**Clan Culture and Patterns of Industrial Specialization in China**

**Haichao Fan**

Institute of World Economy, Fudan University, Shanghai, China

Email: fan\_haichao@fudan.edu.cn

**Chang Li\***

School of Economics, East China Normal University, Shanghai, China

Email: cli@fem.ecnu.edu.cn

**Chang Xue**

Economics and Management School, Wuhan University, Wuhan, China

Email: xuechang@whu.edu.cn

**Miaojie Yu**

Faculty of Economics, Liaoning University, Shenyang, China

Email: mjyu@lnu.edu.cn

**Acknowledgments**

Haichao Fan acknowledges the financial support by the Shanghai Institute of International Finance and Economics; the Innovative Research Groups Project of the National Natural Science Foundation of China (no. 72121002). Chang Li acknowledges the financial support from the National Natural Science Foundation of China (no. 71803048). Chang Xue acknowledges the financial support from the National Natural Science Foundation of China (no. 7220040764).

**Clan Culture and Patterns of Industrial Specialization in China**

**Abstract:** The clan, an extant social organization formed 2,000 years ago, bears a unique system of values still shaping fundamental institutions of modern society. In this research, we show how traditional clans affect the patterns of industrial specialization in China today. We find that industries dependent on relationship-specific investments tend to cluster in prefectures with strong clans. Our findings remain robust when considering alternative measures, including a set of historical and geographical correlates, and excluding cities in the southeast of China. Clans have a stronger effect on the specialization of the private sector than the state sector. In addition, the clan culture of immigrants matters for the industrial specialization of host regions. Our firm-level analysis further shows that the effects mainly originate from an overall improvement of the contracting environment by the clan culture.

**Key Words**: clans; industrial specialization; informal contracting institutions

**JEL codes**: Z10, O10, R12

It is of considerable economic consequence whether or not confidence, which is basic to business, rests upon purely personal, familial, or semi-familial relationships as was largely the case in China.

———Max Weber[[1]](#footnote-1)

1. **Introduction**

Previous studies have well recognized that contracting institutions, which regulate transactions between private parties and enable private contracts, are crucial to economic growth (see, e.g., Klein, Crawford, and Alchian, 1978; Williamson, 1979; Hart and Moore, 1990). Thus, institutional factors, such as the quality of a judicial system, have been extensively used to explain the cross-country and within-country variations in productivity, firm investment, capital allocation efficiency, and financial development (e.g., Gopalan, Mukherjee, and Singh, 2016; Giacomelli and Menon, 2017). The consensus among several studies is that a good contractual environment creates a comparative advantage for industries whose production is heavily dependent on contract enforcement. For example, Nunn (2007) show that countries with better contracting institutions generally export more in industries in which relationship-specific investments are important to the production of goods.[[2]](#footnote-2)

Notably, well-functioning formal contracting institutions are usually absent in emerging economies. The quality of China’s legal system is not on par with its rapid economic growth (e.g., Allen, Qian, and Qian, 2005; Xu, 2011). According to the Worldwide Governance Indicators (WGI), the index of “rule of law” ranks China at approximately 44 of 227 countries in 2017. Nonetheless, China’s exports have explosively increased in industries that are heavily dependent on contracting institutions; for example, the exports of electric machinery industry increased 29 times over the period of 2000-2013 (Ding, Fan, and Lin, 2018).[[3]](#footnote-3) While the coexistence of both China’s unprecedented growth of contract-intensive sectors and weak contracting institutions seems to be a contradiction, Acemoglu and Johnson (2005) provides an explanation: private parties have resources to develop alternative informal arrangements and alleviate the adverse effect of poor contracting institutions. Greif and Tabellini (2010) similarly argue that different societies rely on a combination of formal and informal enforcement institutions to sustain cooperation; particularly, in most developing countries characterized with collectivist societies, contract enforcement is mainly achieved through informal institutions (Greif, 1994).[[4]](#footnote-4)

To reconcile the argument that contracting institutions determine comparative advantage in the broad literature with China’s unique experience, our work aims to offer new insights into informal institutions that are of particular importance to China. A long history is a card of China in the world. In this paper, we focus on the clan, an incredibly important social organization built on kinship in ancient China, to explain the pattern of industrial specialization nowadays. The clan is a group consisting of agnate households with one common ancestor (Freedman, 1966; Greif and Tabellini, 2017). Historians believe that clans played predominate roles in managing grassroots affairs for thousands of years (Fei, 1946; Chu, 1962). Clans doled out financial provisions and informational resources to their members, constructed infrastructures, collected tax, and educated the young to approach officialdom. Studies have shown that clans promote the development of private sector, restrain the development of formal finance, and save lives in natural disasters (Cao, Xu, and Zhang, 2022; Chen, Ma, and Sinclair, 2021; Zhang, 2020). What makes clans attractive in explaining patterns of specialization is the institutional elements and cultural values embedded.

The first feature of clans relevant to informal institutions is the codification of rules. Clans typically establish their own rules and discipline members who violate the rules. The rules determine the rights and responsibilities of clan members in socioeconomic affairs, such as property ownership, the provision of offering succor to vulnerable groups, the punishment and reward for certain behaviors. The specific rules, particularly those recorded in the genealogy books, function in a similar manner as legal provisions featured in formal institutions. In this way, the kinship-based organization, which is complementary to bureaucracy, can protect property rights of their members and administer justice locally (See, e.g., Ruskola, 2000; Faure, 2007; Greif and Tabellini, 2017). The other elements include the construction of physical places for holding clan activities, that is, ancestral halls (*citang*, in Chinese), and substantial land and properties owned. Clan members regularly gather in the ancestral hall to perform ancestor-worship rituals or discuss property and business issues. The clan leader often announces punishments and rewards in the ancestral hall. In this sense, the hall functions like a legal court in modern society. The income generated from estate and property owned by clans guarantees the expenses on rituals, succor to the poor and unfortunate, and clan school, as well as constituting an internal financial market (Chen et al., 2021).

The clan norms and cultural values persist in modern society. Confucianism is centered around clan organizations for thousands of years. Opportunism—the pursuit of personal benefits at the sacrifice of the interests of counterparties—is not morally acceptable in clan culture. Greif and Tabellini (2010, p. 136) point out that, “cooperation within the clan is sustained mainly by moral obligations and reputational incentives that discourage cheating and free riding. Enforcement through formal institutions plays a small role.” Clans require a high level of trust among members to sustain cooperation. Kin-based trust, which might affect the level of societal trust in a region, reduces the transaction costs and improves the institutional environment.

These elements and norms institutionalize clans as organizations that enforce informal rules, sustain cooperation, and exhibit a strong sense of collectivism. It is worth noting that this form of social organization has survived after approximately one hundred years of extreme social changes, and in some regions, is flourishing in the reform era (Tsai, 2007; Cao et al., 2022). Our research goal is to examine whether well-organized clans create a comparative advantage for certain industries facing challenges in enforcing contracts.

Recently, several studies have emphasized the social capital and institutional features of clans and identified their roles in financial development and disasters (Chen et al., 2021; Cao et al., 2022). The most related paper to ours is Zhang (2020), which shows that clans help private firms in overcoming financial constraints and escaping from the “grabbing hand” of governments and, thereby, promote the development of the private sector; however, his paper takes all industries in one prefecture as a whole, while our paper concentrates on how the strength of local clans affects the development of each industry differentially.

To scrutinize the role of clans in regional specialization, we collect firm data from the Annual Survey of Industrial Firms to calculate the gross output of each industry in each prefecture, and we extract the genealogy information from the *General Catalog of Chinese Genealogy* (in Chinese, *Zhong Guo Jia Pu Zong Mu*) to measure the strength of local clans. The genealogy is a book that records the important information of clans, including their pedigrees from the oldest common ancestors, marriages, offspring, and excellent achievements of members, as well as clan rules and the public wealth. Genealogy itself, particularly the length and editions of genealogy, reflects the activity and local influence of a clan. An extensive compilation of genealogies exists in China.

Using the number of genealogies (normalized by population) to measure clan strength, we find that contract-intensive industries tend to cluster in the prefectures with strong clans. To pin down the relation between clan culture and formal contracting institutions, we count the number of civil and commercial cases (legal case) over the period of 2003–2007 in each prefecture and roughly assume that a large number of cases represents a high probability of a broken contract, as well as a poor contracting environment. We find evidence of a negative relation between the intensity of clans and the number of cases, supporting our hypothesis that clans promote the agglomeration of contract-intensive industries by providing alternative institutional arrangements.[[5]](#footnote-5)

The empirical findings are virtually unchanged when we construct a measure of China-specific contract intensity using China’s Input-Output table, count the number of genealogies compiled over different periods, address the concern of survivor bias driven by the Cultural Revolution, and control for a variety of important historical correlates and geographic factors. Clans are not randomly distributed across regions. To address the endogeneity concerns, we perform instrumented estimations by using the number of migrants settled in each prefecture during the 1127–1130 southward migration and the minimum distance to Zhu Xi academies as our instrument. Two-stage least squares (2SLS) regressions produce consistent results.

We further corroborate that the present pattern of industrial specialization is driven by the market force that incorporates the influence of clans, rather than a persistent pattern inherited from the centrally planned economy. Using data from the Second Industrial Census, we separately examine the effect of clans on industrial specialization of 85 cities covered in this census for the year 1985 and for our baseline year 2007. We find that clans play an essential role nowadays, while imposing a trivial effect on industrial development in the 1980s. To account for other factors as the foundation for social capital, we adopt the amount of donation and nongovernmental organization (NGO) participation as the proxies and find that our main findings remain robust when controlling for several measures of social capital.

Compared with privileged state-owned enterprises (SOEs), private-controlled firms are typically discriminated against in terms of access to legal services and the credit market. The dependence on informal institutions to enforce contracts are expected to be different in SOE and non-SOE groups. We separately calculate the specialization measures for SOE and non-SOE in each prefecture and run analogous regressions. The results show that clans have a strong effect on the specialization of non-SOE but have a negligible effect on the specialization of SOE.

Going further, we explore potential mechanisms. The positive effect of clans on the agglomeration of contract-intensive industries could be explained in two ways. One is that the strength of local clans shapes the informal contracting institutions at the regional level and has a general effect on local firms. The other is that the effect is specific to firms having connections with local clans. Using the Annual Survey of Industrial Firms, we extract surname information for each entrepreneur and construct a surname-based measure of a firm’s clan strength. We find that the overall density of local clans helps firms increase their operating profits and reduce their overhead expenses. However, we have no evidence of an individual effect, that is, an entrepreneur benefiting because of sharing a surname with a specific clan. The results favor a broad-brush explanation, namely, that the long-lived clans affect the institutional framework as a whole.

Finally, we investigate whether the clan background of immigrants can affect patterns of industrial specialization at host regions. China has witnessed a great inter-region migration in past decades. Evidence shows that immigrants have a long-term effect on the productivity, taxes, population growth, innovation, wages, native outflows, and foreign direct investments of host countries (See, e.g., Card and DiNardo, 2000; Hunt and Gauthier-Loiselle, 2010; Javorcik, Özden, Spatareanu, and Neagu, 2011). Drawing on the 2000 National Population Census data, we construct a population-weighted clan variable for immigrants and interact it with the measure of contract intensity in our estimations. The inclusion of the new clan variable does not alter our baseline findings, and particularly, we find that a strong clan background of immigrants also promotes the development of contract-intensive industries.

Our work contributes to the literature in several ways. First, our findings provide new insights into the determinants of industrial specialization and unbalanced regional development. Prior studies have captured several important factors that affect the agglomeration of industries in countries or in subnational regions, such as trade barriers (Krugman and Venable, 1996), financial development (Svaleryd and Vlachos, 2005; He, Xue, and Zhu, 2017), local protectionism (Bai, Du, Tao, and Tong, 2004; Lu and Tao, 2009), the efficiency of the legal system (He et al., 2019), and transport networks (Jaimovich, 2019). The present study contributes to this strand of literature by showing that clans, a special cultural tradition of China, act as an alternative informal arrangement and imbue contract-intensive industries with a long-run regional comparative advantage. Our paper also helps contextualize the role of informal institutions in China’s unconventional path of economic growth over the past few decades (e.g., Allen, Qian, and Qian, 2019).

Second, our paper explores the persistent effect of a historical social group, the clan, on today’s industrial development. Therefore, we contribute to the literature on the persistence of historical institutions (e.g., Jia, 2014; Guiso, Sapienza, and Zingales, 2016; Chen, Kung, and Ma, 2020; Zhang, 2021), as well as to the literature on the relation between culture and economic outcomes (e.g., Guiso et al., 2006; Tabellini, 2008; Zhang, 2020). This paper provides empirical evidence for Acemoglu and Johnson’s (2005) arguments about the development of alternative informal arrangements in the face of weak contracting institutions and the loose law-finance-growth nexus in Allen, Qian, and Qian (2005).

The rest of the paper is structured as follows. Section 2 discusses the background of clans in China and why clans matter for industrial specialization. Section 3 describes the data and research designs. Section 4 presents the empirical results. Section 5 concludes.

1. **Background**

*2.1. The clan and its persistence*

Clans are kinship-based groups consisting of agnate households in China (Freedman, 1966; Greif and Tabellini, 2017). The clan formally came into existence in 1100 BC, during the Western Zhou Dynasty (Peng, 2004), when polity relied on a patriarchal clan system and aristocracy. In imperial China, though the imperial regime primarily depended on bureaucracy, clans acted as an “informal” government and governed through grassroots efforts, especially in rural areas, for thousands of years (Chu, 1962). Clans provided public goods and services, such as education and the construction of infrastructures, administered justice, and offered succor to vulnerable groups (Faure, 2007). In doing so, they substituted for the functions of government to a certain extent (Fei, 1946). In ancient China, clans served to build and sustain people’s cooperation similar to independent cities and guilds in premodern Europe (Greif and Tabellini, 2017).

Historians point out that clan organizations experienced an accelerating development in the Song Dynasty (AD 960-1279) and flourished in the Qing Dynasty (1636-1911). Particularly, the southward migration driven by Jurchens’ invasion in 1100s and the Northern Song’s collapse dramatically shaped the spatial distribution of clans.[[6]](#footnote-6) The south-north division of China, between the Southern Song (1127–1279) and the Jurchens’ Jin (1115–1234), lasted about 150 years until the Mongols’ conquest in 1279. Hence, the migrated clans were less likely to move back. China has not experienced a South-North division for such a long time since then, and accordingly this wave of migration imposes a persistent effect on the population structure across regions. As shown in Fig. 1, clans nowadays are still concentrated in the southeast of China, like Zhejiang and Fujian provinces.

Clan organizations and relevant activities were nearly banned over the thirty years of Mao’s era, and many genealogies and ancestral halls were destroyed. Evidence shows that clans survived, resurged, and even flourished in several regions after the market reforms (Tsai, 2007; Cao et al., 2022). 21% of genealogies in our data set were compiled after 1949. We believe that the variations in the strength of clans across regions are rather persistent, and thereby, the historical institution can shape the industrial specialization today.

[Fig. 1 are about here]

An important reason for the boom of clan organizations in pre-industrial China is that clans worked as autonomous organizations and substituted for state power in several aspects, such as collecting tax and settling conflicts (Fei, 1946; Rowe, 1998). In pre-modern China, state authorities stopped at the county level, and matters at the grassroots (except murders) were usually handled by clans and some notables. Chen et al. (2021) show that clans managed commercial affairs and acted as the *de facto* enforcement rulers during the Qing Dynasty, which restrained China’s formal financial development. Even nowadays, clans still function as grassroots organizations and contribute to public governance, for example, mobilizing villagers to vote (Su et al., 2011), or increasing public goods expenditure (Xu and Yao, 2015). Governments, either in pre-industrial society or in modern society, need such grassroot organizations to maintain stability.

Long-lived social relationships among lineage groups also contribute to the persistent variations in clans among regions. China is a family- and relationship-centered society. Hometown connections are valuable to individuals and firms (Fisman et al., 2020). It is worth noting that clan member identity is inherently determined, and the clan lineage, that is, a member’s relationship to the clan, is rarely erased. Therefore, except for large shocks eliminating clans in a region, clan members today undoubtedly inherit the resources of the social networks established by their fathers, grandfathers, or prior generations.

Another factor behind the persistence are the cultural norms of clans. Tabellini (2008) points out that individual values provide a “missing link” between historical traditions and current institutions, as morality of good conduct was widespread in premodern society. The cultural values embedded in clans, particularly the Confucian norms, can be internalized in the spiritual world of their members, as well as local people, in the long run. In other words, offspring might inherit the eldership’s cultural values. Thus, regions with strong clans in pre-modern society are likely to continue compiling genealogies and holding ancestor-worship rituals nowadays, and their residents’ behaviors are dictated by the rules of local clans. For example, Chen et al. (2021) find that people from regions with strong clans tend to trust relatives more. Also, people with strong family ties are less likely to migrate (Alesina et al., 2015). Fig. 2 depicts the relationship among the number of genealogies compiled over different periods, and we find a strong and positive relation between any two periods.

 [Fig. 2 is about here]

*2.2. The clan and informal institutions*

Informal institutions refer to informal constraints encompassing norms of behavior, conventions, and self-imposed codes of conduct (North, 1990). The clan, a long-lived and dominant institution of premodern China, is very likely to affect norms and beliefs nowadays, and to influence, albeit informally, economic interactions, despite the fact that the institution itself has markedly declined of late. A crucial assumption of new institutional economics is that the opportunism of private parties may lead to broken contracts and result in hold-up problems (Williamson, 1985). Opportunism—the pursuit of personal benefits at the sacrifice of the interests of counterparties—is not morally acceptable in traditional Chinese culture. If the institutional elements and cultural norms associated with clans affect private parties’ ethical standards and increase the costs of default, the contractual environment should be improved. Through this lens, we posit that clans shape contracting institutions in several ways and affect the industrial specialization.

First, clans enforce their own informal rules to organize daily life. The rules determined the rights and responsibilities of clan members; particularly, clans punish members who violate the rules, or even eliminate a clan member’s identity for severe cases, such as murder or betraying clans. Also, clans build ancestral halls to regularly gather members to discuss socioeconomic affairs and hold important ceremonies, such as ancestor-worship rituals or weddings. The clan leader often announces punishments to its members in the ancestral hall. The combination of specific rules and physical facilities institutionalized clans and functioned similar to the combination of legal provisions and courts in a modern judicial system. Liu (1959) argues that “a clan should always watch its reputation by preventing its members from harming outsiders”. In this way, clan members are effectively monitored and less likely to engage in opportunistic behavior.

Second, clans are characterized by strong collectivism among clan members (Greif and Tabellini, 2013). In ancient China, clans protect their members from government or outside clan disputes. Information can be dispersedly and effectively transmitted through clan-based networks. Then, the opportunistic behaviors of outsiders would be considered damaging to the interests of the whole clan, which, in turn, create substantial obstacles for outsiders who are doing business in the located region. If an entity breaks a contract with a clan member intentionally, the opportunistic reputation of the counterparty would be disseminated throughout the entire clan, and the likelihood of other clan members transacting with the counterpart would be reduced. Therefore, the increase in opportunity costs associated with information transmission in the clan network largely impede any opportunistic behavior of outsiders and promote contract enforcement between private parties.

We have mentioned cultural norms embedded in clans to explain their persistence in Section 2.1. Following a similar line of argument, moral obligations and individual values (such as trust and cooperation) of clan members, which are inherited from older generations, can affect their economic behaviors today. Economic entities with a strong clan background instinctively discriminate against opportunistic behaviors and thus are less likely to engage in opportunistic behavior. Norms typically have a spillover to the entire region, rather than being limited solely to those entities in clan organizations. Under such conditions, the norms of clans can raise the ethical standards of local people and establish societal trust. In this sense, we expect that clans exert an influence beyond the entities having direct connections with local clans and can shape the overall contractual environment at the regional level. We will further discuss the issues and test the predictions in Section 4.5.

1. **Data and empirical designs**

***3.1 Key variables and sources***

**Industrial specialization**. We collect disaggregated firm-level data from the Chinese Annual Survey of Industrial Firms (ASIF) to calculate the specialization measure. This database has been widely used in prior studies on the Chinese economy (see, e.g., Huang, Li, et al., 2017; Fan, Li and Yeaple., 2018). By 2007, this database covered all state-owned enterprises (SOEs) and all other manufacturing firms with annual sales above 5 million yuan. However, small SOEs with annual sales below 5 (20) million yuan were excluded from 2008 (2011) onward. To ensure a relatively complete coverage of industrial output, we choose the year **2007** as our benchmark sample period.[[7]](#footnote-7)

To measure the degree of specific industry specialization in each prefecture, we define industrial specialization as the proportion of a specific industry’s output in a prefecture’s total production scaled by the average proportion of the industry’s output among all prefectures (See, e.g., Frésard, Hege and Phillips, 2017). Specifically, we define the specialization of industry *i* in prefecture *p* as , where *N* is the number of prefectures.[[8]](#footnote-8) is gross output from all manufacturing firms in industry *i* in prefecture *p*. The firms are distributed across 28 industries, as coded by the industrial classification for National Economic Activities, and across 330 prefectures in mainland China to yield a sample of 9,240 prefecture-industry observations.[[9]](#footnote-9)

**Clan culture**. In premodern China, clans were kinship-based social groups consisting of members with a common patrilineal ancestor. Like other social groups, a clan needs specific traditions and rituals to maintain the common identity of its members (clansmen). The genealogy is a book that records clan history, including pedigrees from the oldest common ancestors, marriage, offspring, and excellent achievements, as well as clan rules and public wealth. The compilation of genealogies could strengthen the sense of clan members’ belonging and accordingly the influence of the entire clan in the settled region. Following Greif and Tabellini (2010) and Zhang (2020), we use the number of genealogies in each prefecture to measure the strength of local clans.

The genealogy data are manually collected from *The General Catalog of Chinese Genealogies*, which is compiled by the Shanghai Library and was published by the Shanghai Ancient Books Publishing House in 2008. For each genealogy, we extract surname, compilation year, and location information. The primary measure of clans in the present study is the number of genealogies normalized by population, that is, . Fig. 1 depicts the geographic distribution of genealogies across regions. Cities with strong clans are concentrated in the southeast. Genealogy could be found in 283 prefectures in mainland China, among which 71 prefectures have more than 100 editions of genealogy. Jinhua in the Zhejiang province is the prefecture with the highest records, presenting 3,487 genealogies, whereas no records are available for most cities in Inner Mongolia and Xinjiang. Large variations in the number of genealogies across prefectures exist.

The *Catalog* records extant genealogies by the end of 2004. China experienced political and social turmoil in late Qing dynasty and in the period of Republic of China (1912-1949); also, traditional activities were severely suppressed during the Cultural Revolution (1966-1976). Here we show that, though clan activities were affected now and then, the variations in clans across regions are persistent. We observe a strong correlation between the number of genealogies compiled before 1850 and those after 1949, between those compiled before 1912 and those after 1949, between those compiled before 1949 and those after 1949, and between those compiled before 1949 and those after 1976 across prefectures, respectively (See Fig. 2).

**Industrial contract intensity**. We employ the index of contract intensity constructed by Nunn (2007) to measure the level of industry dependence on the contractual environment. The index is rationalized in the share of relationship-specific investments in the production of each industry. According to Rauch (1999), production inputs are coded as one of three classifications: sold on an organized exchange, reference priced, or neither. The former two have a relatively limited scope for hold-up and thus are less dependent on the quality of the contracting institutions. Nunn (2007) provides a measure of contract intensity that captures the relationship specificity for each industry *i* as where is the share of the value of input  used in the production of industry , and is the share of input that is relationship specific. Two measures are provided:  and . The former denotes inputs that are neither sold on the exchange nor reference priced as relationship specific, and the latter extends the scope to reference-priced inputs.[[10]](#footnote-10)

***3.2 Specification***

To empirically investigate whether clan intensity creates a comparative advantage for certain industries and shapes the industrial specialization in each prefecture, we estimate the following specification:

 (1)

where *i* and *p* corresponds to industry and prefecture, respectively. is the logarithm of gross output of industry *i* in prefecture *p.*[[11]](#footnote-11) is the logarithm of the cumulative genealogy number per 10,000 persons at prefecture *p*; is the measure of contract intensity, namely, the industrial dependence on contracting institutions. denotes the interaction between the strength of clans and industrial dependence on contract enforcement. The positive coefficient for implies that in China, prefectures with strong clans tend to specialize in industries with higher contract intensity. A negative sign implies that strong local clans impede the development of contract-intensive industries.

Like Nunn (2007), we include several interactions between regional factor endowments and their intensities in the production of goods in an industry. denotes the vector of industrial intensities, including and which corresponds to the dependence on agricultural inputs, mining inputs, and human capital in the production of industry *i*, respectively. denotes the vector consisting of and , that is, the corresponding factor endowments in prefecture *p*.

The industrial intensities of agriculture and mining are defined as the ratios of agricultural and mineral inputs to all intermediate inputs, the data of which are collected from China’s input-output table for the years 2002 and 2007. Human capital intensity is measured by the ratio of employees with at least a secondary education to the total number of employees in the industry. The endowments and are measured by the per capita output value of agriculture and mining in each prefecture. Human capital endowment, , is the share of the population with at least a secondary education in each prefecture.[[12]](#footnote-12)

We also interact several economic development indicators with industrial contract intensity to control for the possibilities that the ranking of clan strength is highly similar to the ranking based on economic margins and developed regions might specialize in certain industries. We adopt six indicators: the logarithm of gross domestic product (GDP) per capita (), the logarithm of population (), the share of urban population (*Urbanization*), the ratio of bank credit to GDP (*Loan/GDP*), the amount of fixed investments (*ln Fixed Inv*), and the ratio of outputs of tertiary sector to GDP (*Share 3rd*). The spatial distribution of clans is far from random, and geography and climate conditions obviously affect factor endowment and comparative advantage. We control for the potential importance of geographic factors by including the interactions of contract intensity with the logarithm of the nearest distance between a prefecture’s centroid to the coast, the average gradient, the river density, and the land size. We include industry and prefecture fixed effects, and , in all regressions.

Table 1 reports the summary statistics of all variables used in this study. The statistics show a large variation in the number of genealogy books across cities. The mean and median number of genealogies are 130.9 and 19.5. As shown in Fig. 1, prefectures with high records of genealogies concentrate in the southeast, like Jinhua and Shaoxing from Zhejiang province. However, we do find that several cities with strong clans are from central regions, such as Anhui, Jiangxi, Hunan, and Henan provinces. The skewed distribution of clans makes the sensitivity checks of excluding cities without clans or excluding cities in the southeast essential. Appendix Table A1 defines the variables and summarizes data sources.

[Table 1 is about here]

***3.3 Preliminary patterns***

Before rigorous regression estimations, we begin with some suggestive evidence. We first show a roughly positive relationship between an output-weighted average of industrial contract intensity and clan intensity for all prefectures in Fig. 3. Moreover, we decompose our prefecture–industry data set into four subsamples based on contract intensity and the intensity of local clans. An industry is classified as more (less) dependent on relationship-specific investments if its contract intensity score is above (below) the sample median. A prefecture is classified as having strong (weak) clans if its number of genealogies normalized by the population is above (below) the sample median. We then calculate the mean ratio of the industry output scaled by the total output of the located prefecture in each subsample and present the comparisons in Table 2.

[Fig. 3 is about here]

The results from both panels show that the economic importance of contract-intensive sectors is much better in prefectures with strong clan culture (row 1): take Panel A, for example. There, the mean ratios of industrial output in the subsamples of weak clan and strong clan are 0.0130 and 0.0218, respectively, which are statistically different at the level of 1%. By contrast, industries with less dependence on relationship-specific investments produce higher output in prefectures with weak clan culture (row 2). The univariate analysis paints a picture that prefectures with high genealogy density are more likely to specialize in industries with a high reliance on contracting enforcement.

[Table 2 is about here]

1. **Empirical results**

***4.1 Baseline results***

Table 3 reports the OLS regression results for model (1). The first column estimates the effect of the interaction between clan and contract intensity (*z1*) on the pattern of industrial specialization without the inclusion of other endowment intensities and controls. The estimated coefficient for is positive and statistically significant at the 1% level, suggesting that strong clan culture creates a comparative advantage for the development of contract-intensive industries. Next, in column 2, we control for factor endowment interactions that may influence industry agglomeration, that is, , , and , all of which are introduced in model (1). The clan interaction with contract intensity, our key interest, remains significantly positive. The economic magnitude of the comparative advantage generated by the clan is substantial. If Hanzhong, a historical city in the Shaanxi province, held the number of genealogies equal to Shantou’s, that is, from the 25th percentile to the 75th percentile in our sample, then its output of the “food production” industry would increase by 15.46%, and the share of food production to total output would increase by 257 basis points. Also, agriculture, mining, and human capital interactions produce significantly positive signs, in agreement with Romalis (2004) and Nunn (2007), who find that factor endowments are the source of comparative advantage for production.

In addition to the endowment interactions, we control for other factors that may bias the importance of clans in driving the specialization if omitted. We interact contract intensity with a variety of economic and geographic indicators. The results in column 2 show that regions with a prosperous tertiary sector tend to specialize in contract-intensive industries. Consistent with our conjecture, geographical factors are important for the pattern of specialization. Estimates show that contract-intensive industries are less likely to cluster in prefectures with a long distance to the coast or in mountainous areas.

As clans are unevenly distributed across regions in China, in column 3 we further restrict our analysis to prefectures with at least one genealogy. We do so to check whether our results are driven by many zero observations. In our sample, 58 prefectures do not have any stored genealogies. Excluding these omissions, the subsample consists of 7,616 observations (28 × (330 − 58)). Consistently, we find that strong clans promote the agglomeration of contract-intensive industries at the prefecture level. Moreover, when we adopt another measure of contract intensity from Nunn (2007), *z2*, a broader definition of relationship-specific investments, our new estimates reported in columns 4–6 are consistent with those obtained using the narrow measure, *z1*.[[13]](#footnote-13) To better fit the production from Chinese firms, we construct a China-specific contract intensity measure drawing on the data from the China Economic Census Yearbook and China’s input-output table.[[14]](#footnote-14) The results in Appendix Table A2 show that the adoption of the new contract intensity does not alter our estimates: the coefficients for are statistically significant and positive.

[Table 3 is about here]

 In Section 2, we demonstrated that clan culture substantially increases the costs of opportunistic behaviors, and thereby, entities are less likely to engage in such activities. We put those arguments into an empirical test of the relation between the strength of local clans and the quality of contracting environment. By manually collecting the data fromstatistical or comprehensive yearbooksthat are compiled by each prefecture, we count the number of civil and commercial cases () over the period of 2003–2007 in each prefecture. We roughly assume that fewer legal cases could be a) a sign of less dependency on formal institution to deal with contract disputes; b) a lower probability of a broken contract, and thus represent a better contracting environment.

 Column 1 of Table 4 shows a significantly negative relation between the number of civil and commercial cases (normalized by population) and the intensity of local clans. These results confirm that firms are less likely to engage in contractual disputes in the prefectures with strong clans. Clans, as a special purveyor of social capital in premodern China, act as informal but effective middlemen to handle contractual issues or act as an effective mechanism to improve the contracting environment. We then regress the industrial output on the interaction between industrial contract intensity and the predicted value of derived from the regression on clans in column 1. Unsurprisingly, the interactions produce significantly negative signs, suggesting that a poor contracting environment is harmful to the development of contract-intensive industries.[[15]](#footnote-15) The evidence supports our prediction that clans promote the agglomeration of contract-intensive industries by improving contracting institutions. We also present a simple partial equilibrium model in Appendix B, which predicts that firms in sectors dependent on the contract enforcement gain a comparative advantage in regions with strong clans.

[Table 4 is about here]

***4.2 Robustness checks***

*4.2.1 Alternative measures of clan strength and the Cultural Revolution*

One potential concern in our measurement of the clan intensity is the date of genealogy compilation. *The General Catalog of Chinese Genealogy* records all genealogies compiled before 2004. China notably suffered dozens of wars in the late Qing dynasty and over the period of the Republic of China (1912-1949). Violent conflicts, accompanied by tens of thousands of deaths, may heavily affect the priorities of those who would normally store and compile genealogies across regions. Additionally, some prefectures or even provinces were governed by foreign authorities for quite a long time.

To check the robustness of our results, we separately employ the number of genealogies written before the years 1850, 1912, and 1949 and those after 1949 to measure the intensity of clans. The People’s Republic of China and Republic of China were founded in 1949 and 1912, respectively. Year 1850 is the last year of the reign of Qing’s Emperor Dao-Guang, during whose dominion occurred the famous Opium War (1840–1842). The outbreak of this war plunged China into a hundred years of wars. Appendix Table A2 report the estimation results. All the coefficients on clan interactions are statistically and significantly positive, suggesting that the compilation of genealogies, as the embodiment of clan culture, consistently provides industries that require intensive relationship-specific investments a comparative advantage.

Moreover, during 1966–1976, the Cultural Revolution (CR) embraced an elimination of the old (*po si jiu* in Chinese), namely, old thoughts, old culture, old customs, and old habits (MacFarquhar and Schoenhals, 2006). Clan activities were strongly suppressed during this period; in particular, many genealogies were destroyed in violence. It is possible that our measure of clan strength suffers from survivor bias (Cao et al., 2022). However, Fig. 2(d) shows a strong and positive relation between the number of genealogies compiled before 1949 and those after 1976, implying that the distribution of clans did not experience a structural change during the period of severe government suppression.

To address the concern of survivor bias thoroughly, we first investigate whether the extent of violence during the CR indeed affects the storage and compilation of genealogy books. We collect the data on the number of victims in the CR violence at the county level from Walder (2015). The book assembles relevant figures from the county chronicles. We aggregate the data into the prefecture level and employ the number of victims scaled by 1964 population as the proxy for the extent of CR violence. In Panel A of Table 5, we regress the total number of genealogies, the number of genealogies written before 1949, and the number of those written after 1976 on the CR violence measure in columns 1-3, respectively. The high violence of CR significantly reduces the number of genealogies stored nowadays. However, the magnitude of the effect on the genealogies written after 1976 is much smaller than that on those written before 1949. To further clarify the pattern, we divide the cities into high violence and low violence subgroups according to the median number of victims. The results in columns 4-5 suggest that, in the low-violence subsample, the extent of CR violence has a negligible effect on either the total number of genealogies or the number of those compiled before 1949, both of which are more likely to suffer from survivor bias. In other words, the negative effect of CR violence on the compilation of genealogies is concentrated in high-violence regions.

Next, we additionally add an interaction between CR violence and industrial contract intensity into our model (1). The results are reported in columns 1 and 4, Panel B of Table 5 (*z1* and *z2*, respectively). We find that the estimated coefficient on the CR violence interaction is statistically and significantly negative, suggesting that prefectures experiencing high violence in CR are less likely to specialize in contract-intensive sectors. We conjecture that the violence severely destroyed societal trust that is important to contract enforcement. In particular, we find that, after controlling for the CR violence, the strength of local clans continues to impose a significant effect on industrial specialization. Stepping further, we employ the number of genealogies written after CR (1976) to measure the strength of clans and report the results in columns 2 and 5. The coefficients on the alternative measure that is less likely to suffer from survivor bias remain significant. Last, we run analogous regressions using the low-violence subsample in which CR violence has a negligible effect on the compilation of genealogies. All the estimated coefficients on clan interactions are significantly positive.

[Table 5 is about here]

*4.2.3 Controlling for historical and geographical factors*

Historical institutions have been shown to exert a persistent effect on several aspects of economic development, even nowadays. In this subsection, we test whether our findings are robust to the inclusion of a variety of historical factors.

The first correlate is the success of *Keju*, that is, China’s imperial examination system for civil service before 1905. Chen et al. (2020) show that the long-lived *Keju* has a strong and positive impact on human capital in modern China. The familial transmission of *Keju* culture and today’s endowment of human capital across regions obviously affect patterns of industrial specialization. We then employ the total number of *jinshi* during the Ming and Qing dynasties normalized by population to measure the success of historical education for each prefecture. The data come from the *Distribution of Jinshi in Ming Dynasty* and *List of Jinshi in Qing Dynasty*. A second set of historical correlates refer to openness and invasion of foreign cultural norms. We construct a dummy variable, *Treaty port*, indicating whether the city was a treaty port from 1840s to 1910s. Western countries were more likely to establish commercial and manufacturing facilities in the treaty ports, and evidence shows that treaty ports grew faster than non-treaty ports in the reform era (Jia, 2014). We also control for the number of Catholic churches in a prefecture, which represents the penetration of western cultural norms.

Next, we include the historical urbanization rate, measured by the proportion of the urban population in 1920, and the ratio of the non-agricultural population to total population in 1964. These two factors characterize the population distribution and industrial structure before the market reform. Last, we consider the amount of capital for modern firms over the period of 1840-1927 in each prefecture, a proxy for the economic prosperity of the late-Qing episode. The data are manually extracted from Du (1991) who compiled the information of industrial firms before 1927 from various sources, such as government reports and firm archives. We incorporate the six historical correlates into the estimation of model (1) and report the results in first two columns of Appendix Table A3. The inclusion of historical correlates does not alter our findings.

Prefectures with intensive clan culture are concentrated in the southeast, which spreads out along the coast to the north and along the Yangtze River to the west. To tackle the issue of skewed distribution, we exclude cities in the southeast of China to ensure robustness.[[16]](#footnote-16) Specifically, we exclude all cities from nine provincial judications in the southeast, that is, Jiangsu, Shanghai, Zhejiang, Guangdong, Guangxi, Fujian, Hainan, Anhui, and Jiangxi provinces, and report the estimates in columns 3-4 of Table A3. We find that clans are still an important determinant of comparative advantage. The magnitudes become smaller than those obtained from baseline estimations, for example, 0.772 vis-à-vis 1.120 for the coefficients on *z1× Clan*. The decrease in the magnitude reinforces our argument that clans can shape regional specialization, as powerful clans are concentrated in the removed cities.

Weather conditions are important to the development of agriculture, which is the dominant sector in pre-industrial society. Also, climate might affect cultural norms and cooperation. Talhelm et al. (2014) show that a history of farming rice makes cultures more interdependent, and the cultural values further shape the behaviors of local people in modern society. Farming rice or wheat is naturally determined by weather conditions. In this sense, we additionally control for a set of weather conditions to ensure robustness. In columns 5-6, we interact the contract intensity with the average values of rainfall, sunshine, and temperature over 2000-2007. We again find that contract-intensive sectors tend to cluster in cities with strong clans, and estimates on controls show that these sectors are likely to agglomerate in rainy and cold cities.

*4.2.4 The role of clans before market reform*

One could argue that clans shaped China’s industrial layout decades ago, even before the foundation of the People’s Republic of China in 1949, and that the pattern persists today. This would make the observed relation between regional clan intensity and current industrial specialization spurious. China’s transformation from a planned economy to a market economy provides an appropriate setting to address this concern. In a planned economy, firms mainly follow the national strategy in deciding where to locate their headquarters, whereas in a market economy firms enjoy more freedom to incorporate many factors into this decision. The reform and opening-up policy, in place since 1978, jumpstarted China’s explosive economic growth over the past forty years. In particular, several policies under the reform of the urban economic system, starting from 1984, such as the establishment of special economic zones and open coastal cities, strongly promoted the development of private sectors across cities (Worden et al., 1987). Here, we roughly peg the reform of the urban economic systemas the end point of China’s planned economy.

To check whether the pattern is persistent, we turn to the Second Industrial Census carried out in 1985, when the reform for urban and industrial sectors had just started. The industrial layout during that time can be regarded as an artifact of central planning. On the contrary, the specialization pattern of our benchmark year 2007 reflects many factors present in a market-oriented economy. Industrial data fromthe Second Industrial Census are available for 85 cities only. To ensure a good comparison, we separately estimate model (1) for the covered 85 cities using 1985 and 2007 industrial production data. Table 6 summarizes the results. We find that the intensity of local clans does not affect regional industrial specialization in 1985, whereas clans are again an essential determinant of industrial agglomeration in the 2007 subsample. Therefore, we conclude that the pattern of current industrial specialization across prefectures evolved after the economic reform, and that the distribution of clans played an important role during this time.

[Table 6 is about here]

To further exclude the possibilities that our findings are mainly driven by government policies, we additionally control for two sets of controls. The establishment of Special Economic Zones (SEZ) has been shown to affect the economic development, capital accumulation, and human capital (Alder, Shao, and Zilibotti, 2016). We count the number of SEZ in each prefecture and augment our model by interacting the measure with contract intensity. Moreover, we control for the fiscal expenditure and government transfer, which may manifest the policy support from the central government. The results reported in Appendix Table A4 provide robust evidence for our findings.

*4.2.5 Social capital*

Social capital can be defined as “the ability of actors to secure benefits by virtue of membership in social networks” (Bourdieu, 1985). Clans are organizations that enforce informal rules, sustain cooperation, and provide public goods to their members. Thus, clans characterized by strong collectivism among members can act as a foundation for social capital, especially in rural areas (Greif and Tabellini, 2013). Cao et al. (2022) concentrate on clan organizations to examine how social capital saves lives in disasters. However, other factors might constitute social capital that sustains cooperation and enhances societal trust, and the ranking of regions based on these omitted factors are unexpectedly similar to that of clan strength.

To account for other foundation for social capital, we follow Hasan, He, and Lu (2022) to construct two proxies for regional social capital in China. Hansen et al. (2022) adopt voluntary blood donation and nongovernmental organization (NGO) participation to construct a composite social capital index at the provincial level. People who donate blood or register with NGOs are typically not driven by economic incentives; thus, an active participation here demonstrates high levels of societal trust and cooperation. As the data on blood donation is not available at the prefecture level, we instead adopt the amount of donation as one proxy. NGO participation is measured by the number of people registered with the NGO in each prefecture. Both the donation and NGO data are collected from the appendix of Chinese Civil Affairs Statistical Yearbook.

We conduct a cross-sectional analysis to examine whether the strength of clans is related to social capital measures. The results are reported in Panel A, Table 7. We find a significantly positive relation between these two, corroborating our argument that clans act as a foundation for social capital. Then, we include the interaction between clans and contract intensity and two interactions between social capital proxies and contract intensity simultaneously into the estimations. The results show that the coefficients on social capital interactions are insignificant in 3 out of 4 columns, while those on clan interactions are all statistically and significantly positive. This suggests that our main findings of clan effect remain robust when controlling for several other measures of social capital.

[Table 7 is about here]

***4.3. State sector and private sector***

The legal system is underdeveloped in China. Private sector, especially small- and middle-size enterprises, are discriminated against in terms of access to legal protection and credit resources (Allen et al., 2005); by contrast, large and state-owned firms (SOEs) have privileged access and are less likely to suffer from the “grabbing hand” of governments. In the abovementioned analysis, we argue that clans, as a key informal arrangement, can create a comparative advantage for industries whose production heavily relies on contracting institutions. The clan-based network and clan rules effectively hamper opportunistic behaviors and reduce enforcement costs; cultural norms, such as moral obligations, facilitate the establishment of societal trust and sustain cooperation. In a nutshell, clans play a role when access to legal services is limited. The private sector is hypothesized to depend more on informal arrangements to conduct business than the state sector.[[17]](#footnote-17) In addition, many SOEs follow the national strategy in deciding where to locate their headquarters and factories, which makes these sites exogeneous to certain attributes of cities.

In this section, we separately examine the effect of clans on the specialization of SOEs and non-SOEs to strengthen our arguments. The Chinese Annual Survey of Industrial Firms database provides the information on the registration category of each firm, such as state-controlled enterprise (110), wholly foreign-owned enterprise (330), or private partnership (172). We then classify the firms into SOE and non-SOE subgroups[[18]](#footnote-18), and correspondingly calculate the measures of SOE and non-SOE specialization for each prefecture.

We report the estimation results of model (1) by regressing the SOE and non-SOE industrial outputs on clans in columns 1-2 of Table 8. We find that the coefficient on the interaction between clans and contract intensity is only significant when taking the non-SOE specialization measure as the dependent variable, while the estimated coefficient is insignificant in SOE regression. This suggests that clans have a trivial effect on the specialization of SOEs across regions. Also, the coefficient for non-SOEs is much larger than that for SOEs (1.102 vs 0.223). To better compare the patterns in two samples, we perform a simulation procedure to determine the significance of differences in coefficients (e.g., Cleary, 1999). The empirical *p*-value from the bootstrapping procedure is 0.000, indicating that SOEs and non-SOEs perform significantly differently at the 1% level. Following the same sequence, we report the estimated results of using another measure of contract intensity, *z2*, in columns 3-4. The comparison again implies that clans as grassroots organizations are more likely to affect the development of the private sector.

[Table 8 is about here]

***4.4. Instrumented results***

A prefecture’s industrial specialization is likely associated with a wide variety of factors, especially historical factors, that cannot be exhaustively controlled for. Thus, concerns that omitted variables are simultaneously correlated with regional industrial specialization and the intensity of clans should be further addressed. To mitigate this concern, we perform instrumental variable estimations and choose two instruments (IV).

The first IV is the southward migration during 1127–1130, which was driven by the Jurchen’s invasion and the Northern Song’s collapse. The migration wave typically intensified competition between the migrated and native clans, which resulted in two outcomes that enhanced their local influence: a) quantity, more children for enlarging clan size; b) quality, investing more resources in education for the future success of *Keju* and governmental connections (Bai, 2022). Both actions would be positively associated with the length and the number of editions of a clan’s genealogy. The information on the number of migrants settled in each prefecture is obtained from the *History of Migration in China* (Wu, 1997), and we manually count the number of migrants who settled in each prefecture during this wave of migration by matching the city name in the Song Dynasty with present jurisdictions.

Our second IV is the minimum distance to any of three Zhu Xi academies in Southern Song Dynasty, that is, *Bailudong* Academy, *Yuelu* Academy, and *Hanquan* Academy. As explained carefully in Chen et al. (2021) and Cheng et al. (2021), Zhu Xi is the most important person popularizing clan culture and establishing detailed institutional designs for clan organizations. Given that transportation costs were considerably high in ancient China, people in nearby regions were more likely to hear the lectures on clan-related doctrines held in these academies and therefore, were exposed to the influence of Zhu Xi and his followers historically. Moreover, the geography distance to historical places should be orthogonal to the economic prosperity in 21st century, as the location choice of the three academies is by coincidence.[[19]](#footnote-19)

To formally test the relevance condition, we regress each prefecture’s number of genealogies on the number of migrants during 1127–1130 and the minimum distance to Zhu Xi academies. The first-stage estimations in Panel B of Table 9 validate the rationale of our instruments. The southward migration significantly encourages the compilation of genealogies in the settled regions,[[20]](#footnote-20) but the intensity of clans decreases with the distance to Zhu Xi academies. In Panel A, after instrumenting by the migrant volume and distance to Zhu Xi academies and controlling for a variety of economic and geographic controls, we again find that the intensity of local clans has a strong explanatory power for the patterns of industrial specialization across regions. The results remain unchanged when using the subsample of cities with nonzero genealogies in columns 3-4. The coefficients on the instrumented clan interaction are greater than the OLS estimates (1.633 vis-à-vis 1.120 for *z1* in the full sample), implying that the endogenous genealogy variable underestimates the long-run effect of clans due to potential measurement errors and omitted factors. F-statistics show that our IVs are far from weak instruments, and Hansen J statistics suggest that they pass the over-identification tests. Taken together, the results from IV estimations assure us that industries with a high level of reliance on contracting institution are more likely to cluster in prefectures with powerful clans.

[Table 9 is about here]

***4.5. Firm-level analysis***

We now attempt to examine how the intensity of local clans encourages the development of contract-intensive industries. As discussed in Section 2, the characteristics of social organizations and cultural norms embedded in clans can shape the contractual environment by motivating private parties to eschew opportunistic behavior, but we posit that the spheres of influence are different. If clans act as a foundation for social capital, opportunistic behaviors are constrained, because a) clans discipline members who violate the rules and thereby, clansmen engage less in activities that damage the interests and reputation of the whole clan; b) clansmen are less likely to do business with outsiders who have broken a contract with other members of their clan due to the effective information transmission through the network. In this way, the benefits of strong clans for contract enforcement tend to concentrate on firms that have a specific relationship with a powerful clan. If the effect of clans is mainly driven by cultural norms, we expect that norms typically have a spillover to the entire region. The beliefs and values held by local people can be viewed as internalized customs passed on from long-lived clans. Clans raise ethical standards and establish societal trust at the regional level in the long run; thus, the culture protects local firms dependent on relationship-specific investments from default in a general manner.

To differentiate between these two explanations, we perform a firm-level analysis and create a new clan measure, , which is defined as the logarithm of the number of genealogies for the surname of each firm’s entrepreneur (i.e., the legal representative in the data set), in the prefecture where the headquarter is located. Again, the data are obtained from the Chinese Annual Survey of Industrial Firms, which covers more than 300,000 industrial firms. We collect a series of financial statement items, including size, leverage, returns, fixed assets, number of employees, operating profits, and overhead expenses, and we estimate the following specification:

where the subscripts *i*, *j*, and *p* correspond to the surname of the entrepreneur, the industry type, and the prefecture, respectively. We use two measures as a proxy for firm performance: the operating profit ratio and the overhead expenses ratio. If firms directly benefit from the intensity of surname-related clans, rather than the overall intensity of the clan tradition at the regional level, we should observe a statistically significant and positive and an insignificant ; on the contrary, if the overall density of clans creates a comparative advantage for all firms heavily dependent on contract enforcement, then the significance of should dominate that of .

 Panel A, Table 10 reports the results of firms’ profitability. In column 1, we first include and a set of firm controls in the estimation. Surprisingly, we find that the strength of the firm-specific clan cannot impose a positive effect. However, the pattern changes when the interactions for the surname-prefecture count of genealogies with industrial contract intensity are included in columns 2 and 3 (*z1* and *z2*, respectively). The results show that the intensity of the firm owner’s clan significantly increases the operating profit of firms in contract-intensive industries.

Columns 4 and 5 confirm our baseline results using highly disaggregated firm-level data. Our key clan interaction term in model (1) maintains its significance, ensuring a positive effect of an overall intensity of local clans on the industrial specialization. Interestingly, when two interactions for the overall clan intensity and for the strength of the firm owner’s clan simultaneously enter the estimations in columns 6 and 7, we find no evidence of a significant effect generated from a specific clan. Nevertheless, we continue to see that the regional intensity of all clans substantially boosts the operating income of firms facing more challenge in contract enforcement. This contrast suggests that all firms in contract-intensive sectors gain a comparative advantage from the overall density of local clans, in favor of a regional improvement hypothesis.

In a similar sequence, we examine the effect of clans from the perspective of expenses. China is a family- and relationship-centered society. Either a political or a hometown connection is valuable to individuals and firms (Ding et al., 2018; Fisman et al., 2020). The establishment and maintenance of relationships largely rely on frequent communications like business and entertainment activities. Firms in contract-intensive industries may devote resources to negotiating with their suppliers for input prices and customized needs or maintaining valuable relationships with their suppliers, both of which incur overhead expenses unrelated to the production.

If strong clans improve the informal contractual environment, firms are less likely to suffer from a broken contract and accordingly reduce their expenses on traveling or business entertainment. Panel B reports the results. Likewise, we first separately include the number of genealogies for a surname-prefecture unit and for a prefecture unit in our regressions and then simultaneously incorporate the two interactions into our model. The patterns are almost the same to those observed in Panel A. The estimated coefficients for the surname-prefecture clan variables have a moderate effect on expenses after we control for the total number of genealogies (see columns 6 and 7). Thus, we conclude that the beneficial effect is not specific to a sparse set of firms whose owners have the same surnames as the local clans; instead, the intensity of clans could generate a comparative advantage for all firms from contract-intensive sectors in this region.

[Table 10 is about here]

***4.6. Further discussion on Migration***

The impact of immigrants on host countries gains much attention in international economics literature. Evidence shows that immigrants can have a long-term effect on productivity, taxes, population growth, innovation, wages, native outflows, and foreign direct investments (See, e.g., Card and DiNardo, 2000; Hunt and Gauthier-Loiselle, 2010; Javorcik, Özden, Spatareanu, and Neagu, 2011). Immigrants not only directly contribute to the economy of the host country, but also indirectly affect a variety of economic outcomes through spillovers on local citizens. For example, skilled immigrants can enrich the critical understanding of local researchers on a specialized field and boost innovation together (Hunt and Gauthier-Loiselle, 2010).

It is worth noting that immigrants tend to take the occupations that they had in their home countries and probably form an industry clustering in certain cities. Hornung (2014) shows that most Huguenots settled in Prussia resumed their occupation and concentrated on textile and apparel industries. They brought many skills that were unknown to the host regions, and consequently, most towns hosting the first generation of Huguenot immigrants exhibited high productivity of textile industries and developed large-scale production. In this sense, immigrants can transfer the technology unevenly across industries and affect the industrial agglomeration of host countries.

Despite the restrictions of the household registration system (known as “*Hukou*”), past decades have witnessed a great internal migration in China. According to the survey of the National Bureau of Statistics, the number of workers migrating from rural areas increased to 168 million in 2015. It is important to emphasize that inter-region migration in China is usually organized by families, kinsmen, and villages.

The clan background of immigrants may affect their employment distribution and further the industrial specialization of host regions for several reasons. First, following a similar line of argument to Hornung (2014), immigrants tend to resume the occupations that they have at home, where the distribution of occupations (or say industrial specialization) is affected by the strength of home-region clans. A large volume of immigrants flowing into a sparse set of industries may boost the development of certain industries. Second, a clan network is a kind of social capital. The first-mover immigrants tend to help those following them into a city to find jobs if they have social connections, which concentrates the employment of immigrants within certain industries. Third, the cultural values and beliefs held by immigrants may have a spillover effect on their social networks and further shape the contractual environment at host regions.

To investigate the effect of immigrant-backed clan strength on host-region industrial specialization, we collected the migration data from the 2000 National Population Census and calculated the number of immigrants at the prefecture-pair level. We constructed a weighted clan variable, *Clan\_Migration*, as follows:

where is the number of immigrants from prefecture *j* to prefecture *p*. We then incorporated the interaction term between and industrial contract intensity in model (1).

 The results are reported in Table 11. Three columns correspond to Nunn’s contract intensity *z1*, *z2*, and China’s contract intensity, respectively. We observe that the strength of local clans maintains its significant effect on the industrial specialization. More importantly, two out of three coefficients on the weighted migration-backed clan variable are statistically and significantly positive, suggesting that a large number of immigrants from prefectures with strong clans also promote the development of contract-intensive industries at host regions.

[Table 11 is about here]

1. **Conclusion**

Nowadays, institutions are seen as channels through which history matters (Acemoglu et al., 2001; Tabellini, 2008). Using data on genealogies across Chinese prefectures, this paper examines the role of the clan, which we argue is an informal institution formed by historical tradition, in shaping the specialization of manufacturing industries. We find that industries heavily relying on relationship-specific investments tend to cluster in prefectures with a high intensity of clans. Our results remain robust after we control for a variety of determinants of industrial specialization and perform several robustness checks. We also find that clans have a stronger effect on the specialization of the private sector than the state sector. By constructing a measure of the firm-specific clan relationship, we further show that the clan effect is not specific to a sparse set of firms whose owners have the same surnames as the local clans. In this sense, clan culture improves the contracting environment for *all* local firms. Finally, we show that the strength of clan background of immigrants also affects the patterns of industrial specialization at host regions.

This paper builds on the ongoing literature that explores how informal institutions play a role in China’s unconventional path of economic growth. We focus on a strong and institutionalized social organization, clans, which formed 2000 years ago, flourished during the Qing Dynasty, and resurged in past decades in China. We provide evidence that clan culture shapes the long-run pattern of industrial development. Our work also adds to the literature on the relation between culture and economic performance, as well as the interplay between formal and informal institutions. Variations in the industrial specialization driven by the strength of long-lived clans provide a new explanation for the unbalanced regional development within China. Our findings provide policy implications for economies where good formal institutions are absent that industrial policies can suit local traditions and cultural norms.

Future research could consider a close scrutiny of the relationship between historical institutions and industry agglomeration along other dimensions, such as Confucian or temple associations. It is also worth noting that the findings in this paper do not implicitly suggest that historical kinship-based clans outperform a modern judicial system. A social group dependent on kinship and moral standards naturally confers disadvantages under a modern economic regime in which contracts are largely transacted among strangers. However, an old system in work can be a second best.

**References**

Acemoglu, D., Johnson, S. (2005). Unbundling Institutions. *Journal of Political Economy*, vol.113, 949–995.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation, *American Economic Review*, vol.91, 1369–1401.

Alder, S., Shao, L., and Zilibotti, F. (2016). Economic reforms and industrial policy in a panel of Chinese cities, Journal of Economic Growth, vol. 2, 305–349.

Allen, F., Qian, J., and Qian, M. (2005). Law, finance and economic growth in China, *Journal of Financial Economics*, vol.77, 57–116.

Allen, F., Qian, J., and Qian, M. (2019). A Review of China’s Institutions, *Annual Review of Financial Economics*, vol.11, 39–64.

Alesina, A., Algan., Y., Cahuc, P., Giuliano, P. (2015). Family values and the regulation of labor. *Journal of the European Economic Association*, vol.13 (4), 599–630.

Araujo, L., Mion, G., Ornelas, E. (2016). Institutions and Export Dynamics. *Journal of International Economics,* vol.98 (1), 2–20.

Bai, C. E., Du, Y., Tao, Z., and Tong, S. Y. (2004) Local protectionism and regional specialization: evidence from China’s industry, *Journal of International Economics*, vol. 63, 397–417.

Bai, Y. (2022). The struggle for existence: migration, competition and human capital accumulation in historic China, *International Economic Review*, vol.63, 1239–1269.

Bourdieu, P. (1985). “The Forms of Capital.” In Handbook of Theory and Research for the Sociology of Education, J. G. Richardson, ed. New York, NY: Greenwood Press.

Cao, J., Xu, Y., Zhang, C. (2022). Clans and calamity: How social capital saved lives during China’s Great Famine. *Journal of Development Economics*, vol.157, 102865.

Card, D., DiNardo, J. (2000). Do Immigrant Inflows Lead to Native Outflows? *American Economic Review*, vol.90, 360–367.

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, vol.104 (11), 3600–3634.

Chen, T., Kung, J. K. and Ma, C. (2020). Long Live Keju! The Persistent Effects of China's Imperial Examination System, *The Economic Journal*, vol.130, 2030–2064.

Chen, Z., Ma, C., Sinclair, A.J. (2021). Banking on the Confucian Clan: Why China Developed Financial Markets So Late? *The Economic Journal,* vol.132, 1378–1413.

Cheng, J., Dai, Y., Lin, S., Ye, H. (2021). Clan Culture and Family Ownership Concentration: Evidence from China. *China Economic Review*, vol.70, 101692.

Chisik, R. (2003). Export Industry Policy and Reputational Comparative Advantage. *Journal of International Economics*, vol.59, 423–451.

Chu, T. (1962). *Local Government in China under the Ch'ing*, Cambridge, MA.: Harvard University Press.

Cingano, F., Pinotti, P. (2016). Trust, Firm Organization, and the Pattern of Comparative Advantage. *Journal of International Economics*, vol.100, 1–13.

Cleary, S., 1999. The Relationship between Firm Investment and Financial Status. *Journal of Finance*, vol.54 (2): 673–692.

Ding, H., Fan, H., and Lin, S. (2018). Connect to trade, *Journal of International Economics*, vol.110, 50–62.

Du, X., 1991, *National Capitalism and the Government of Old China (1840-1937)* (Minzu Zibenzhuyi he Jiuzhongguo zhengfu), Shanghai: Shanghai Academy of Social Science Press.

Elman, B. A. (1991). Political, Social, and Cultural Reproduction via Civil Service Examinations in Late Imperial China, *Journal of Asian Studies*, vol.50, 7–28.

Fan, H., Li, Y. A., Yeaple, S. (2018), On the Relationship Between Quality and Productivity: Evidence from China's Accession to the WTO, *Journal of International Economics*, vol.110, 28–49.

Faure, D. (2007). *Emperor and Ancestor: State and Lineage in South China*, Stanford: Stanford University Press.

Feenstra, R., Hong, C., Ma, H., and Spencer, B.J. (2013). Contractual Versus Non-Contractual Trade: The Role of Institutions in China, *Journal of Economic Behavior & Organization*, vol.94, 281–294.

Fei, H. (1946). Peasantry and Gentry: An Interpretation of Chinese Social Structure and Its Changes, *American Journal of Sociology*, vol.52, 1-17.

Fisman, R., Shi, J., Wang, Y., and Wu, W. (2020) Social Ties and the Selection of China's Political Elite, *American Economic Review*, vol.110(6): 1752–1781.

Frésard, L., Hege, U., and Phillips, G. (2017). Extending Industry Specialization through Cross-Border Acquisitions, *Review of Financial Studies*, vol.30, 1539–1582.

Freedman, M. (1966). *Chinese Lineage and Society: Fukien and Kwangtung*, London: The Ahtlone Press.

Giacomelli, S., Menon, C. (2017). Does weak contract enforcement affect firm size? Evidence from the neighbor’s court, *Journal of Economic Geography*, vol.17, 1251–1282.

Gopalan, R., Mukherjee, A., and Singh, M. (2016) Do Debt Contract Enforcement Costs Affect Financing and Asset Structure?, *Review of Financial Studies*, vol.29, 2774–2813.

Greif, A. (1994). Cultural beliefs and the organization of society: a historical and theoretical reflection on collectivist and individualist societies, *Journal of Political Economy*, vol. 102 (5), 912–50.

Greif, A., Tabellini, G. (2010). Cultural and institutional bifurcation: China and Europe compared. *American Economic Review: Papers & Proceeding*, vol.100 (2), 135–140.

Greif, A., Tabellini, G. (2013). Social organizations, violence and modern growth, *American Economic Review*, vol.103 (3), 534–538.

Greif, A., Tabellini, G. (2017). The clan and the corporation: sustaining cooperation in China and Europe, *Journal of Comparative Economics*, vol.45, 1–35.

Guiso, L., Herrera, H., and Morelli, M. (2006). Does Culture Affect Economic Outcomes?, *Journal of Economic Perspectives,* vol.20, 23–48.

Guiso, L., Herrera, H., and Morelli, M. (2016). Cultural Differences and Institutional Integration, *Journal of International Economics*, vol.99, S97–S113.

Guiso, L., Sapienza, P., Zingales, L. (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics,* vol.124 (3), 1095–1131.

Hart, O., Moore, J. (1990). Property rights and the nature of the firm, *Journal of Political Economy,* vol.98, 1119-1158.

Hansen, I., He, Q., Lu, H. (2022). Social Capital, Trusting, and Trustworthiness: Evidence from Peer-to-Peer Lending. *Journal of Financial and Quantitative Analysis* 57(4), 1409–1453.

He, Q., Xue, C., and Zhou, S. (2019) Does contracting institution affect the patterns of industrial specialization in China?, *China Economic Review*, vol.54, 191–203.

He, Q., Xue, C., and Zhu, C. (2017). Financial development and patterns of industrial specialization: Evidence from China, *Review of Finance*, vol.21, 1593–1638.

Hornung, E. (2014). Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia, *American Economic Review*, vol.104, 84–122.

Huang, Z., Li, L., Ma, G., and Xu, L. C. (2017). Hayek, Local Information, and Commanding Heights: Decentralizing State-Owned Enterprises in China, *American Economic Review*, vol.107, 2455–2478.

Hunt, J., Gauthier-Loiselle, M. (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics*, vol.2, 31–56.

Jaimovich, E. (2019). Roadways, input sourcing, and patterns of specialization, *European Economic Review*, vol.120, 103319.

Javorcik, B. S., Özden, C., Spatareanu, M., Neagu, C. (2011). Migrant networks and foreign direct investment, *Journal of Development Economics*, vol. 94, 231–241.

Jia, R. (2014). The legacies of forced freedom: China's treaty ports, *Review of Economics and Statistics*, vol.96, 596–608.

Kalemli-Ozcan, S., Sørensen, B. E. and Yosha, O. (2003). Risk Sharing and Industrial Specialization: Regional and International Evidence, *American Economic Review*, vol.93, 903–918.

Klein, B., Crawford, R. G., and Alchian, A. A. (1978) Vertical Integration, Appropriable Rents, and the Competitive Contracting Process, *Journal of Law and Economics*, vol. 21, 297–326.

Krugman, P., Venables, A. J. (1996). Integration, specialization, and adjustment, *European Economic Review*, vol.40, 959–967.

Liu, H. W., (1959), *The Traditional Chinese Clan Rules*, Locust Valley, NY.: J.J. Augustin.

Lu, J., Tao, Z. (2009). Trends and determinants of China’s industrial specialization, *Journal of Urban Economics,* vol. 65, 167–180.

MacFarquhar, R., Schoenhals, M. (2006). *Mao's Last Revolution*, Cambridge, MA.: Harvard University Press

North, D. (1990). *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade, *Quarterly Journal of Economics,* vol.122, 569–600.

Peng, Y. (2004). Kinship Networks and Entrepreneurs in China’s Transitional Economy, *American Journal of Sociology*, vol.109, 1045–1074.

Rajan, R. G., Zingales, L. (1998). Financial dependence and growth, *American Economic Review*, vol.88, 559–586.

Rauch, J. E. (1999). Networks versus markets in international trade, *Journal of International Economics*, vol.48, 7–35

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade, *American Economic Review*, vol.94, 67–97.

Rowe, W.T. (1998). Ancestral Rites and Political Authority in Late Imperial China: Chen Hongmou in Jiangxi, *Modern China,* vol.24(4), 378–407.

Ruskola, T. (2000). Conceptualizing Corporations and Kinship: Comparative Law and Development Theory in a Chinese Perspective, *Stanford Law Review,* Vol. 52, 1599–1729.

Su, F., Ran, T., Sun, X., and Liu, M. (2011). Clans, Electoral Procedures and Voter Turnout: Evidence from Villagers' Committee Elections in Transitional China, *Political Studies*, vol.59, 432–457.

Svaleryd, H., Vlachos. J. (2005) Financial markets, the pattern of industrial specialization and comparative advantage: Evidence from OECD countries, *European Economic Review*, vol.49(1), 113–144.

Tabellini, G. (2008). Institutions and Culture, *Journal of the European Economic Association*, vol. 6, 255–294.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., and Kitayama, S. (2014). Large-scale Psychological Differences within China Explained by Rice versus Wheat Agriculture. *Science,* vol.344, 603–608.

Tsai, L.L. (2007). Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China, *American Political Science Review*, vol.101, 355–372.

Walder, A. G. (2015). China under Mao: A Revolution Derailed. Harvard University Press.

Williamson, O. E. (1979). Transaction Cost Economics: The Governance of Contractual Relations, *Journal of Law and Economics*, vol.22, 233–261.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*, New York: The Free Press.

Worden, R. L., Savada, A. M. and Dolan, R. E. (1987). *China: A Country Study*, Washington: GPO for the Library of Congress.

Wu, S. (1997). *History of Migration in China (Zhongguo yimin shi)*, vol.4, Fuzhou: Fujian People’s Publishing House

Xu, C. (2011). The fundamental institutions of China's reforms and development, *Journal of Economic Literature*, vol.49, 1076–1151.

Xu, Y., Yao, Y. (2015). Informal institutions, collective action, and public investment in rural China. *American Political Science Review*, vol.109, 371-391.

Yan, Z., 1955, Selected Statistical Materials on Modern Chinese Economic History (zhongguo jndai Jingjishi tongji ziliao xuanji), Beijing: Science Press.

Zhang, C. (2020). Clans, entrepreneurship, and development of the private sector in China, *Journal of Comparative Economics*, vol.48, 100–123.

Zhang, J. (2021). Is the Confucian ethic a hindrance to economic development in China?, *Economic and Political Studies*, vol.9, 255-314.



**Fig. 1** The geography distribution of clans

*Notes*: A prefecture with more (less) genealogies is shaded in darker (lighter) blue. Regions with no available data are colored in white.

 

1. (b)

 

1. (d)

**Fig. 2 Persistence of clans**

*Notes*: The *x*-axis and *y*-axis present the cumulative number of genealogies (normalized by population) during the specified period. Each point corresponds to a prefecture. Prefectures with no genealogies are omitted.



(A)



(B)

**Fig. 3** Strength of the clan and weighted average of industrial contract intensity

*Note*: The *x*-axis represents the logarithm of the number of genealogies normalized by population; the *y*-axis represents the output-weighted average of industrial contract intensity at the prefecture level. Each point corresponds to a prefecture. Z1 and Z2 are two measures of contract intensity in Nunn (2007).

**Table 1 Summary statistics**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Obs | Mean | Median | P25 | P75 | Std |
| 　 | (1) | (2) | (3) | (4) | (5) | (6) |
| *SP* | 9240 | 10.45 | 12.74 | 9.352 | 14.54 | 5.916 |
| *# Genealogies* | 330 | 130.89 | 19.5 | 3 | 81 | 340.05 |
| *Clan* | 330 | -2.697 | -2.896 | -3.967 | -1.584 | 1.705 |
| *Economic controls* |  |  |  |  |  |  |
| *ln GDPpc* | 330 | 9.683 | 9.646 | 9.208 | 10.1 | 0.665 |
| *ln Pop* | 330 | 5.691 | 5.817 | 5.263 | 6.267 | 0.825 |
| *ln Fixed inv* | 330 | 5.447 | 5.424 | 4.794 | 6.063 | 1.049 |
| *Share 3rd* | 330 | 0.357 | 0.351 | 0.31 | 0.402 | 0.0819 |
| *Loan/GDP* | 330 | 0.029 | 0.019 | 0.010 | 0.030 | 0.039 |
| *Urbanization* | 330 | 0.343 | 0.291 | 0.197 | 0.451 | 0.191 |
| *ln AGRpc* | 330 | -1.062 | -1.025 | -1.291 | -0.787 | 0.428 |
| *ln Minepc* | 330 | 8.249 | 8.63 | 7.314 | 9.99 | 3.116 |
| *Edu* | 330 | 0.021 | 0.015 | 0.010 | 0.023 | 0.020 |
| *Geography controls* |  |  |  |  |  |  |
| *ln Distance to Coast* | 330 | 5.797 | 6.162 | 5.098 | 6.795 | 1.365 |
| *Gradient* | 330 | 11.26 | 10.64 | 6.383 | 15.36 | 5.895 |
| *River Density* | 330 | 0.231 | 0.228 | 0.184 | 0.267 | 0.0767 |
| *Land Size* | 330 | 24.9 | 13.24 | 7.974 | 22.35 | 44.69 |
| *Other variables* |  |  |  |  |  |  |
| *ln Jinshi* | 330 | 3.526 | 4.103 | 1.946 | 5.147 | 2.176 |
| *CR Violence* | 330 | -0.468 | -0.219 | -1.679 | 0.786 | 1.831 |
| *Treaty Port* | 330 | 0.236 | 0 | 0 | 0 | 0.425 |
| *ln Catholic Church* | 330 | 1.844 | 1.792 | 1.099 | 2.565 | 1.192 |
| *Urbanization in 1920* | 282 | 0.057 | 0.031 | 0 | 0.072 | 0.082 |
| *Non-Agr in 1964* | 330 | 0.17 | 0.112 | 0.0684 | 0.218 | 0.159 |
| *ln Modern Firms’ Cap* | 330 | 3.985 | 4.615 | 0 | 6.77 | 3.34 |
| *ln NGO employees* | 326 | 7.442 | 7.514 | 6.633 | 8.417 | 1.593 |
| *ln Value of Donation* | 325 | 6.837 | 6.66 | 5.802 | 7.82 | 1.517 |
| *Migration* | 330 | 0.564 | 0 | 0 | 0.693 | 1.087 |
| *ln Dist Academy* | 330 | 6.556 | 6.672 | 6.136 | 7.164 | 0.876 |

*Notes*: This table reports the summary statistics of all variables used in this paper. See variable descriptions in Appendix Table A1.

**Table 2 A simple comparison of industrial output**

|  |  |  |  |
| --- | --- | --- | --- |
| 　 | (1) | (2) | (3) |
| Panel A: *Z1*  | 　 | 　 | 　 |
|  | Strong clans | Weak clans | Diff. |
| High intensity | 0.0218 | 0.0130 | 0.0087\*\*\* |
| Low intensity | 0.0278 | 0.0327 | -0.0049\*\*\* |
|  |  |  |  |
| Panel B: *Z2* | 　 | 　 | 　 |
|  | Strong clans | Weak clans | Diff. |
| High intensity | 0.0238 | 0.0147 | 0.0091\*\*\* |
| Low intensity | 0.0267 | 0.0318 | -0.0051\*\*\* |

*Notes*: Each number in columns 1 and 2 is the mean of industrial output scaled by total output of the located prefecture in each category. An industry is classified as high (low) dependent on relationship-specific investments if its score of contract intensity is above (below) sample median. A prefecture is classified as having strong (weak) clans if its number of genealogies normalized by population is above (below) sample median. Panels A and B employ Z1 and Z2 in Nunn (2007) to measure the contract intensity of each industry, respectively. Column 3 present the differences between columns 1 and 2. \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, 1% levels, respectively.

**Table 3 Clans and industrial specialization: OLS regressions**

|  |  |  |  |
| --- | --- | --- | --- |
| 　 | *Z1* | 　 | *Z2* |
|  | Full Sample | Full Sample | Clan>0 |  | Full Sample | Full Sample | Clan>0 |
| 　 | (1) | (2) | (3) | 　 | (4) | (5) | (6) |
| *Z × Clan* | 1.705\*\*\* | 1.120\*\*\* | 0.974\*\*\* |  | 2.007\*\*\* | 1.251\*\*\* | 1.086\*\*\* |
|  | (0.121) | (0.148) | (0.167) |  | (0.163) | (0.193) | (0.213) |
| *Z × ln Pop* |  | 0.815 | 1.328\* |  |  | 1.821\* | 2.222\*\* |
|  |  | (0.715) | (0.793) |  |  | (0.941) | (1.055) |
| *Z × ln GDPpc* |  | -0.531 | 0.0598 |  |  | 2.413\* | 3.411\*\* |
|  |  | (0.907) | (0.973) |  |  | (1.255) | (1.340) |
| *Z × Urbanization* |  | 1.751 | 2.548 |  |  | 1.756 | -0.330 |
|  |  | (1.784) | (2.018) |  |  | (2.483) | (2.772) |
| *Z × Share 3rd* |  | 6.913\*\* | 4.170 |  |  | -4.581 | -4.180 |
|  |  | (3.376) | (4.224) |  |  | (4.505) | (5.610) |
| *Z × Loan/GDP* |  | -10.02 | -10.70 |  |  | -25.03\*\* | -26.88\*\* |
|  |  | (6.510) | (7.642) |  |  | (10.04) | (11.83) |
| *Z × ln Fixed Inv* |  | 0.359 | 0.114 |  |  | -0.562 | -0.578 |
|  |  | (0.702) | (0.744) |  |  | (0.942) | (1.021) |
| *Z × Land Size* |  | -0.0142\*\* | -0.0402\*\* |  |  | -0.01 | -0.0398\*\* |
|  |  | (0.0063) | (0.0157) |  |  | (0.0093) | (0.0182) |
| *Z × River Density* |  | 3.843 | 4.676 |  |  | 5.921 | 3.452 |
|  |  | (3.484) | (3.585) |  |  | (4.286) | (4.383) |
| *Z × ln Distance to Coast* |  | -0.732\*\*\* | -0.480\*\* |  |  | -0.836\*\*\* | -0.699\*\* |
|  |  | (0.218) | (0.230) |  |  | (0.277) | (0.290) |
| *Z × Gradient* |  | -0.144\*\*\* | -0.115\*\* |  |  | -0.235\*\*\* | -0.206\*\*\* |
|  |  | (0.0467) | (0.0530) |  |  | (0.0598) | (0.0655) |
| *agr × ln AGRpc* |  | 2.147\*\*\* | 0.796 |  |  | 3.253\*\*\* | 2.173\*\*\* |
|  |  | (0.571) | (0.592) |  |  | (0.590) | (0.608) |
| *mine × ln Minepc* |  | 0.604\*\*\* | 0.606\*\*\* |  |  | 0.738\*\*\* | 0.720\*\*\* |
|  |  | (0.122) | (0.146) |  |  | (0.130) | (0.151) |
| *Skill × Edu* |  | 394.2\*\*\* | 426.4\*\*\* |  |  | 422.2\*\*\* | 456.7\*\*\* |
|  |  | (66.85) | (69.79) |  |  | (66.94) | (68.80) |
| Prefecture FE | Yes | Yes | Yes |  | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes |  | Yes | Yes | Yes |
| Observations | 9,240 | 9,240 | 7,616 |  | 9,240 | 9,240 | 7,616 |
| R-squared | 0.622 | 0.632 | 0.599 | 　 | 0.620 | 0.633 | 0.603 |

*Notes*: This table reports the OLS regressions of the relation between the strength of local clans and industrial specialization. The dependent variable is the logarithm of the output of a specific industry in a specific prefecture. *Clan* is the logarithm of cumulative genealogy number per 10,000 persons in a prefecture, and *Z* corresponds to contract intensity. Control variables include the interaction of contract intensity with population, GDP per capita, urbanization, the share of tertiary sector outputs, the ratio of bank credits to GDP, fixed investments, land size, river density, the distance to the coast, and average gradients, the interaction of agricultural endowments and agricultural intensity, the interaction of mining endowments and intensity, and the interaction of human capital stock and the industrial human capital dependence. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 4 Clans and contracting institutions**

|  |  |  |  |
| --- | --- | --- | --- |
| 　 | legal cases | 　 | Industrial output |
| 　 | (1) | 　 | (2) | (3) |
| *Clan* | -0.0544\*\*\* |  |  |  |
|  | (0.0172) |  |  |  |
| *Z1 × predicted\_legal case* |  |  | -14.92\*\*\* |  |
|  |  |  | (2.851) |  |
| *Z2 × predicted\_legal case* |  |  |  | -14.34\*\*\* |
|  |  |  |  | (3.917) |
| Controls | Yes |  | Yes | Yes |
| Prefecture FE | No |  | Yes | Yes |
| Industry FE | No |  | Yes | Yes |
| N | 238 |  | 6664 | 6664 |
| R-squared | 0.636 | 　 | 0.618 | 0.619 |

*Notes*: The dependent variable in column 1 is the total number of civil and commercial cases over the period of 2003–2007 normalized by the 2007 population. *predicted\_legal case* is the predicted value of *legal case* from the estimation in column 1. In column 1, the control variables include GDP per capita, population, agricultural output per capita, mining output per capita, and human capital stock and other economic controls. In columns 2-3, the control variables are the same as those in Table 3. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 5 Robustness: The Cultural Revolution**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Panel A | All genealogies | Genealogies before 1949 | Genealogies after 1976 | All genealogies;Low-violence subsample | Genealogies before 1949;Low-violence subsample |
|  | (1) | (2) | (3) | (4) | (5) |
| *CR Violence* | -0.138\*\*\* | -0.152\*\*\* | -0.0836\*\* | 0.0423 | 0.0803 |
|  | (0.0408) | (0.0422) | (0.0349) | (0.0725) | (0.0768) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 330 | 330 | 330 | 165 | 165 |
| R-squared | 0.405 | 0.358 | 0.255 | 0.487 | 0.444 |

|  |  |  |  |
| --- | --- | --- | --- |
| Panel B | *Z1* |  | *Z2* |
|  | Full Sample | Full Sample | Low-violence subsample |  | Full Sample | Full Sample | Low-violence subsample |
| 　 | (1) | (2) | (3) | 　 | (4) | (5) | (6) |
| *Z × CR Violence* | -0.228\* |  |  |  | -0.324\*\* |  |  |
|  | (0.128) |  |  |  | (0.164) |  |  |
| *Z × Clan* | 1.066\*\*\* |  | 1.119\*\*\* |  | 1.175\*\*\* |  | 1.390\*\*\* |
|  | (0.149) |  | (0.219) |  | (0.197) |  | (0.277) |
| *Z × Clan after 1976* |  | 1.221\*\*\* |  |  |  | 1.116\*\*\* |  |
|  |  | (0.176) |  |  |  | (0.230) |  |
| Controls | Yes | Yes | Yes |  | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes |  | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes |  | Yes | Yes | Yes |
| N | 9,240 | 9,240 | 4,620 |  | 9,240 | 9,240 | 4,620 |
| R-squared | 0.633 | 0.632 | 0.657 | 　 | 0.634 | 0.633 | 0.66 |

*Notes*: In Panel A, we regress the number of genealogies on the extent of violence in the Cultural Revolution (CR), which is defined as the number of victims in CR violence normalized by the 1964 population. The data are collected from Walder (2015). Columns 1-3 report the estimation results of all genealogies, the genealogies compiled before 1949, and those compiled after 1976. Columns 4-5 report the estimations using the subsample of cities with below-median number of victims (low-violence subsample). In Panel B, we include the interaction between CR violence and contract intensity in our estimations in columns 1 and 4. Columns 2 and 5 adopt the number of genealogies compiled after 1976 to proxy for the strength of clans. Columns 3 and 6 report the results using the low-violence subsample. The control variables in Panel B are the same as those in Table 3. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 6 Role of clans before the market reform**

|  |  |  |  |
| --- | --- | --- | --- |
| 　 | 2007 output | 　 | 1985 output |
|  | *Z1* | *Z2* |  | *Z1* | *Z2* |
| 　 | (1) | (2) | 　 | (3) | (4) |
| *Z × Clan* | 0.592\*\* | 0.793\*\* |  | 0.164 | -0.0436 |
|  | (0.238) | (0.396) |  | (0.167) | (0.262) |
| Controls | Yes | Yes |  | Yes | Yes |
| Prefecture FE | Yes | Yes |  | Yes | Yes |
| Industry FE | Yes | Yes |  | Yes | Yes |
| N | 2,268 | 2,268 |  | 2,268 | 2,268 |
| R-squared | 0.570 | 0.571 |  | 0.627 | 0.625 |

*Notes*: This table reports the estimation results using the data of 85 cities covered in *the Second Industrial Census* in 1985. The dependent variables in columns 1-2 are the output of an industry in a prefecture at year 2007 and in columns 3-4 are the industrial output in 1985. *Clan* is the logarithm of cumulative genealogy number per 10,000 persons in a prefecture, and *Z* corresponds to contract intensity. The control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 7 Social Capital**

|  |  |  |
| --- | --- | --- |
| Panel A | *ln NGO employees* | *ln Value of Donation* |
|  | (1) | (2) |
| *Clan* | 0.140\*\*\* | 0.117\*\*\* |
|  | (0.0351) | (0.0429) |
| Controls | Yes | Yes |
| Observations | 326 | 325 |
| R-squared | 0.724 | 0.584 |

|  |  |  |
| --- | --- | --- |
| Panel B | *Z1* | *Z2* |
|  | (1) | (2) | (3) | (4) |
| *Z × ln Value of Donation* | -0.010 |  | 0.301 |  |
|  | (0.210) |  | (0.277) |  |
| *Z × ln NGO employees* |  | 0.836\*\*\* |  | 0.487 |
|  |  | (0.253) |  | (0.340) |
| *Z × Clan* | 1.076\*\*\* | 0.992\*\*\* | 1.159\*\*\* | 1.160\*\*\* |
|  | (0.152) | (0.153) | (0.198) | (0.199) |
| Controls | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Observations | 9,100 | 9,128 | 9,100 | 9,128 |
| R-squared | 0.627 | 0.630 | 0.628 | 0.631 |

*Notes*: Panel A examines the relation between social capital and the strength of clans. *ln NGO employees* is the logarithm of the number of people registered with the NGO; *ln Value of Donation* is the logarithm of total amount of donation for each prefecture. Panel B reports the estimation results of model (1) by additionally incorporating the interaction between social capital and contract intensity. The control variables are the same as those in Table 3. All regressions in Panel B include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 8 SOE and non-SOE**

|  |  |  |
| --- | --- | --- |
| 　 | *Z1* | *Z2* |
|  | SOE | Non-SOE | SOE | Non-SOE |
|  |
| 　 | (1) | (2) | (3) | (4) |
| *Z × Clan* | 0.223 | 1.102\*\*\* | 0.436\* | 1.042\*\*\* |
|  | (0.177) | (0.148) | (0.234) | (0.174) |
| Controls | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Observations | 9,240 | 9,240 | 9,240 | 9,240 |
| R-squared | 0.379 | 0.649 | 0.378 | 0.650 |
| Empirical *p*-value | 0.000 | 0.000 |

*Notes*: The dependent variables are the logarithm of industrial outputs aggregated from state-controlled enterprises (SOEs) in columns 1 and 3 and are the logarithm of industrial outputs aggregated from non-SOEs in columns 2 and 4. The last row reports the empirical *p*-values which are calculated in a simulation procedure to test the significance of difference in coefficient estimates of different samples. The control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 9 Instrumented results**

|  |  |  |
| --- | --- | --- |
| Panel A: second-stage results |  |  |
|  | Full Sample | *Clan*>0 |
|  | *Z1* | *Z2* | *Z1* | *Z2* |
|  | (1) | (2) | (3) | (4) |
| *Z × Clan* | 1.633\*\*\* | 1.667\*\*\* | 1.408\*\*\* | 1.367\*\*\* |
|  | (0.191) | (0.220) | (0.270) | (0.310) |
| Controls | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Observations | 9,240 | 9,240 | 7,616 | 7,616 |
| R-squared | 0.632 | 0.634 | 0.599 | 0.603 |
| Panel B: first-stage results |  |  |  |  |
| *Z × Migration* | 0.382\*\*\* | 0.384\*\*\* | 0.384\*\*\* | 0.386\*\*\* |
|  | (0.0190) | (0.0261) | (0.0187) | (0.0258) |
| *Z × ln Dist Academy* | -0.986\*\*\* | -0.983\*\*\* | -0.833\*\*\* | -0.833\*\*\* |
|  | (0.0304) | (0.0419) | (0.0321) | (0.0443) |
| Kleibergen-Paap rk LM statistic | 987.1 | 518.4 | 905.2 | 479.2 |
| Kleibergen-Paap rk Wald F statistic | 1304.6 | 686.7 | 1181.6 | 623.4 |
| Hansen Test | 0.42 | 0.12 | 0.28 | 0.13 |

*Notes*: This table presents two-stage-least-squared estimations for the relation between the strength of local clans and industrial specialization. We use two instruments. The first one is the number of migrants settled in the prefecture during 1127-1130 southward migration (*Migration*); the second is the minimum distance to any Zhu Xi academies in the Southern Song Dynasty (*ln Dist Academy*). *Clan* is the logarithm of cumulative genealogy number per 10,000 persons in a prefecture, and *Z* corresponds to contract intensity. The data of migrants are manually collected from *History of Migration in China* (Wu, 1997). Columns 3 and 4 use the sample in which the prefectures have stored at least one genealogy. The control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 10 Firm level analysis: clan and operating performance**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 　 | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: Operating profit ratio* |  |  |  |  |  |  |  |
| *Clan Surname* | 0.0002 | -0.0007\*\* | -0.0022\*\*\* |  |  | -0.0000 | -0.0002 |
|  | (0.0002) | (0.0003) | (0.0007) |  |  | (0.0004) | (0.0009) |
| *Z1 × Clan Surname* |  | 0.0018\*\*\* |  |  |  | 0.0005 |  |
|  |  | (0.0005) |  |  |  | (0.0006) |  |
| *Z2 × Clan Surname* |  |  | 0.0027\*\*\* |  |  |  | 0.0004 |
|  |  |  | (0.0007) |  |  |  | (0.0010) |
| *Z1 × Clan* |  |  |  | 0.0020\*\*\* |  | 0.0017\*\*\* |  |
|  |  |  |  | (0.0005) |  | (0.0006) |  |
| *Z2 × Clan* |  |  |  |  | 0.0030\*\*\* |  | 0.0028\*\*\* |
|  |  |  |  |  | (0.0006) |  | (0.0008) |
| N | 296631 | 296631 | 296631 | 296631 | 296631 | 296631 | 296631 |
| R-squared | 0.370 | 0.370 | 0.370 | 0.370 | 0.370 | 0.370 | 0.370 |
|  |  |  |  |  |  |  |  |
| *Panel B: overhead expenses* |  |  |  |  |  |  |  |
| *Clan Surname* | -0.0005\*\*\* | 0.0005\*\* | 0.0017\*\*\* |  |  | -0.0004 | -0.0011\* |
|  | (0.0001) | (0.0002) | (0.0005) |  |  | (0.0003) | (0.0006) |
| *Z1 × Clan Surname* |  | -0.0021\*\*\* |  |  |  | -0.0002 |  |
|  |  | (0.0004) |  |  |  | (0.0005) |  |
| *Z2 × Clan Surname* |  |  | -0.0024\*\*\* |  |  |  | 0.0007 |
|  |  |  | (0.0005) |  |  |  | (0.0007) |
| *Z1 × Clan* |  |  |  | -0.0026\*\*\* |  | -0.0025\*\*\* |  |
|  |  |  |  | (0.0004) |  | (0.0005) |  |
| *Z2 × Clan* |  |  |  |  | -0.0035\*\*\* |  | -0.0039\*\*\* |
|  |  |  |  |  | (0.0005) |  | (0.0006) |
| N | 296533 | 296533 | 296533 | 296533 | 296533 | 296533 | 296533 |
| R-squared | 0.212 | 0.212 | 0.212 | 0.212 | 0.212 | 0.212 | 0.212 |
| Firm controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ownership FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Surname FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* This table reports the estimation results using the firm level data from 2007 Chinese Annual Survey of Industrial Firms (ASIF). The dependent variable is the ratio of operating profits to operating income in Panel A and the ratio of overhead expenses to operating income in Panel B. *Clan Surname* is the logarithm of the number of genealogies for the surname of each firm’s entrepreneur (i.e., the legal representative in the data set) in a prefecture. Z1 and Z2 are two indices of contract intensity in Nunn (2007). *Clan* is the logarithm of cumulative genealogy number per 10,000 persons in a prefecture. Controls include firm leverage, ROA, fixed assets, firm size, the number of employees. All regressions include prefecture, industry, ownership, and surname fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table 11 Migration-backed clan strength**

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Z1* | *Z2* | *China's Z*  |
|  | (1) | (2) | (3) |
| *Z × Clan\_Migration* | 15.63\*\* | 8.391 | 17.45\*\* |
|  | (6.609) | (10.61) | (7.741) |
| *Z × Clan* | 0.901\*\*\* | 1.398\*\*\* | 0.789\*\*\* |
|  | (0.229) | (0.294) | (0.264) |
| Controls | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes |
| N | 8,540 | 8,540 | 8,540 |
| R-squared | 0.625 | 0.628 | 0.621 |

*Notes: Clan\_Migration* is the weighted clan strength for immigrants, which is calculated by the equation: , where is the number of immigrants from prefecture *j* to prefecture *p*. The data of immigrants are from 2000 National Population Census. *Clan* is the cumulative genealogy number per 10,000 persons in a prefecture, and *Z* corresponds to contract intensity. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Online Appendix (not for publication)**

**Appendix A**

**Table A1 Variable definitions**

|  |  |  |
| --- | --- | --- |
| Variables | Definitions | Sources |
| *SP* | Ln (1+ 2007 outputs of a specific industry in a prefecture) | Chinese Annual Survey of Industrial Firms (ASIF) |
| *Clan* | Ln [(# Genealogies+1)/population 2007] | The General Catalog of Chinese Genealogies |
| *Z1* | Contract intensity, where is the share of the value of input used in the production of industry and is the share of input that is neither sold on the exchange nor reference priced. | Nunn (2007) |
| *Z2* | Contract intensity, where is the share of the value of input used in the production of industry and is the share of input that is not sold on the exchange. | Nunn (2007) |
| *China’s Z* | Contract intensity of China | He et al. (2019) |
| *ln GDPpc* | Ln (GDP per capita) | China Statistics Yearbook for Regional Economy |
| *ln pop* | Ln (population 2007) | China Statistics Yearbook for Regional Economy |
| *Urbanization* | The proportion of the non-agricultural population | Demographic Data of Counties and Cities in the People's Republic of China, 2007 |
| *Share 3rd* | The ratio of tertiary sector outputs to GDP | China Statistics Yearbook for Regional Economy |
| *Loan/GDP* | The ratio of bank credits to GDP | China Statistics Yearbook for Regional Economy |
| *ln Fixed Inv* | The amount of fixed investments | China Statistics Yearbook for Regional Economy |
| *ln Distance to Coast* | Ln (the nearest distance between a prefecture’s centroid to the coast) | National Geomatics Center of China |
| *Gradient* | the average gradient | National Geomatics Center of China |
| *River Density* | The length of river within the jurisdiction scaled by area | National Geomatics Center of China |
| *Land Size* | Ln (area of each prefecture) | China Statistics Yearbook for Regional Economy |
| *arg* | The ratio of agricultural inputs to total inputs in the production of an industry | China’s Input-Output Table in 2002 and 2007 |
| *ln AGRpc* | Ln (Agricultural output per capita) | China Statistics Yearbook for Regional Economy |
| *mine* | The ratio of mineral inputs to total inputs in the production of an industry | China’s Input-Output Table in 2002 and 2007 |
| *ln Minepc* | Ln (Mineral output per capita) | China Statistics Yearbook for Regional Economy |
| *Skill* | The share of population with at least secondary education in each prefecture | China Population Statistics Yearbook; China Statistics Yearbook |
| *Edu* | The ratio of employees with at least secondary education scaled by total number of employees in an industry | China Economic Census Yearbook |
| *Clan before1850* | Ln [(1+#Genealogies in or before 1850)/ population 2007] | The General Catalog of Chinese Genealogies |
| *Clan before1912* | Ln [(1+#Genealogies in or before 1912)/ population 2007] | The General Catalog of Chinese Genealogies |
| *Clan before1949* | Ln [(1+#Genealogies in or before 1949)/ population 2007] | The General Catalog of Chinese Genealogies |
| *Clan after 1976* | Ln [(1+#Genealogies after 1976)/ population 2007] | The General Catalog of Chinese Genealogies |
| *Jinshi* | Ln [(1+ #jinshi in Ming and Qing Dynasty)/ population 2007] | Distribution of Jinshi in Ming Dynasty; List of Jinshi in Qing Dynast |
| *Treaty Port* | A dummy, equal to 1 if the city was a treaty port from 1840s to 1910s and 0 otherwise. | Yan (1955) |
| *Catholic Church* | Ln (the number of Catholic churches) | Chinese Research Data Services |
| *Urbanization in 1920* | The proportion of the urban population in 1920 | Stauffer (1922) |
| *Non-Agr in 1964* | The ratio of the non-agricultural population to total population in 1964 | Almanac of China’s Population, 1988 |
| *ln Modern Firms’ Cap* | Ln (the amount of capital for modern firms over the period of 1840-1927) | Du (1991) |
| *ln NGO employees* | Ln (the number of people registered with the NGO in each prefecture) | Chinese Civil Affairs Statistical Yearbook |
| *ln Value of Donation* | Ln (the amount of donation) | Chinese Civil Affairs Statistical Yearbook |
| *Legal case* | Ln [1(+ total number of resolved civil and commercial cases over the period of 2003-2007) / population 2007] | The yearbooks annually compiled by each prefecture |
| *Migration* | Ln (1+#migrants settled in a prefecture during 1127-1130 southward migration) | Wu (1997), History of Migration in China  |
| *Migration\_surname* | Ln (1+# migrants of each surname settled in a prefecture during 1127-1130 southward migration) | Wu (1997), History of Migration in China  |
| *Operating profit ratio* | Operating profits / operating income | Chinese Annual Survey of Industrial Firms  |
| *Overhead expense ratio* | Overhead expenses / operating income | Chinese Annual Survey of Industrial Firms |
| *Clan Surname* | Ln (1+#Genealogies of the clan with the surname of an entrepreneur) | The General Catalog of Chinese Genealogies |
| *Leverage*  | Total liabilities / total assets | Chinese Annual Survey of Industrial Firms |
| *ROA* | Net profits / total assets | Chinese Annual Survey of Industrial Firms |
| *Fixed assets* | Net fixed assets / total assets | Chinese Annual Survey of Industrial Firms |
| *Size* | Ln (total assets) | Chinese Annual Survey of Industrial Firms |
| *Employee* | Ln (total number of employees) | Chinese Annual Survey of Industrial Firms |

**Table A2 Robustness: alternative measures of contract intensity and clans**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 　 | (1) | (2) | (3) | (4) | (5) |
| *China's Z × Clan* | 0.892\*\*\* |  |  |  |  |
|  | (0.163) |  |  |  |  |
| *Z × Clan before1850* |  | 0.981\*\*\* |  |  |  |
|  |  | (0.139) |  |  |  |
| *Z × Clan before1912* |  |  | 0.927\*\*\* |  |  |
|  |  |  | (0.140) |  |  |
| *Z × Clan before1949* |  |  |  | 0.744\*\*\* |  |
|  |  |  |  | (0.158) |  |
| *Z × Clan after 1949* |  |  |  |  | 1.245\*\*\* |
|  |  |  |  |  | (0.175) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| N | 9,240 | 9,240 | 9,240 | 9,240 | 9,240 |
| R-squared | 0.635 | 0.632 | 0.632 | 0.631 | 0.632 |

*Notes*: Column 1 reports the estimation results of using an alternative measure of China’s contract intensity. The data used to calculate the China-specific industrial contract intensity are collected from *China Economic Census Yearbook* and China’s *Input-Output Table*. Columns 2-5 report the estimation results by counting the number of genealogies before 1850, 1912, and 1949, and after 1949, respectively. Z refers to the contract intensity measure Z1 in Nunn (2007). The control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table A3 Robustness: Historical and geographic factors**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  History & Culture |  | Excluding 9 provinces  |  | Climate |
|  | *Z1* | *Z2* |  | *Z1* | *Z2* |  | *Z1* | *Z2* |
|  | (1) | (2) |  | (3) | (4) |  | (5) | (6) |
| *Z × Clan* | 1.242\*\*\* | 1.237\*\*\* | *Z × Clan* | 0.772\*\*\* | 0.896\*\*\* | *Z × Clan* | 0.748\*\*\* | 1.296\*\*\* |
|  | (0.198) | (0.274) |  | (0.202) | (0.277) |  | (0.179) | (0.232) |
| *Z × Jinshi* | -0.203 | -0.005 |  |  |  | *Z × Temperature* | -0.362\*\*\* | -0.139 |
|  | (0.229) | (0.302) |  |  |  |  | (0.116) | (0.156) |
| *Z × Treaty Port* | -0.591 | -1.220 |  |  |  | *Z × Rain* | 0.001\*\*\* | -0.000 |
|  | (0.605) | (0.846) |  |  |  |  | (0.00015) | (0.00019) |
| *Z × Catholic Church* | 0.268 | 0.511 |  |  |  | *Z × Sunshine* | -0.002 | -0.002 |
|  | (0.242) | (0.323) |  |  |  |  | (0.00121) | (0.00159) |
| *Z × Urbanization in 1920* | -1.464 | -6.185 |  |  |  |  |  |  |
|  | (3.844) | (4.464) |  |  |  |  |  |  |
| *Z × Non-Agr in 1964* | 5.990\*\* | 3.975 |  |  |  |  |  |  |
|  | (2.991) | (4.180) |  |  |  |  |  |  |
| *Z × ln Modern Firms’ Cap* | -0.0764 | -0.172 |  |  |  |  |  |  |
|  | (0.0998) | (0.126) |  |  |  |  |  |  |
| Controls | Yes | Yes |  | Yes | Yes |  | Yes | Yes |
| Prefecture FE | Yes | Yes |  | Yes | Yes |  | Yes | Yes |
| Industry FE | Yes | Yes |  | Yes | Yes |  | Yes | Yes |
| Observations | 7,896 | 7,896 |  | 6,468 | 6,468 |  | 9,240 | 9,240 |
| R-squared | 0.608 | 0.610 |  | 0.638 | 0.639 |  | 0.634 | 0.634 |

*Notes*: In columns 1-2, we additionally control for a set of historical factors, including the number of *jinshi* during Ming and Qing dynasties, *Treaty port* dummy indicating whether the city was a treaty port from 1840s to 1910s, the number of Catholic churches, the urbanization rate measured by the proportion of the urban population in 1920, the ratio of the non-agricultural population to total population in 1964, and the amount of capital for modern firms over the period of 1840-1927. Columns 3-4 report the estimation results of model (1) by excluding nine provinces in the southeast of China. In columns 5-6, we control for climate conditions. *Rainfall, Sunshine* and *temperature* are the average values during the period of 2000-2007. Other control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table A4 Robustness: the role of government**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 　 | Z1 | Z2 | Z1 | Z2 |
| 　 | (1) | (2) | (3) | (4) |
| *Z × Clan* | 1.091\*\*\* | 1.157\*\*\* | 1.135\*\*\* | 1.213\*\*\* |
|  | (0.15) | (0.19) | (0.151) | (0.197) |
| *Z × CSEZ*  | -1.889 | -8.092\*\*\* |  |  |
|  | (2.25) | (3.02) |  |  |
| *Z × ETDZ*  | -1.234\* | -1.953\* |  |  |
|  | (0.69) | (1.01) |  |  |
| *Z × HIDZ*  | 1.822\*\* | 0.90  |  |  |
|  | (0.71) | (1.00) |  |  |
| *Z × BECZ*  | -1.637 | -3.573\*\* |  |  |
|  | (1.16) | (1.60) |  |  |
| *Z × EPZ*  | -0.528 | -0.298 |  |  |
|  | (0.48) | (0.71) |  |  |
| *Z × BZ*  | -0.726 | -0.381 |  |  |
|  | (0.98) | (1.56) |  |  |
| *Z × ln (Fiscal Expenditure)* |  |  | -1.776\* | -2.008 |
|  |  |  | (0.939) | (1.284) |
| *Z × ln (Fiscal Transfer)* |  |  | 0.971 | -0.276 |
|  |  |  | (0.617) | (0.849) |
| Controls | Yes | Yes | Yes | Yes |
| Prefecture FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Observations | 9240  | 9240  | 9,128 | 9,128 |
| R-squared | 0.633  | 0.635  | 0.629 | 0.631 |

*Notes*:The definitions of special economic zones are from Alder et al. (2016). *CSEZ* is Comprehensive Special Economic Zones; *ETDZ* is Economic and Technological Development Zones; *HIDZ* is High-tech Industry Development Zones; BECZ is Border Economic Cooperation Zones; EPZ is Export Processing Zones; BZ is Bonded Zones. We count the number of each type of SEZ for each prefecture. Other control variables are the same as those in Table 3. All regressions include prefecture and industry fixed effects. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.

**Table A5 The relation between clans and 1127-1130 southward migration**

|  |  |  |  |
| --- | --- | --- | --- |
| 　 | Cross-sectional data | 　 | Prefecture-surname panel data |
|  | Full sample | Clan>0 | Migration>0 |  | Full sample | Clan>0 | Migration>0 |
| 　 | (1) | (2) | (3) | 　 | (4) | (5) | (6) |
| *Migration* | 0.7306\*\*\* | 0.6791\*\*\* | 0.4520\*\*\* |  |  |  |  |
|  | (0.0642) | (0.0591) | (0.1095) |  |  |  |  |
| *Migration\_surname* |  |  |  |  | 0.7399\*\*\* | 0.2586\*\*\* | 0.5119\*\*\* |
|  |  |  |  |  | (0.0444) | (0.0397) | (0.0392) |
| Controls | Yes | Yes | Yes |  | No | No | No |
| Prefecture FE | No | No | No |  | Yes | Yes | Yes |
| Surname FE | No | No | No |  | Yes | Yes | Yes |
| N | 330 | 272 | 95 |  | 200640 | 8935 | 57152 |
| R-squared | 0.525 | 0.508 | 0.477 | 　 | 0.273 | 0.423 | 0.411 |

*Notes:* This table reports the results on the relation between the strength of local clans and 1127-1130 southward migration. *Clan* and *Migration* refer to the number of genealogies and the number of historical migrants settled in a prefecture in columns 1-3, respectively. In columns 4-6, the dependent variable, *Clan,* refers to the number of genealogies for each surname in a prefecture. *Migration\_surname* is the number of migrants with a specific surname settled in a prefecture. The data of migrants are manually collected from *History of Migration in China* (Wu, 1997). Columns 2 and 5 use the sample in which the prefectures have stored at least one genealogy; columns 3 and 6 use the sample in which the prefectures have received at least one migrant in the migration wave. Robust standard errors are reported in parentheses. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01.



(A)



(B)

**Fig. A1** Distribution of migrants in the 1127–1130 migration wave

*Note*: Panels A and B represent the geography distribution of southward migrants in China and Southern Song, respectively.

**Appendix B：A Simple Model**

In this section, we present a partial equilibrium model to study the impact of clan culture on each region’s comparative advantage. To be simplified, we ignore the labor mobility across regions.

**B1. Preference and Market Structure**

Consumers in region has the access to a potentially different set of goods . We assume that a representative consumer in region has the constant-elasticity-of-substitution (CES) utility function:

where indexes varieties in the product set , is the quantity of variety from region demanded by the representative consumer in region and captures the elasticity of substitution between varieties. Then, consumer optimization yields the following demand for variety :

where is the price of variety , is an aggregate price index, and represents the total expenditure of region . To simplify notation, the subscripts, as well as the index for variety, are suppressed hereafter.

**B2. Producers**

A variety is produced by a firm with productivity () according to the following production function:

where , and are positive and . and respectively denote capital and labor inputs employed, and denotes the intermediate inputs bundle. The intermediates bundle is assembled by combining one bundle of continuum relationship-specific intermediate inputs, , and another bundle of intermediate inputs without specific relationship, , according to the following CES aggregator:

where is the elasticity of substitution. These component input bundles themselves are also CES aggregates:

where is the elasticity of substitution. deontes the set of inputs which is relationship specific; deontes the set of inputs which is not relationship specific. One unit of intermediate input or requires units of and units of . We use to denote the rental rate for capital and to represent the wage payment for unit labor. The rental rate and wage payment are exogenously determined.

For the relationship-specific inputs , after the production unit is formed, can only recover a fraction of the investment. The parameter captures quality of contract enforcement. Better institutions thus correspond to higher values of . In other words, if contracts and property rights are well-enforced, each agent will be able to recoup its ex-ante investment to a greater degree. In order to induce to form the production unit, it must be compensated with a share of the surplus, which is given by the revenue minus the ex-post opportunity costs of the factors . We adopt the assumption that ex post the parties reach a Nash bargaining solution and each receive one half of the surplus. Thus, will only enter the -good production if the individual rationality constraint, , holds. This implies that the price of satisfies .

The clan culture improves the producers’ contracting environment in several ways that are explained in Section 2. Therefore, is an increasing function of the intensity of clan culture () (i.e., ).[[21]](#footnote-21) That is, the material cost to produce the relationship-specific inputs, , is lower for the firms in the regions with the clan culture.

To produce the inputs , which are not relationship-specific, can recover its whole investment. Hence, the material cost to produce inputs without specific relationship satisfies .

The firm chooses labor , capital , and the amounts of intermediate inputs and , given the wage rate , the rental rate , and the prices of intermediate inputs and . Given the above production function and the firm’s productivity , the marginal cost of inputs when producing the final variety satisfies:

where is the price index for the intermediate inputs bundle . is the price index for the bundle of the relationship-specific intermediate inputs. is the price index for the bundle of intermediate inputs, which are not relationship specific.

**B3. Comparative Statics**

Given the marginal production cost, firm would maximize its profits as follows,

which implies that the selling revenue () is equal to:

By totally differentiating, the effect of the clan culture on the selling revenue is given by:

where corresponds to the degree of dependence on contract enforcement, with a larger value indicating a higher degree of contract intensity.[[22]](#footnote-22) The previous equation implies that:

Hence, we have the following proposition:

**Proposition 1**. *The firms in the sectors with higher contract intensity (i.e., higher ) sell relatively more in regions with strong clans.*

1. This quotation comes from *The Religion of China: Confucius and Taoism*, published in 1915. [↑](#footnote-ref-1)
2. Feenstra, Hong, Ma, and Spencer (2013) and He, Xue, and Zhou (2019) also find that contracting institutions affect the patterns of exports and industrial specializations across Chinese provinces. [↑](#footnote-ref-2)
3. The electric machinery industry has the highest score of contract intensity according to the calculation of Nunn (2007), that is, the industry most dependent on contracting institutions. [↑](#footnote-ref-3)
4. A nascent literature has examined the effects of various alternative informal arrangements on bilateral trade, such as network (Chaney, 2014; Rauch, 1999), experience (Araujo, Mion, and Ornelas, 2016), and cultural bias (Guiso, Sapienza, and Zingales, 2009). However, hitherto few studies have introduced informal institutional factors to explain the industrial specialization across countries or subnational regions. Among the few, Chisik (2003) captures a “reputational comparative advantage” and shows that the reputation of origin countries determines the type of products in which countries specialize; a more related paper, Cingano and Pinotti (2016), shows that high-trust regions tend to specialize in delegation-intensive industries in Italy. [↑](#footnote-ref-4)
5. To reinforce the argument, we provide a partial equilibrium model in Appendix B to demonstrate that firms in sectors dependent on contract enforcement gain a comparative advantage in regions with strong clans. [↑](#footnote-ref-5)
6. Jurchen tribes were a northern ethnic group affiliated with the Khitan Liao dynasty. In 1115, the Jurchens rebelled against the Liao dynasty (916–1125) and declared the formation of the Jin Dynasty (1115–1234). In 1127, Jurchens conquered the capital of Song and imprisoned two emperors and most of the imperial family, ending the Northern Song Dynasty (960–1127). Zhao Gou, the brother of the last emperor in Northern Song, together with several senior officials, surrendered the northern regions and founded Southern Song (1127–1279) with Lin-An (Hangzhou), a city in south China, as its capital. Because of fierce military conflicts in northern China at that time, around 5 million northerners fled south between 1127 and 1130 (Wu, 1997; Bai, 2022). [↑](#footnote-ref-6)
7. Our results throughout this paper are very robust to the data of other years, that is, from 1998 to 2006. [↑](#footnote-ref-7)
8. The measure is similar to the specialization index in Kalemli-Ozcan et al. (2003), who provide international evidence on the relation between industrial specialization and risk sharing. [↑](#footnote-ref-8)
9. The Chinese Annual Survey of Industrial Firms originally covered 40 industries in 2007. We exclude mining and public utilities industries. Our analysis includes four municipalities (namely, Beijing, Shanghai, Chongqing, Tianjin) and a prefecture (namely, Chaohu) that was abolished in 2010. [↑](#footnote-ref-9)
10. Nunn (2007) provides detailed illustrations of the methodology. Industries are categorized in accordance with the BEA’s 1997 I-O industry classification (381 industries) in Nunn’s data set, so we manually match the contract intensity data to China’s two-digit industrial classification for National Economic Activities (28 industries). Equal weights are used when a Chinese industry maps to more than one industry in Nunn’s study. [↑](#footnote-ref-10)
11. When taking the natural logarithm form, our original measure of industrial specialization, $SP\_{ip}={\frac{output\_{ip}}{output\_{p}}}/{\frac{1}{N}}\sum\_{q=1}^{N}\frac{output\_{iq}}{output\_{q}}$, is simplified to $ln⁡(output\_{ip}) $in our specification, as the prefecture fixed effect absorbs the factor $output\_{p}$ and the industry fixed effect absorbs the factor $\frac{1}{N}\sum\_{q=1}^{N}\frac{output\_{iq}}{output\_{q}}, q=1,2,..N$. Following Nunn (2007), we employ $ln⁡(1+output\_{ip})$ to avoid missing values. [↑](#footnote-ref-11)
12. In this setting, we explain the industrial specialization using a series of interactions between industry characteristics and regional characteristics. This research design is widely relied on to explore the source of a comparative advantage in cross-country studies. Rajan and Zingales (1998) primarily use the interaction term of industrial external finance dependence with the level of financial market development to examine whether finance can be a source of comparative advantage. Romalis (2004) uses a similar setting to study whether countries with abundant skilled labor concentrate their efforts on the export and production of skill-intensive goods. Nunn (2007) tests whether judicial quality generates a comparative advantage for contract-intensive industries in their exports. Recently, some studies extended the functional form to a subnational or firm-level investigation. Among studies on the Chinese economy, He et al. (2017) and Ding et al. (2018) test the role of financial development and political connections, respectively, in shaping a comparative advantage. [↑](#footnote-ref-12)
13. In this paper, we primarily use the modern population (year 2007) to normalize the number of genealogies. It is possible that the explosive growth of China’s economy in recent years has shaped the distribution of populations across regions. To ensure robustness, we use 1953 population to normalize the number of genealogies and find that all the regression results in this paper remain unchanged. Moreover, to address the concern that ASIF database might not capture the pattern of cities with a large number of small-size firms, we draw on the 2004 economic census data to calculate the specialization measure. Our results remain unchanged, including the benchmark, robustness, subsample analyses, and the instrumental variable (IV) approach. To save space, we do not tabulate tables which can be provided upon request. [↑](#footnote-ref-13)
14. We refer interested readers to He et al. (2019), who provide more details about the methodology. [↑](#footnote-ref-14)
15. In unreported tables, we simultaneously incorporate the legal case and clan interactions into model (1). The results show that, when clans are controlled, the role of formal contracting institutions in the industrial specialization pales. [↑](#footnote-ref-15)
16. Thanks to an anonymous reviewer for the valuable suggestion. [↑](#footnote-ref-16)
17. We are grateful to an anonymous reviewer for the insightful suggestion. [↑](#footnote-ref-17)
18. SOEs are defined as the firms with registration codes of 110 (state-controlled enterprise), 141 (state-controlled joint venture), and 151 (state-owned enterprise), and non-SOEs refer to left firms. [↑](#footnote-ref-18)
19. For example, Zhu Xi built the *Hanquan* Academy because his mother was buried there. Chen et al. (2021) and Cheng et al. (2021) show evidence that the distance measure has a negligible relation with the level of GDP, modern human capital, average financial development, commercial tax, population density, or urbanization. [↑](#footnote-ref-19)
20. In Appendix Table A5, we further show that migration could encourage the compilation of genealogies specific to the migrated clan. In other words, the number of migrants with a specific surname is related to the number of genealogies with this surname in the settled prefecture. [↑](#footnote-ref-20)
21. We provide the empirical evidence in Table 3. Using the number of civil and commercial legal cases to proxy for the quality of contractual environment, we find that intense clan culture significantly decreases the probability that firms engage in contractual disputes. [↑](#footnote-ref-21)
22. This is consistent with Nunn’s measure: $\frac{P\_{D}^{1-ς}}{P\_{D}^{1-ς}+P\_{Z}^{1-ς}}=\frac{P\_{D}^{1-ς}}{P\_{D}^{1-ς}+P\_{Z}^{1-ς}}×1+\frac{P\_{Z}^{1-ς}}{P\_{D}^{1-ς}+P\_{Z}^{1-ς}}×0$, where $\frac{P\_{D}^{1-ς}}{P\_{D}^{1-ς}+P\_{Z}^{1-ς}}$ denotes the expenditure share on the intermediates inputs whose production needs relationship-specific investments and $\frac{P\_{Z}^{1-ς}}{P\_{D}^{1-ς}+P\_{Z}^{1-ς}}$ reflects the expenditure share on the non-relationship-specific intermediates inputs whose production does not need relationship-specific investments. [↑](#footnote-ref-22)