2 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 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.8 and Table A.9). 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.10.

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.11 in the Appendix, the estimates become even more pronounced after excluding counties without any genealogies.

In addition, 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.12 demonstrates that the effects of genealogy density are even more pronounced in these relatively smaller cities with few migrants.

5.3 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.13. 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.14. 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.15. We also report standard errors clustered at the prefecture level and province level in the same table. Finally, we exclude the two outliers in genealogy density and use $\log(1+\text{genealogy density})$ to mitigate the bias from extreme values, as in Table A.16. In brief, all these alternative methods confirm our baseline results.

5.4 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, we amend the threshold to be 100km and 300 km as robustness checks, using the same equation. As one can see in Table A.18 and A.19, the estimates are robust. Additionally, we adopt the count of Confucian academies within 200 km as another robustness check in Table A.20. It suggests that our results are not driven by the particular formula which computes the influence of academies.

Our baseline NNM 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 a small sample size, we find each county's nearest neighbor to form a pair (not necessarily mutually nearest). It 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 for multiple counties. Table A.17 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.

5.5 A Placebo Test

To further address the concern that genealogies might reflect historical economic prosperity, we employ Buddhist sites density in 1820⁴³ as a placebo test. Although Buddhism was originated 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.21](#), historical Buddhism basically exerted no influence on contemporary family firms. Overall speaking, 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.

6 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. His-

⁴³Number of Buddhist temples and monasteries in 1820 in each county, normalized by 1953 population.

torical 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, Panunzi and Shleifer, 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.

6.1 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:

$$Y_{ij} = \alpha + \beta Genealogy\ Density_j + X'_{ij}\gamma + \mu_p + \varepsilon_{ij} \quad (5)$$

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 during 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. Specially, 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 5. 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.

6.2 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 that 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 5: 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*** (0.004) | 0.021*** (0.004) | 0.028*** (0.009) | 0.026*** (0.009) | 0.012*** (0.003) | 0.010*** (0.003) | 0.019** (0.009) | 0.017* (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 /Pseudo 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 that 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 number of genealogies during 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 6. 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. Further, the results imply 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.

Table 6: 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* (0.007) | 0.017*** (0.006) | -0.015* (0.008) | -0.023*** (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 during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6.3 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). 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 7, historical family culture is positively related to willingness to inherit by children and the tendency to pass on the firms to offsprings by founders. Specifically, in column (2) with full controls, a one standard deviation increase in genealogy density is associated with a 1.4% increase in the willingness to inherit firms from parents. Column (4) suggests that founders' tendency to pass on the firms to their children increases by 1.9% 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.

Table 7: Mechanism: inheritance (Probit, marginal effects)

| Dependent variable | Willing to inherit | | Plan to pass on | |
|------------------------|--------------------|-------------------|------------------|-------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.013** (0.006) | 0.014* (0.008) | 0.015 (0.010) | 0.019* (0.011) |
| 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 | 1677 | 1677 | 1828 | 1828 |
| Pseudo R^2 | 0.141 | 0.147 | 0.103 | 0.109 |
| 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 during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

7 Firm Performance and Additional Results on Listed Firms

In this section, we demonstrate the private firm performance implications of historical family culture and analyze its role in listed firms 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 the firm performance. Furthermore, we explore the role of family culture on the share of listed firms and find no statistically significant correlation.

7.1 Historical Family Culture and Firm Performance

The findings in previous sections reveal the effects of historical family clans on the share of family firms and the underlying mechanism. Unlike past studies directly comparing the performance difference between family and non-family firms, we explore the deep role of culture – specifically family culture – in the performance of firms. The empirical framework we adopt is the same as a 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 8 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 for a positive impact of genealogy density 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.

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.⁴⁴ One mechanism we exploit here is that firms in regions with 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 the performance of firms. In addition, as discussed in Section 6.2, firms in regions with a stronger family culture raise a higher share of initial capital from family members and lower share from banks. We would, therefore, expect

⁴⁴It also depends on the level of creditor monitoring.

Table 8: Firm performance and the underlying mechanism

| Dependent variable | ROA | | Leverage | | ROA | |
|------------------------|---------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genealogy density | 0.073*** (0.018) | 0.087*** (0.020) | -0.075** (0.031) | -0.079*** (0.020) | 0.062*** (0.015) | 0.075*** (0.017) |
| Leverage | | | | | -0.210*** (0.038) | -0.207*** (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 during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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 the level of debt in a firm. We regress *Leverage* on genealogy density with and without controlling for historical and socioeconomic variables separately.

The results are displayed in columns (3) and (4) in Table 8. 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 8. 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%. This suggests 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 compared with professional managers because their interests are more aligned with firm heads (Cai et al., 2013).

7.2 Additional Results 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 county level.⁴⁵ 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.23 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.⁴⁶ The sub-sample 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

⁴⁵The 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.

⁴⁶Notice 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.

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 6.1. The results show that there is no consistently positive impact of genealogy density on all four measures of family engagement in listed family firms.⁴⁷

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 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 by decreasing the leverage ratio. The effects of historical family culture are muted in listed firms because outside investors could more easily obtain equity shares through the stock market, and the self-selection of going public for firms.

Our study inspires topics of general interests, even though it is based on data from China. First, there are deep historical roots for contemporary industry structure in terms of the fraction of family firms. This implies differential prevalence levels of family firms in countries emphasizing or dampening family culture, and different probabilities of establishing family firms for immigrants to the same country from various cultural backgrounds. Second, historical family culture could be another channel for firm creation in the absence of modern financial systems and regulations.

⁴⁷See Appendix Table A.24

Third, family culture influences the management of firms in various respects due to entrepreneurs' family and culture concerns, indicating that culture not only shapes outcomes of countries and individuals, but also deeply affects that of firms. This has deeper implications for economic development.

References

- Alesina, Alberto and Paola Giuliano. 2010. “The power of the family.” *Journal of Economic Growth* 15(2):93–125.
- Alesina, Alberto and Paola Giuliano. 2011. “Family ties and political participation.” *Journal of the European Economic Association* 9(5):817–839.
- Alesina, Alberto and Paola Giuliano. 2014. Family ties. In *Handbook of Economic Growth*. Vol. 2 Elsevier pp. 177–215.
- Alesina, Alberto, Yann Algan, Pierre Cahuc and Paola Giuliano. 2015. “Family values and the regulation of labor.” *Journal of the European Economic Association* 13(4):599–630.
- Algan, Yann and Pierre Cahuc. 2010. “Inherited trust and growth.” *American Economic Review* 100(5):2060–92.
- Allen, Franklin, Jun Qian and Meijun Qian. 2005. “Law, finance, and economic growth in China.” *Journal of Financial Economics* 77(1):57–116.
- Altonji, Joseph G and Charles R Pierret. 2001. “Employer learning and statistical discrimination.” *The Quarterly Journal of Economics* 116(1):313–350.
- Amit, Raphael and Belén Villalonga. 2014. “Financial performance of family firms.” *The SAGE Handbook of Family Business* pp. 157–178.
- Amore, Mario Daniele, Orsola Garofalo and Alessandro Minichilli. 2014. “Gender interactions within the family firm.” *Management Science* 60(5):1083–1097.
- Anderson, Ronald C, Augustine Duru and David M Reeb. 2009. “Founders, heirs, and corporate opacity in the United States.” *Journal of Financial Economics* 92(2):205–222.
- Anderson, Ronald C and David M Reeb. 2003. “Founding-family ownership and firm performance: evidence from the S&P 500.” *The Journal of Finance* 58(3):1301–1328.
- Anderson, Ronald C, David M Reeb and Wanli Zhao. 2012. “Family-controlled firms and informed trading: Evidence from short sales.” *The Journal of Finance* 67(1):351–385.

- Ayres, Glenn R. 1990. "Rough family justice: Equity in family business succession planning." *Family Business Review* 3(1):3–22.
- Bai, Xinliang. 2012. *The study of academies in the Ming and Qing dynasties (Mingqing Shuyuan Yanjiu)*. The Forbidden City Press (Beijing).
- Bai, Ying and Ruixue Jia. 2016. "Elite recruitment and political stability: the impact of the abolition of China's civil service exam." *Econometrica* 84(2):677–733.
- Baker, Hugh DR. 1979. *Chinese family and kinship*. Columbia University Press.
- Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, institutions, and economic performance: The legacy of colonial land tenure systems in India." *American Economic Review* 95(4):1190–1213.
- Becker, Sascha O and Hans K Hvide. 2019. "Entrepreneur death and startup performance." *Available at SSRN 2199645* .
- Bennedsen, Morten, Kasper Meisner Nielsen, Francisco Pérez-González and Daniel Wolfenzon. 2007. "Inside the family firm: The role of families in succession decisions and performance." *The Quarterly Journal of Economics* 122(2):647–691.
- Bertrand, Marianne and Antoinette Schoar. 2006. "The role of family in family firms." *Journal of Economic Perspectives* 20(2):73–96.
- Bertrand, Marianne, Simon Johnson, Krislert Samphantharak and Antoinette Schoar. 2008. "Mixing family with business: A study of Thai business groups and the families behind them." *Journal of Financial Economics* 88(3):466–498.
- Bian, Yanjie. 1994. "Guanxi and the allocation of urban jobs in China." *The China Quarterly* 140:971–999.
- Biao, YOU. 2007. "Clan, ancestral hall, and sacrifice in the Song Dynasty." *Frontiers of History in China* 2(2):166–180.
- Bloom, Nicholas and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The Quarterly Journal of Economics* 122(4):1351–1408.
- Bøhren, Øyvind, Bogdan Stacescu, Line F Almlı and Kathrine L Søndergaard. 2019. "When does the family govern the family firm?" *Journal of Financial and Quantitative Analysis* 54(5):2085–2117.

- Bol, Peter Kees. 2008. *Neo-Confucianism in history*. Harvard University Press.
- Buggle, Johannes and Ruben Durante. 2017. “Climate risk, cooperation, and the co-evolution of culture and institutions.”
- Burkart, Mike, Fausto Panunzi and Andrei Shleifer. 2003. “Family firms.” *The Journal of Finance* 58(5):2167–2201.
- Cai, Hongbin, Hongbin Li, Albert Park and Li-An Zhou. 2013. “Family ties and organizational design: Evidence from Chinese private firms.” *Review of Economics and Statistics* 95(3):850–867.
- Campello, Murillo. 2006. “Debt financing: Does it boost or hurt firm performance in product markets?” *Journal of Financial Economics* 82(1):135–172.
- Cao, Jerry, Douglas Cumming and Xiaoming Wang. 2015. “One-child policy and family firms in China.” *Journal of Corporate Finance* 33:317–329.
- Card, David and Alan B Krueger. 2015. *Myth and Measurement: The New Economics of the Minimum Wage-Twentieth-Anniversary Edition*. Princeton University Press.
- Chen, Shuping, Xia Chen, Qiang Cheng and Terry Shevlin. 2010. “Are family firms more tax aggressive than non-family firms?” *Journal of Financial Economics* 95(1):41–61.
- Chen, Ting, James Kai-sing Kung and Chicheng Ma. 2017. “Long live Keju! The persistent effects of China’s imperial examination system.” *The Persistent Effects of China’s Imperial Examination System (June 26, 2017)* .
- Chen, Zhiwu, Chicheng Ma and Andrew Sinclair. 2020. “Banking on the Confucian Clan: Why Did China Miss the Financial Revolution?” *Available at SSRN 3671280* .
- Claessens, Stijn, Simeon Djankov, Joseph PH Fan and Larry HP Lang. 2002. “Disentangling the incentive and entrenchment effects of large shareholdings.” *The Journal of Finance* 57(6):2741–2771.
- Claessens, Stijn, Simeon Djankov and Larry HP Lang. 2000. “The separation of ownership and control in East Asian corporations.” *Journal of Financial Economics* 58(1-2):81–112.

- Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92(1):1–45.
- Dalci, Ilhan. 2018. "Impact of financial leverage on profitability of listed manufacturing firms in China." *Pacific Accounting Review* .
- Davis, John A, Marion McCollom Hampton, Ivan Lansberg et al. 1997. *Generation to generation: Life cycles of the family business*. Harvard Business Press.
- Doepke, Matthias and Fabrizio Zilibotti. 2008. "Occupational choice and the spirit of capitalism." *The Quarterly Journal of Economics* 123(2):747–793.
- Ebrey, Patricia Buckley, James L Watson et al. 1986. *Kinship Organization in Late Imperial China, 1000-1940*. Vol. 5 Univ of California Press.
- Ellul, Andrew, Marco Pagano and Fausto Panunzi. 2010. "Inheritance law and investment in family firms." *American Economic Review* 100(5):2414–50.
- Enke, Benjamin. 2019. "Kinship, cooperation, and the evolution of moral systems." *The Quarterly Journal of Economics* 134(2):953–1019.
- Fan, Joseph PH, Qiankun Gu and Xin Yu. 2021. "Collectivist Cultures and the Emergence of Family Firms." *Forthcoming in Journal of Law and Economics* .
- Fan, Xiaoping. 2004. *Confucian temples in China (Zhongguo Kongmiao)*. Sichuang Wenyi Chubanshe.
- Fei, John CH and Tsui-Jung Liu. 1982. "The growth and decline of Chinese family clans." *Journal of Interdisciplinary History* pp. 375–408.
- Feng, Erkang and Jianhua Chang. 2008. *The history of Chinese Clans (Zhongguo Zongzu Shi)* . Shanghai Renmin Chubanshe (Beijing).
- Fernández, Raquel and Alessandra Fogli. 2009. "Culture: An empirical investigation of beliefs, work, and fertility." *American Economic Journal: Macroeconomics* 1(1):146–77.
- Fincher, Corey L, Randy Thornhill, Damian R Murray and Mark Schaller. 2008. "Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism." *Proceedings of the Royal Society B: Biological Sciences* 275(1640):1279–1285.

- Galasso, Vincenzo and Paola Profeta. 2018. "When the state mirrors the family: the design of pension systems." *Journal of the European Economic Association* 16(6):1712–1763.
- Galor, Oded and Ömer Özak. 2016. "The agricultural origins of time preference." *American Economic Review* 106(10):3064–3103.
- Gernet, Jacques. 1995. *Buddhism in Chinese society: an economic history from the fifth to the tenth centuries*. Columbia University Press.
- Gethin, Rupert et al. 1998. *The foundations of Buddhism*. Oxford University Press.
- Giavazzi, Francesco, Fabio Schiantarelli and Michel Serafinelli. 2013. "Attitudes, policies, and work." *Journal of the European Economic Association* 11(6):1256–1289.
- Giuliano, Paola. 2007. "Living arrangements in western europe: Does cultural origin matter?" *Journal of the European Economic Association* 5(5):927–952.
- Gómez-Mejía, Luis R, Katalin Takács Haynes, Manuel Núñez-Nickel, Kathryn JL Jacobson and José Moyano-Fuentes. 2007. "Socioemotional wealth and business risks in family-controlled firms: Evidence from Spanish olive oil mills." *Administrative Science Quarterly* 52(1):106–137.
- Gowdy, John and Lisi Krall. 2016. "The economic origins of ultrasociality." *Behavioral and Brain Sciences* 39.
- Greif, Avner and Guido Tabellini. 2010. "Cultural and institutional bifurcation: China and Europe compared." *American Economic Review* 100(2):135–40.
- Greif, Avner and Guido Tabellini. 2017. "The clan and the corporation: Sustaining cooperation in China and Europe." *Journal of Comparative Economics* 45(1):1–35.
- Harrell, Stevan. 1987. "On the holes in Chinese genealogies." *Late Imperial China* 8(2):53–79.
- Harrell, Stevan et al. 1995. *Chinese historical microdemography*. Vol. 20 Univ of California Press.
- Huillery, Elise. 2009. "History matters: The long-term impact of colonial public investments in French West Africa." *American Economic Journal: Applied Economics* 1(2):176–215.

- Jensen, Michael C and William H Meckling. 1976. "Theory of the firm: Managerial behavior, agency costs and ownership structure." *Journal of Financial Economics* 3(4):305–360.
- Ji, Xiaofeng. 1996. *A dictionary of academies in China (Zhongguo Shuyuan Cidian)*. Zhejiang Educational Press.
- Kelly, Morgan. 2019. "The standard errors of persistence."
- Kong, Xianglin and Zhe Kong. 2011. *The study of Confucian temples in the world (Shijie Kongzimiao Yanjiu)*. Zhongyang Bianyi Chubanshe.
- La Porta, Rafael, Florencio Lopez-de Silanes and Andrei Shleifer. 1999. "Corporate ownership around the world." *The Journal of Finance* 54(2):471–517.
- Lang, Larry, Eli Ofek and RenéM Stulz. 1996. "Leverage, investment, and firm growth." *Journal of Financial Economics* 40(1):3–29.
- Li, Hongbin, Lingsheng Meng, Qian Wang and Li-An Zhou. 2008. "Political connections, financing and firm performance: Evidence from Chinese private firms." *Journal of Development Economics* 87(2):283–299.
- Lins, Karl V, Paolo Volpin and Hannes F Wagner. 2013. "Does family control matter? International evidence from the 2008–2009 financial crisis." *The Review of Financial Studies* 26(10):2583–2619.
- Liu, Hui-chen Wang, Wang Liu et al. 1959. *The traditional Chinese clan rules*. Vol. 7 JJ Augustin.
- Lv, Simian. 2018. *A brief of Chinese clan institutions (Zhongguo Zongzu Zhidu Xiaoshi)*. Zhishi Chanquan Chubanshe (Beijing).
- Mehrotra, Vikas, Randall Morck, Jungwook Shim and Yupana Wiwattanakantang. 2013. "Adoptive expectations: Rising sons in Japanese family firms." *Journal of Financial Economics* 108(3):840–854.
- Michalopoulos, Stelios and Elias Papaioannou. 2013. "Pre-colonial ethnic institutions and contemporary African development." *Econometrica* 81(1):113–152.
- Miller, Danny, Isabelle Le Breton-Miller, Richard H Lester and Albert A Cannella Jr. 2007. "Are family firms really superior performers?" *Journal of Corporate Finance* 13(5):829–858.

- Mueller, Holger M and Thomas Philippon. 2011. "Family firms and labor relations." *American Economic Journal: Macroeconomics* 3(2):218–45.
- Nee, Victor. 1992. "Organizational dynamics of market transition: Hybrid forms, property rights, and mixed economy in China." *Administrative Science Quarterly* pp. 1–27.
- Nicholson, N. 2008. "Evolutionary psychology, corporate culture and family business." *Academy of Management Perspectives* 22:73–84.
- Nunn, Nathan and Diego Puga. 2012. "Ruggedness: The blessing of bad geography in Africa." *Review of Economics and Statistics* 94(1):20–36.
- Opler, Tim C and Sheridan Titman. 1994. "Financial distress and corporate performance." *The Journal of Finance* 49(3):1015–1040.
- Padró i Miquel, Gerard, Nancy Qian, Yiqing Xu and Yang Yao. 2015. "Making democracy work: Culture, social capital and elections in china." *Social Capital and Elections in China (April 2015)*. NBER Working Paper (w21058).
- Peng, Yusheng. 2004. "Kinship networks and entrepreneurs in China's transitional economy." *American Journal of Sociology* 109(5):1045–1074.
- Pérez-González, Francisco. 2006. "Inherited control and firm performance." *American Economic Review* 96(5):1559–1588.
- Robb, Alicia M and David T Robinson. 2014. "The capital structure decisions of new firms." *The Review of Financial Studies* 27(1):153–179.
- Schmid, Thomas. 2013. "Control considerations, creditor monitoring, and the capital structure of family firms." *Journal of Banking & Finance* 37(2):257–272.
- Shen, Na and Jun Su. 2017. "Religion and succession intention-Evidence from Chinese family firms." *Journal of Corporate Finance* 45:150–161.
- Shiue, Carol H. 2016. "A culture of kinship: Chinese genealogies as a source for research in demographic economics." *Journal of Demographic Economics* 82(4):459–482.
- Sraer, David and David Thesmar. 2007. "Performance and behavior of family firms: Evidence from the French stock market." *Journal of the European Economic Association* 5(4):709–751.

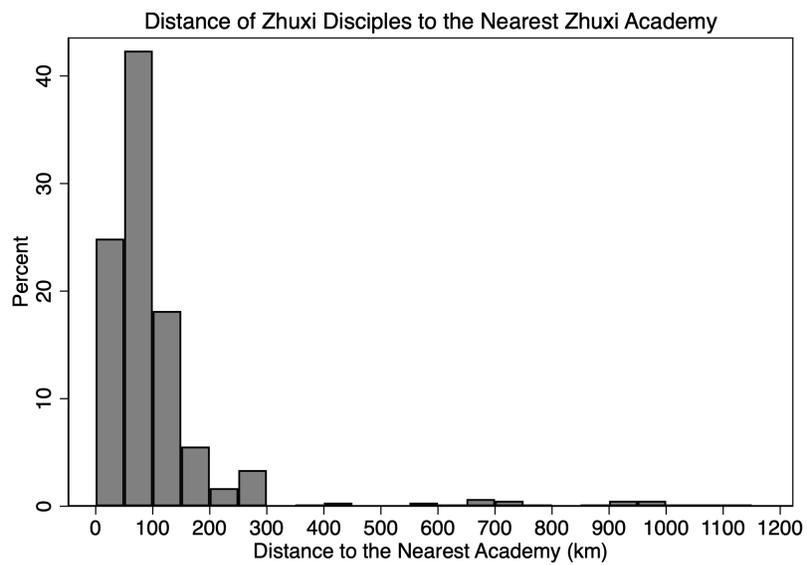
- Tabellini, Guido. 2010. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic Association* 8(4):677–716.
- Telford, Ted A. 1986. "Survey of social demographic data in Chinese genealogies." *Late Imperial China* 7(2):118–148.
- Tsoutsoura, Margarita. 2015. "The effect of succession taxes on family firm investment: Evidence from a natural experiment." *The Journal of Finance* 70(2):649–688.
- Twitchett, Denis. 1959. "The Fan clan's charitable estate, 1050-1760." *Confucianism in Action* 97:133.
- Villalonga, Belen and Raphael Amit. 2006. "How do family ownership, control and management affect firm value?" *Journal of Financial Economics* 80(2):385–417.
- Voigtländer, Nico and Hans-Joachim Voth. 2012. "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany." *The Quarterly Journal of Economics* 127(3):1339–1392.
- Wang, Shanjun. 2013. "The spatial variation of Song clans and its reasons." *Anhui Shixue* 1.
- Watson, James L. 1982. "Chinese kinship reconsidered: Anthropological perspectives on historical research." *The China Quarterly* 92:589–622.
- Whyte, Martin King. 1995. "The social roots of China's economic development." *The China Quarterly* 144:999–1019.
- Whyte, Martin King. 1996. "The Chinese family and economic development: Obstacle or engine?" *Economic Development and Cultural Change* 45(1):1–30.
- Zhang, Chuanchuan. 2020. "Clans, entrepreneurship, and development of the private sector in China." *Journal of Comparative Economics* 48(1):100–123.

Appendices

Appendix A Appendix

A.1 Figures

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



A.2 Summary Statistics

Table A.1: 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 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 |

A.3 Validation of Genealogy

Table A.2: 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 book | 0.009 (0.011) | 0.020* (0.011) | 0.027** (0.012) | 0.024* (0.013) | 0.069** (0.027) | 0.055** (0.026) |
| Observations | 33289 | 33285 | 33289 | 33285 | 30632 | 30627 |
| R^2 | 0.000 | 0.152 | 0.000 | 0.154 | 0.001 | 0.427 |
| Dependent variable | Visit grave last year | | # Relatives visit spring festival | | Principal component of family ties | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genealogy book | 0.066*** (0.010) | 0.060*** (0.010) | 1.696*** (0.158) | 0.921*** (0.153) | 0.104*** (0.025) | 0.096*** (0.026) |
| Observations | 33281 | 33277 | 33099 | 33096 | 33293 | 33289 |
| R^2 | 0.004 | 0.344 | 0.011 | 0.321 | 0.001 | 0.223 |

Notes. OLS regressions. 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.3: Genealogy and Trust: CFPS Data

| Dependent variable | Trust | |
|--------------------|------------------|------------------|
| | (1) | (2) |
| Genealogy density | 0.001 (0.005) | 0.005 (0.005) |
| Controls | No | Yes |
| Observations | 22196 | 21566 |
| R^2 | 0.000 | 0.038 |

Notes. OLS regressions. 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.4: Validation of Genealogy: Controlling for Trust

| Dependent variable | Happy family important | | Child. having achieve. important | | Family relation. are close | |
|--------------------|------------------------|---------------------|-----------------------------------|---------------------|------------------------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genealogy density | 0.014*** (0.005) | 0.012*** (0.005) | 0.016*** (0.006) | 0.012* (0.007) | 0.048*** (0.011) | 0.027* (0.014) |
| Trust | 0.029** (0.013) | 0.001 (0.011) | -0.020 (0.015) | -0.016 (0.013) | 0.104*** (0.029) | 0.056** (0.025) |
| Controls | No | Yes | No | Yes | No | Yes |
| Observations | 22194 | 21564 | 22194 | 21564 | 20383 | 19816 |
| R^2 | 0.001 | 0.046 | 0.001 | 0.052 | 0.006 | 0.096 |
| Dependent variable | Visit grave last year | | # Relatives visit spring festival | | Principal component of family ties | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genealogy density | 0.020*** (0.006) | 0.013** (0.005) | 0.391*** (0.099) | 0.373*** (0.100) | 0.054*** (0.012) | 0.041*** (0.015) |
| Trust | -0.015 (0.011) | -0.006 (0.009) | 0.388*** (0.144) | 0.333** (0.132) | 0.071** (0.028) | 0.030 (0.022) |
| Controls | No | Yes | No | Yes | No | Yes |
| Observations | 22185 | 21555 | 22079 | 21449 | 22196 | 21566 |
| R^2 | 0.005 | 0.068 | 0.008 | 0.086 | 0.004 | 0.074 |

Notes. OLS regressions. 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.4 IV: Exclusion Restrictions

Table A.5: 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 (3.449) | -2.434 (2.970) | 0.207 (0.187) | 0.072 (0.075) | 1.759* (1.020) | 1.395 (1.117) | 0.597 (0.507) | 0.481 (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.6: IV: exclusion restrictions

| Dependent variable | Confucianism in 2010 | | Fraction of family firms | |
|------------------------|-----------------------|---------------------|--------------------------|------------------|
| | (1) | (2) | (3) | (4) |
| Academy (200 km) | 317.7*** (111.741) | 215.6** (96.746) | | |
| Confucianism in 2010 | | | 0.031 (0.025) | 0.024 (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.5 Robustness: Alternative Definitions of Family Firms

Table A.7: 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*** (0.018) | 0.047*** (0.018) | 0.046** (0.019) | 0.048** (0.021) |
| Dependent variable | Fraction of family firms (20% share + family engage.) | | | |
| Genealogy density | 0.050*** (0.018) | 0.048** (0.019) | 0.047** (0.020) | 0.049** (0.021) |
| Dependent variable | Fraction of family firms (33.3% share + family engage.) | | | |
| Genealogy density | 0.051*** (0.018) | 0.049*** (0.018) | 0.048** (0.020) | 0.050** (0.021) |
| Dependent variable | Fraction of family firms (66.7% share + family engage.) | | | |
| Genealogy density | 0.052*** (0.015) | 0.050*** (0.016) | 0.050*** (0.016) | 0.052*** (0.018) |
| Dependent variable | Fraction of family firms (50% share) | | | |
| Genealogy density | 0.018** (0.009) | 0.017* (0.010) | 0.016* (0.009) | 0.015 (0.010) |
| Dependent variable | Fraction of family firms (66.7% share) | | | |
| Genealogy density | 0.027** (0.012) | 0.026** (0.012) | 0.026** (0.012) | 0.024* (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 during 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.6 Robustness: Alternative Samples

Table A.8: 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** (0.023) | 0.049** (0.024) | 0.056** (0.023) | 0.055** (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 during 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.9: 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*** (0.013) | 0.040*** (0.014) | 0.042*** (0.014) | 0.049*** (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 during 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.10: Genealogy density and the distribution of family firms (Probit, marginal effects)

| Dependent variable | Family firm dummy | | | |
|------------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.039*** (0.010) | 0.039*** (0.010) | 0.040*** (0.010) | 0.044*** (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 during 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.11: 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*** (0.020) | 0.063*** (0.021) | 0.061*** (0.019) | 0.071*** (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 during 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.12: 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*** (0.019) | 0.059*** (0.020) | 0.061*** (0.022) | 0.066*** (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 during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

A.7 Robustness: Alternative Methods

Table A.13: Genealogy density and fraction of family firms (Tobit)

| Dependent variable | Fraction of family firms | | | |
|------------------------|--------------------------|--------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.054** (0.024) | 0.052** (0.024) | 0.051** (0.024) | 0.053** (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 during 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.14: Genealogy density and fraction of family firms (dropping province FEs)

| Dependent variable | Fraction of family firms | | | |
|------------------------|--------------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.043*** (0.011) | 0.038*** (0.012) | 0.039*** (0.012) | 0.039*** (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 during 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.15: 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 during 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.16: Genealogy density and fraction of family firms (Take log)

| Dependent variable | Fraction of family firms | | | |
|---------------------------|--------------------------|---------------------|--------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| log (Genealogy density+1) | 0.049*** (0.018) | 0.049*** (0.018) | 0.049** (0.019) | 0.057*** (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 | 202 | 202 | 202 | 202 |
| R^2 | 0.255 | 0.260 | 0.269 | 0.279 |
| Mean dependent vars. | 0.495 | 0.495 | 0.495 | 0.495 |

Notes. County-level OLS regressions. The two extreme values of genealogy density are removed. 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 during 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.8 Robustness: Alternative Nearest Neighbor Matching

Table A.17: Genealogy density and fraction of family firms (NNM)

| Dependent variable | Fraction of family firms | | | |
|------------------------|--------------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| Genealogy density | 0.046*** (0.015) | 0.046*** (0.016) | 0.041*** (0.015) | 0.038*** (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 |
| Observations | 408 | 408 | 408 | 408 |
| R^2 | 0.606 | 0.608 | 0.616 | 0.640 |
| Mean dependent vars. | 0.466 | 0.466 | 0.466 | 0.466 |

Notes. County-level nearest neighbor matching. For each county, we find its nearest county and match them to be a pair (As a result, a county could appear several times in the sample). 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 during 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.9 Robustness: IV

Table A.18: Genealogy density and fraction of family firms (IV)

| <i>Second stage results</i> | | | | |
|-----------------------------|--------------------------|----------|----------|----------|
| 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 |
| <i>First stage results</i> | | | | |
| 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 during 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.19: Genealogy density and fraction of family firms (IV)

| <i>Second stage results</i> | | | | |
|-----------------------------|--------------------------|----------|----------|----------|
| 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 |
| <i>First stage results</i> | | | | |
| 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 during 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.20: Genealogy density and fraction of family firms (IV)

| <i>Second stage results</i> | | | | |
|--------------------------------|--------------------------|----------|----------|----------|
| 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 |
| <i>First stage results</i> | | | | |
| Dependent variable | Genealogy density | | | |
| | (1) | (2) | (3) | (4) |
| N.O. 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 during 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.10 Robustness: A Placebo Test

Table A.21: Buddhist temples and fraction of family firms (OLS)

| Dependent variable | Fraction of family firms | | | |
|-------------------------|--------------------------|-------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) |
| Buddhist temple density | -0.012 (0.009) | -0.013 (0.010) | -0.004 (0.016) | -0.002 (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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.22: Mechanisms: Distrust in Professional CEOs

| Dependent variable | Distrust professional CEOs | | |
|------------------------|----------------------------|-------------------|-------------------|
| | (1) | (2) | (3) |
| Genealogy density | -0.056* (0.033) | -0.051 (0.036) | -0.036 (0.039) |
| Individual controls | No | Yes | Yes |
| Firm controls | No | Yes | Yes |
| Geographic controls | No | Yes | Yes |
| Historical controls | No | No | Yes |
| Socioeconomic controls | No | No | Yes |
| Province FEs | Yes | Yes | Yes |
| Observations | 3088 | 2697 | 2691 |
| Pseudo R^2 | 0.037 | 0.044 | 0.050 |
| Mean dependent vars. | 0.207 | 0.206 | 0.205 |

Notes. Firm-level probit regressions (marginal effects reported). Dependent variable is a dummy equaling to 1 if the owner reports distrust in professional CEOs. Genealogy density is number of genealogies during 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.23: Genealogy Density and Fraction of Listed Family Firms

| Dependent variable | Fraction of Family Firms | | | |
|---|--------------------------|-------------------|-------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| <i>Panel A: Full sample</i> | | | | |
| Genealogy density | 0.015 (0.012) | 0.018 (0.011) | 0.015 (0.011) | -0.009 (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 |
| <i>Panel B: Excluding provincial capitals and sub-provincial cities</i> | | | | |
| Genealogy density | -0.004 (0.009) | -0.007 (0.009) | -0.012 (0.010) | -0.020** (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 during 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.24: Mechanisms: Family Engagement in the Listed Firm

| 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 (0.004) | -0.006 (0.004) | -0.002 (0.003) | -0.001 (0.003) | -0.007 (0.005) | -0.009** (0.004) | -0.004 (0.006) | -0.007 (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, supervisor, managers who are family members. The last dependent variable is the ratio of salary of family members to the total salary. Genealogy density is number of genealogies during 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.