北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

E2021004 2021-07-12

Clan Culture and Patterns of Industrial Specialization in China

Haichao Fan, Chang Li, Chang Xue, Miaojie Yu†

Abstract: A system of values composes historical traditions, which underly the fundamental institutions of a society. We show that the clan, an extant social organization formed 2,000 years ago, affects 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 are robust to alternative measures, controlling for the long-lived civil examination system (*Keju*), and instrumental variable estimations. The 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; migration
JEL codes: Z10, O10, R12

† **Haichao Fan**, Institute of World Economy, School of Economics, Fudan University, Shanghai, China and a research fellow at Shanghai Institute of International Finance and Economics, 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**, Department of Economics and Management, Wuhan University, Wuhan, China, Email: xuechang@whu.edu.cn. **Miaojie Yu**, China Center for Economic Research, Peking University, Beijing, China, Email: mjyu@ccer.pku.edu.cn.

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]

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; Allen, Qian, and Qian, 2005). Thus, institutional factors, such as the quality of 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) and Levchenko (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]

Notably, well-functioning formal contracting institutions are usually absent in emerging economies. Specifically, 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; Ding, Fan, and Lin, 2018). According to the Worldwide Governance Indicators (WGI), the index of "rule of law" ranks China at approximately 44

[1] This quotation comes from *The Religion of China: Confucius and Taoism*, published in 1915.

[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.

of 227 countries in 2017. While the coexistence of both China's unprecedent economic growth and weak contracting institutions seems to be a contradiction, Acemoglu and Johnson (2005) have demonstrated that economies could function well in the absence of strong contracting institutions. They explain that private parties have resources to change the terms of formal contracts and develop alternative informal arrangements to 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. For example, in most of the developing countries characterized with collectivist societies, contract enforcement is mainly achieved through informal institutions (Greif, 1994).[3]

In this paper, we investigate the effect of clans, an incredibly important social organization built on kinship in ancient China, on today's pattern of industrial specialization across Chinese prefectures. Our study is inspired by the role of clans in premodern society. 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 doled out financial provisions and informational resources to their members, constructed infrastructures, collected tax, and educated young to approach the officialdom. In particular, clans were known to establish their own rules to manage their daily affairs and settle conflicts with outsiders; that is said, 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;

[3] Several informal institutions have been captured by the recent work, such as reputation (Macleod, 2007), social network (Chandrasekhar, Kinnan, and Larreguy, 2018), or political connections (Ding, Fan, and Lin, 2018), as alternative mechanisms of contract enforcement.

Faure, 2007; Greif and Tabellini, 2017). Such an important social organization has survived after approximately one hundred years of extreme social changes, and in some regions, proven to flourish (Tsai, 2007).

Our research goals are twofold. First, we aim to examine whether the strength of local clans has a causal relation with regional industrial development and specialization. Second, conditioning on the established relation, we endeavor to explore whether the effect of clans stems from a general improvement in the local contracting institution or from a specific relation between individual firms and local clans. Existing studies have paid limited attention to this traditional cultural tradition. Guiso, Herrera, and Morelli (2016) show that cultural norms generally affect economic outcomes, though their findings are not specific to a certain type of cultural tradition. Few studies examining the clan mainly focus on the comparison of institutions and culture between China and Europe, the history and evolution of two different societies (Greif and Tabellini, 2017), historical correlates, (Shiue, 2017), or the public goods provision and old-age support in modern society (Xu and Yao, 2015; Zhang, 2019). The most related paper to ours is Zhang (2020), who 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 Chinese private sector.

To scrutinize the role of clans in regional specialization, we collect the 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 to measure the strength of local clans. The genealogy is a book that records the important information of a clan, including their pedigrees from the oldest common ancestors, marriages,

offspring, and excellent achievements of members, as well as the public wealth of a clan. 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. According to our novel data set, genealogy records exist for 283 of 330 prefectures in mainland China, of which 71 prefectures have more than 100 editions of their genealogies.

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 civil and commercial cases represents a higher probability of a broken contract, as well as a poor contracting environment. We find evidence of a significantly 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 improving the contracting institutions.[4]

The empirical findings are virtually unchanged when we count the number of genealogies compiled before different time points or measure the strength of local clans using survey question in China Family Panel Studies (CFPS) and when we construct a measure of China-specific contract intensity using China's Input-Output table. We also control for an important historical correlate, the number of *Jinshi* during the Ming and Qing dynasties, as the success of *Keju* has a persistent effect on human capital today (Chen et al., 2020). The strength

[4] To reinforce the argument, we provide a partial equilibrium model in the Appendix to demonstrate that firms in the sectors dependent on contract enforcement gain a comparative advantage in regions with strong clans.

of clans continues to be an important source of a comparative advantage for firms after the inclusion of *Jinshi* density.

We further corroborate that the present pattern of industrial specialization is driven by the force of market incorporating 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 having a trivial effect on the industrial development in the 1980s.

Clans are not randomly distributed across regions. We cannot exhaustively control for all of the vital variables, the omission of which may bias our estimates. To address endogeneity concerns, we employ an instrumental variable approach and use the number of migrants who settled in each prefecture during the 1127–1130 southward migration as our instrument. Historical migration was always organized by family groups involving a number of patrilineal kinsmen (Pasternak, 1969). The rationale undergirding our approach is that the already-fierce competition between newly migrated and native clans intensified clan identity and hence increased the aggregate strengths of clans in the host regions (Bai, 2019). Our results show that the distributions of genealogies and migrants across regions are highly correlated, and the relation even holds at the prefecture-surname level. More importantly, two-stage least squares (2SLS) regressions produce results consistent with those of our benchmark regressions. Similarities in the results reinforce our argument that the strength of local clans is a source of comparative advantage for industries whose production is heavily dependent on a good contractual environment.

Finally, we explore the potential mechanisms by which clan culture creates a comparative advantage for the development of contract-intensive industries. The positive effect of clan culture 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 to 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.

Our work contributes to the literature in several ways. First, our findings provide new insight into the determinants of industrial specialization and unbalanced regional economic development. Prior studies have captured several important factors that affect the agglomeration of industries in countries or in regions within a country, 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), economic policy (Zheng and Kuroda, 2013), the efficiency of legal system (He et al., 2019), and transport network (Jaimovich, 2019). The present study contributes to this strand of literature by showing that clans, a special cultural tradition of China, imbue contract-intensive industries with a long-run regional comparative advantage and affect the pattern of industrial specialization in the

most-representative emerging economy around the world. 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., Dell, 2010; Jia, 2014; Guiso, Sapienza, and Zingales, 2016; Chen, Kung, and Ma, 2020), 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 posits why clans matter for industrial specialization; namely, they act as an informal contracting institution. Section 3 describes the data and research designs. Section 4 presents the empirical results. Section 5 concludes.

2. Background

Clans in China are kin-based groups consisting of agnate households (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 (Cohen, 1990; Szonyi, 2002; 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).

Informal institutions refer to informal constraints encompassing norms of behavior, conventions, and self-imposed codes of conduct (North, 1990). We posit that clans, a long-lived domineering institution of premodern China, are very likely to affect norms and beliefs, even nowadays, and to matter for the economy as an informal arrangement, though the institution itself has distinctively declined. Some recent work has shown that clans have persistent impacts on several aspects of modern Chinese society. Su et al. (2011) document that China's rural democracy has reactivated traditional social ties in villages and in networks associated with clans, and this reactivation has mobilized villagers to attend voting. Xu and Yao (2015) find that, if the leader of a village comes from two of the largest local clans, public investments will be considerably promoted. Zhang (2020) finds that the strength of local clans is positively associated with local entrepreneurship and the prosperity of private sectors.

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 cultural norms associated with clans affect private parties' ethical standards and increase the economic and noneconomic

costs of default, the contractual environment should be improved. Through this lens, we posit that clans could shape contracting institutions in several ways and thereby affect industrial specialization.

First, clans are a social network through which information should be dispersedly and effectively transmitted. Specifically, if an entity breaks the 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 counterparty would be reduced. Therefore, the increase in opportunity costs associated with information transmission in the clan can largely impede opportunistic behavior locally and promote contract enforcement between private parties.

Second, clans are characterized by strong collectivism among clan members (Greif and Tabellini, 2013). In contrast to Western corporations, which emphasize individualism and place greater weight on individual benefits or losses, clans rely on collectivism as an underlying principle. In ancient China, clans functioned to protect their members from government or outside clan disputes. In this sense, the opportunistic behaviors of outsiders would be considered to damage the interests of the whole clan, which, in turn, create substantial obstacles for outsiders who are doing business in the located region. Meanwhile, a clan member is less likely to engage in opportunistic behavior for two reasons: (1) opportunistic behavior may generate a negative externality on the reputation of the belonged clan, and (2) strong clans discipline members who violate the rules. More explicitly, Liu (1959) argues that "a clan should always watch its reputation by preventing its members from harming outsiders and by refusing its offenders clan protection." In particular, clan member identity is inherently determined, and the clan lineage, that is, a member's

relationship to the clan, is rarely erased. Taken together, the clan background of private parties, either the defaulter or its counterpart, would increase the costs of opportunism and facilitate the process of enforcing contracts.

Lastly, the cultural norms of strong clans may be internalized in the spiritual world of their members, as well as local people, in the long run. The point of cultural trait differs from the two above-mentioned points relating to the economic costs of opportunistic behaviors, in order to emphasize the moral motive. 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. Economic entities with a strong clan background instinctively discriminate against opportunistic behaviors and thus are less likely to engage in opportunistic behavior themselves. Both the moral and economic motives behind clan culture are expected to improve contracting institutions, but they create a discrete comparative advantage for firms. If the economic motive dominates, we should observe that the benefits are specific to firms having connections with local clans; on the contrary, if the moral motive dominates, clans shape the overall contractual environment at the regional level and hence impose a general effect on firms in the same region. We will further discuss the issues and test the predictions in Section 4.4.

The next question for our empirical investigation is how to quantify the strength of a clan. Following Greif and Tabellini (2017), Chen et al. (2020), and Zhang (2020), we measure the intensity of the clan using the cumulative number of genealogies normalized by the local population. As a form of social organization, clans rely on specific activities to embody and strengthen the common identity of their members. The rites to ancestors, setup of ancestral halls, and compilation of genealogies are typical clan activities, which all center

around the common ancestors and a clan's lineage. Genealogies are books recording the history of clans, including their pedigrees from the oldest common ancestors, marriages, offspring, and excellent achievements of members, as well as the public wealth of clans, such as land, ancestral graves, and halls. Since the compilation of genealogies serves as a key component and a record of clan activities (Feng, 2011, 2013), the large number of genealogies proxies for frequent historical activities of local clans and particularly an intensity of clan tradition.

3. Data and empirical designs

3.1 Key variables

Industrial specialization. To measure the degree of specific industry specialization in each prefecture, we borrow the idea from the international economics literature. For example, Frésard, Hege and Phillips (2017) analyze the importance of industrial specialization in cross-border investments. They define industrial specialization as the proportion of a specific industry's output in the country's total production scaled by the average proportion of the industry's output in the world. Similarly, we define the specialization of industry i in prefecture p as $SP_{ip} = \frac{output_{ip}}{output_p} \Big/ \frac{1}{N}\sum_{q=1}^{N} \frac{output_{iq}}{output_q}, q = 1,2,\dots,N$, where N is the number of prefectures.[5]

We collect disaggregated firm-level data from the **2007** Chinese Annual Survey of Industrial Firms (ASIF) to calculate the specialization index. This database has been widely used in prior studies on the Chinese economy (see, e.g., Huang, Li, Ma and Xu, 2017; Fan, Li and Yeaple., 2018). By 2007, this

[5] 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.

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 have been excluded from 2008 (2011) onward.

To ensure a relatively complete coverage of industrial output, we choose year **2007** as our benchmark sample period.[6] $Output_{ip}$ is defined as gross output from all manufacturing firms in industry i in prefecture p. The sample 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.[7]

Clan culture. In premodern China, clans were kinship-based social groups consisting of members with a common patrilineal ancestor. Like other social groups or communities, 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, and marriage, offspring, and excellent achievements, as well as public wealth, like land, ancestral graves, and halls. 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

[6] Our results throughout this paper are very robust to the data of other years, that is, from 1998 to 2006.
[7] The Chinese Annual Survey of Industrial Firms originally covered 40 industries in 2007. We further 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.

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, $Clan = ln[(1 + \#geneologies) / 2007\ popultation]$. As the catalog contains genealogies compiled by 2004, we also employ the number of genealogy editions before 1949, 1912, and 1850 as alternative measures. Fig. 1 depicts the distribution of genealogies over different periods. The most-active period of genealogy compilation occurs from 1850 to 1949, during which approximately 63% of genealogies are recorded.

[Place Figure 1 about here]

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 provide a good environment to investigate how clans shape the pattern of industrial specialization (see Fig. 2).

[Place Figure 2 about here]

Industrial contract intensity. We employ the index of contract intensity constructed by Nunn (2007) to measure the level of industry dependence on the contracting 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 $z_i =$

$\sum_k \delta_{ik} R_k$, where δ_{ik} is the share of the value of input k used in the production of industry i, and R_k is the share of input k that is relationship specific. Two measures are provided: ${z_i}^1$ and ${z_i}^2$. 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.[8] 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.

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:

$$SP_{ip} = \beta_1 z_i Clan_p + \beta_2 z_i lnPop_p + \beta_3 z_i \ lnGDP\ pc_p + \boldsymbol{\beta_4 C_p} \cdot \boldsymbol{c_i} + \gamma_i + \delta_p + \varepsilon_{ip}, \quad (1)$$

where *i* and *p* corresponds to industry and prefecture, respectively. SP_{ip} is the logarithm of gross output of industry *i* in prefecture *p*.[9] $Clan_p$ is the logarithm of the cumulative genealogy number per 10,000 persons at prefecture *p*; z_i is the measure of contract intensity, namely, the industrial dependence on contracting institutions. $lnPop$ and $lnGDPpc$ are the logarithm of population and gross domestic product (GDP) per capita in each prefecture. Like Nunn (2007) and He et al. (2017), we include several interactions between regional

[8] Nunn (2007) provides detailed illustrations of the methodology.

[9] 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.

factor endowments and their intensities in the production of goods in an industry. $\boldsymbol{c_i}$ denotes the vector of industrial intensities, including agr_i, $mine_i$, and Edu_i, which corresponds to the dependence on agricultural inputs, mining inputs, and human capital in the production of industry *i*, respectively. $\boldsymbol{C_p}$ denotes the vector consisting of $lnAGRpc_p$, $lnMinepc_p$, and $Skill_p$, 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 scaled by 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 scaled by the total number of employees in the industry. The endowments $lnAGRpc_p$ and $lnMinepc_p$ are measured by the per capita output value of agriculture and mining in each prefecture. Human capital endowment, $Skill_p$, is the share of the population with at least a secondary education in each prefecture. We include industry and prefecture fixed effects, γ_i and δ_p, in all regressions. Table A1 in the Appendix A defines the variables and summarizes the data sources.

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 skilled-intensive goods. Nunn (2007) tests whether judicial

quality generates a comparative advantage for contract-intensive industries in their exports. Recently, some studies extend 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.

In the present study, $z_i Clan_p$ denotes the interaction between the strength of clans, an informal institution, and industrial dependence on contract enforcement. The positive coefficient for β_1 implies that in China, prefectures with strong clans tend to specialize in industries with higher contract intensity. On the contrary, a negative sign implies that strong local clans impede the development of contract-intensive industries.

Considering that the historical tradition of clans is predetermined, our analysis is less vulnerable to a reverse causality concern. Nevertheless, we ensure the robustness of the estimation results by using the number of genealogies compiled prior to several time points in history and formally tackling potential issues of endogeneity using an instrumental variable approach.

3.3 Preliminary patterns

Before the rigorous regression estimations, we begin with some suggestive evidence. We first show a roughly positive relation 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 1.

[Place Figure 3 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 dependent on contracting enforcement. Next, we will further control for regional and industrial characteristics that have been shown to affect the industrial development in the following multivariate setting.

[Table 1 is about here]

4. Empirical results

4.1 Baseline results

Table 2 reports the OLS regression results for model (1). The first column estimates the effect of the interaction between clan and contract intensity (z^I) on the pattern of industrial specialization without the inclusion of other endowment intensities and controls. Our balanced panel data of 9,240 observations is made up of 28 industries and 330 prefectures. The estimated coefficient for $z_i \times Clan_p$ 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 other factor endowment interactions that may influence industry agglomeration, that is, $agr_i \times lnAGRpc_p$, $mine_i \times lnMinepc_p$, and $Edu_i \times Skill_p$, 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 south China, 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 within the "food production" industry would increase by 17.81%, and the share of food production to total output would increase by 296 basis points. Also, agricultural, 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 a comparative advantage for production.

In addition to the endowment interactions, we control for other factors interacted with industrial contract intensity that may bias the importance of clans in shaping the contractual environment. We first interact contract intensity with population to control for the possibility that regions with demographic dividends specialize in certain labor-intensive industries. The positive sign before the estimated coefficient implies that the greater the population, the more developed the contract-intensive industries. Then we include the interaction of GDP per capita to control for the possibility that high-income regions, usually with a strong institutional framework, attract more investments with a high level of reliance on contracting institutions. The outcome matches this conjecture.

[Table 2 is about here]

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 ($28 \times (330 - 58)$). Consistently, we find that strong clans could promote the agglomeration of contract-intensive industries at the prefecture level. Moreover, we adopt another measure of contract intensity from Nunn (2007), z^2, which has a broader definition of relationship-specific investments, and our new estimates, which are reported in Columns 4–6, are consistent with those obtained using the narrow measure, z^1.

In Section 2, we demonstrated that clan culture substantially increases the costs of opportunistic behaviors, and, thereby, firms 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 the contracting environment. By manually collecting the data from statistical or comprehensive yearbooks that are compiled by each prefecture on an annual basis, we count the number of civil and commercial cases ($legal\ \ case$) over the period of 2003–2007 in each prefecture. We roughly assume that a large number of civil and commercial cases represents a higher probability of a broken contract, as well as a poor contracting environment.

Column 1 of Table 3 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. That is said, clans, as a special purveyor of social capital in premodern China, act as informal but

effective middlemen to handle contractual issues. We then regress the industrial output on the interaction between industrial contract intensity and the predicted value of $legal\ case$ 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.[10] The evidence supports our prediction that clans promote the agglomeration of contract-intensive industries by improving the contracting institutions. To deepen the understanding of our empirical results, we also present a simple partial equilibrium model in Appendix B, which predict that firms in the sectors dependent on contract enforcement gain a comparative advantage in regions with strong clans.

[Table 3 is about here]

4.2 Robustness checks

4.2.1 Alternative measures of contract intensity

We now test the sensitivity of our results to an alternative measure of contract intensity. Nunn's contract intensity measures are constructed based on 1997 United States *Input-Output Table*. However, the supply-chain relationship and the share of each input in the production of specific final goods may be different in China. To better fit production by Chinese firms, we construct a China-specific contract intensity measure as a robustness check.[11] However, data on whether inputs are sold on an exchange or reference priced are not available in China. Alternatively, we treat the inputs that lack the data of wholesale or retail sales as relationship specific. Wholesale and retail inputs imply a substantial number of purchasers and sellers, which limits the ability of

[10] In unreported tables, we also 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 industrial specialization pales.

[11] We refer interested readers to He et al. (2019), who provide more details about the methodology.

firms to negotiate on price. In the spirit of Nunn's idea, we define China's contract intensity as $China's\ Z = \sum_k \delta_{ik} S_k$, where δ_{ik} follows the same definition, and S_k is the share of inputs sold neither by wholesale nor by retailers in industry k. The data are collected from the China Economic Census Yearbook and China's Input-Output table.[12] We re-estimate model (1) for our full sample and for the subsample of cities with nonzero genealogy. Table 4 reports the results. The adoption of the new contract intensity does not alter our estimates: the coefficients for $z_i \times Clan_p$ are statistically significant and positive.

[Table 4 is about here]

4.2.2 Alternative measures of clan strength

One potential concern in our measurement of 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. Violent conflict, 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. Moreover, during 1966–1976, the Great Cultural Revolution 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). The disorder during these periods clearly affected the tendency and ability of local people to compile genealogies.

To check the robustness of our results, we separately employ the number of genealogies written before the years 1949, 1912, and 1850 to measure the

[12] We use the average value of whole and retail data from the *China Economic Census Yearbook 2004* and *2008* to calculate $China's\ Z$.

local intensity of clans. The People's Republic of China and Republic of China (PRC) 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. We also focus on the clans complied after 1978, as the spatial distribution of clans might have changed given that such traditional activities were nearly banned by the Chinese Communist Party during the 1949-78 period. Table 5 reports the results of four alternative clan variables for the full sample (columns 1–4) and for the subsample of observations with nonzero genealogies (columns 5–8). The estimations are qualitatively the same but exhibit larger economic magnitudes. Thus, the strength of local clans consistently provides industries that require intensive relationship-specific investments a comparative advantage.

[Table 5 is about here]

Though *General Catalog of Chinese Genealogy* is the largest collection of genealogies so far, the measurement error resulting from record errors still exists. We focus on a survey question from China Family Panel Studies (CFPS): "*does your family have the genealogy?*", and construct an alternative measure of clan intensity.[13] Specifically, we define the variable *Clan_survey* as the number of answers with "Yes" scaled by the total number of surveyed people in this region. We re-estimate model (1) with this new variable to proxy for clan strength and report the results in Appendix Table A2. We continue to find that the strength of local clans creates a comparative advantage for the development of contract-intensive sectors.

[13] CFPS is a nationwide social survey launched by Peking University. The questionnaires cover a variety of topics, such as migration, history, education, marriage, employment, risk attitude, etc. CFPS has been widely used in the studies on Chinese economy (e.g., Yi, 2019; Jia, Lan, and Miqueld, 2021).

4.2.3 Controlling for the success of Keju

Historical control variables are not controlled for in the benchmark analysis, which may bias the importance of clans in shaping industrial specialization. But historical institutions have been shown to exert a persistent effect on several aspects of economic development, even nowadays. In this subsection, we consider the success of *Keju*, that is, China's imperial examination system for civil service before 1905, across prefectures to check the robustness of our results. 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 affects patterns of industrial specialization. Therefore, the omission of historical education can result in an omitted variables problem. Moreover, local clans and the success of *Keju* are influenced by one another. Sui and Tang Emperors established and promulgated the civil examinations, partly in fear of powerful clans (Elman, 1991). The emperors preferred to select their officials who passed a civil examination, rather than have officials be personally recommended by powerful clans. However, a powerful clan could provide enormous educational resources to their members, enabling them to receive systematic education for civil examinations and accordingly achieve a higher probability of success (Greif and Tabellini, 2017). Chen et al. (2020) find a positive relation between *jinshi* density and the number of compiled genealogies, implying that quite a few members entered officialdom through *Keju* and that doing so is an effective channel for maintaining local clan influence and intensifying the sense of belonging and honor of clansmen.

Following Chen et al. (2020), we use the total number of *jinshi* during the Ming and Qing dynasties normalized by the 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*. *Jinshi* is the highest degree that can be obtained in the civil examination system. In Column 1 of Table 6, we include the interaction of *jinshi* with contract intensity instead of our clan interaction. The estimation shows that the density of *jinshi* significantly promotes the development of contract-intensive industries. Taking a step further, we simultaneously incorporate the *jinshi* and clan interactions into model (1). Column 2 shows that, the estimated coefficient for the *jinshi* interaction loses significance, whereas our key interaction, $z_i \times Clan_p$, again obtains a significantly positive coefficient. This being said, the effect of *Keju* success can be largely attributed to powerful local clans. In addition, when we remove the observations of cities without any *jinshi* in Column 3, the intensity of clans continues to be an important advantage for industries dependent on contract enforcement. Likewise, we employ the second measure of contract intensity, z^2, to perform this series of robustness checks, the results of which appear in last three columns of Table 6. As shown, the estimations are qualitatively similar. Taken together, the results in Table 5 indicate that the positive effect of clan culture on the agglomeration of contract-intensive industries does not change after controlling for the confounding historical factor.

[Table 6 is about here]

4.2.4 Role of clans before the market reform

One could argue that predetermined clans already shaped the industrial layout decades ago, even before the foundation of People's Republic of China in 1949, and the pattern persists today. In this context, the observed relation between regional clan intensity and current industrial specialization is 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 firms enjoy more freedom to incorporate many factors into this decision in a market economy. The reform and opening up policy, in place since 1978, jumpstarted China's explosive economic growth over the past 40 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 system as the end point of China's planned economy.

To check whether or not the pattern is persistent, we turn to the Second Industrial Census carried out in 1985, when the reform for urban and industrial sectors just started. The industrial layout during that time can be still considered 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. Industry data from the 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 7 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 the concentration of contract-intensive industries in the 2007 subsample. Therefore, we conclude that the pattern of current industrial specialization across prefectures is a process of evolvement after the economic reform, in which the distribution of clans plays an important role.

[Table 7 is about here]

4.3. An instrumental variable approach: Southward migration during 1127–

1130

A prefecture's industrial specialization is likely associated with a wide variety of factors that cannot be exhaustively controlled for. Thus, concerns that omitted variables are simultaneously associated with regional industrial specialization and the intensity of local clans should be further addressed to establish a causal identification. To mitigate this concern, we employ an instrumental variable, the southward migration of Han during 1127–1130, which was driven by the Jurchen's invasion and Northern Song's collapse.

The rationale of this instrument stems from the essential value that supports the kin-based collectivism of clans. In a sharp contrast to the individualism emphasized by Western corporations, collectivism guides clan members to protect and develop the interests of the clan as a whole. In particular, external shocks could intensify the motives of clansmen. By its very nature, kin-based clans could not freely allow members to join or voluntarily quit the group. When faced with fierce competition for outside resources, clans naturally made efforts to strengthen the common identity of the clan and tended to host clan activities more frequently, like revising or compiling genealogies (Greif and Tabellini, 2017). Therefore, the migration, inevitably intensifying competition between the migrated and native clans, acts as a proper external shock.

Both migrated or native clans make every effort to win more fundamental resources for their survival, such as water and farmland, which indirectly enhance the sense of belonging of clansmen. Competition between the two groups results in two actions that enhance their local influence: (a) quantity, more children for an increased clan size, and (b) quality, investing more resources in education for the future success of *Keju* and governmental

connections (Bai, 2019). Both actions would be positively associated with the length and the number of editions of a clan's genealogy.

The Yellow River Basin is the birthplace of Han nationality. The majority of Han settled in the lower and middle reaches of the Yellow River, which were frequently invaded by northern nomads. As a result, a mass of people migrated to southern areas that were less touched by wars, for example, the regions along Yangtze River or Pearl River, when great conflicts and wars broke out. The southward migration involves civilians, officials, and even the imperial court.

Premodern China experienced three waves of great southward migration, driven by the Revolt of the Five Barbarians (in Chinese, *wu hu luan hua*), An-Lushan Rebellion in Tang Dynasty, and the Jurchens' invasion during the 1100s, respectively.[14] In our study, we focus on the third wave of migration, that is, the 1127–1130 migration following the collapse of the Northern Song Dynasty, for several reasons.[15] First, many of the southward migrants in this wave were imperial kinsmen, noblemen, and officials of great political power and wealth, who could easily settle and form influential clans in their host regions. Also, plenty of large northern clans followed the imperial movement to locate in southern regions, particularly near Lin-An. The migration involving powerful clans and imperial court undoubtedly intensified the competition between natives and migrants for resources, such as land, water, or the opportunities of

[14] 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 north China at that time, around 5 million northerners fled south between 1127 and 1130 (Wu, 1997; Bai, 2019).

[15] The south-north division of China, between Southern *Song* and the Jurchens' *Jin*, lasted about 150 years until the Mongols' conquest in 1279. Hence, the migrated clans were less likely to move back. Moreover, China has not experienced the 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.

promotion in officialdom after *Keju*. For example, Bai (2019) finds that this wave of migration promoted native and migrated clans to make substantial investments in education to offset the intensified competition.

Second, the economic center moved from north to south after the second wave (Maddison, 2007), so the third wave of migration was less likely to affect the economic development of the south in the same way that the former two did. Before the second wave, the south was sparsely populated; for example, only 19% of the population lived in the south around year 2 AD (Wu, 1997). The migration brought not only a substantial amount of labor but also agricultural technology to settlements, which promoted the rapid economic development of southern regions. However, the economic development and population density of south China prior to the third wave was very different. In 1102, before the Jurchens' invasion, the number of households in north and south China was 5.92 and 12.20 million, respectively, and the great influx of migrants was approximately 10% of the native population (Wu, 1997, 2000).[16] The scale of migration in this wave clearly intensified competition but made less of an affect in shaping the economic composition nationwide.

Third, this wave of migration could be interpreted as government-organized activity and the choice of settlement by migrants was exogenously influenced by the government. After the Jurchens took away two emperors to *Jin* Dynasty's capital, Zhao Gou immediately re-founded *Song* as the new emperor in *Yingtian*, a city near *Kaifeng* (the capital of Northern *Song*) in Henan province, supported by the remaining *Song* forces in the north. However, continuous wars between the Jurchens and Southern Song compelled the new emperor to retreat from *Yingtian* to *Yangzhou*, *Ningbo*, and *Wenzhou*, and he

[16] About 50 million people make up 12.20 million households.

finally built the capital at *Lin-An* (*Hangzhou* today). However, *Jiankang* (*Nanjing* today), as the capital of former six southern dynasties during AD 222–589, was generally regarded as a better choice for the new dynasty. Zhao Gou, the cowardly new emperor, insisted on the choice of Lin-An, which was relatively far away from the battlefront (Chen, 1948). During the process of the southward retreat, many powerful persons, such as imperial family members, officials, and soldiers serving the emperor, just followed the emperor's movements and settled in the south.

The information on the number of migrants settled in each prefecture is obtained from the *History of Migration in China* (Wu, 1997), the compilation of which is based on a large number of history classics, especially the *History of Song*, literary works, and personal biographies. The series of books is the most complete and systematic collection of historical migration in ancient China. Volume 4 explicitly records the names of migrated ancestors, sources and settlement locations, migration paths, and information sources. By matching the city name in the Song Dynasty with present jurisdictions, we manually count the number of migrants who settled into each prefecture during this wave of migration. In our data set, 96 of 330 prefectures received migrants between 1127 and 1130. As the migration was prompted by the invasion of the Jurchens from north, southern cities around the new capital of Hangzhou should be the most appealing settlement places. Fig. 4 depicts the migrants' geographical distribution. Our statistics suggest that Hangzhou, Taizhou, and Ningbo received 10.8%, 7.7%, and 5.9% of the recorded migrants.[17]

[Place Figure 4 about here]

[17] Wu's (1997) record of only 1,367 migrants raises concerns about whether the sample is representative for this wave of migration. In fact, a migrant may present a movement of a family, a clan, or even a village. Bai (2019) addresses the issue of representativeness by showing a significant and positive relation between the increase in the share of northern surnames and the number of recorded migrants.

We first confirm the relevance of our instrument to clan strength in Table 8. In column 1, we regress the number of genealogies on the number of migrants for all prefectures and find them to be highly correlated. Then we turn to prefectures with above-zero genealogies and those with above-zero migrants in columns 2 and 3, respectively. The positive relation remains significant. Next, we examine whether migration encourages the compilation of genealogies specific to the migrated clans, that is, whether the number of migrants with a specific surname is related to the number of genealogies with this surname in the settlement location. Results in columns 4–6 confirm a positive and significant relationship. Taken together, the hypothesized importance of southward migration to the intensity of local clans is valid.[18]

[Table 8 is about here]

Table 9 reports 2SLS estimations for the full sample, followed by two subsamples with above-zero genealogies and above-zero migrants. F-statistics and Kleibergen-Paap rk LM statistics suggest that our instrument does not suffer from a weak instrument or an under-identification problem. More importantly, the second-stage estimation results are consistent with our baseline findings. Using the number of southward migrants from 1127 to 1130 as the instrument, we find that the intensity of local clans could significantly explain the patterns of industrial specialization across regions. The estimated coefficient for the instrumented clan interaction (1.3524) is slightly larger than the OLS estimated one (1.2903), implying that the endogenous genealogy variable underestimates the long-run effect of clan culture on the pattern of industrial specialization. The

[18] Before performing 2SLS estimations, we show that migration influences our outcome variable only through its effect on the intensity of local clans. In an unreported table, we first regress regional industrial specialization on the interaction of migration with contract intensity and then include our key clan interaction as well. As expected, the coefficient for the instrument interaction loses significance when two interactions are incorporated into the estimations.

2SLS results 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]

4.4. Firm-level analysis

We now examine how the intensity of local clans encourages the development of contract-intensive industries. As discussed in Section 2, clan culture shapes the contractual environment through motivating private parties to eschew opportunistic behavior. On the one hand, clans, as special purveyors, increase the opportunity costs of default as the information will be effectively transmitted through the whole network; in the spirit of collectivism, clansmen are less likely to either engage in opportunistic behavior or do business with outsiders who have broken a contract with other members of their clan. Here, the benefit of a strong clan for contract enforcement tends to concentrate on firms that have a specific relationship with a powerful clan. On the other hand, the beliefs and values held by local people can be viewed as internalized customs passed on from the long-lived clans. In this way, clans shape the overall contractual environment at the regional level, raise ethical standards in the long run, and generally protect local firms dependent on relationship-specific investments from default.

To differentiate between those two explanations, we perform a firm-level analysis and create a new clan measure, $Clan\ \ Suranme$, 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 a prefecture. 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:

$$Firm\ \ Performance_{ijp} = \beta_1 Clan\ \ Suranme_{ip} + \beta_2 z_j Clan\ \ Suranme_{ip} + \beta_3 z_j Clan_p + \gamma FirmControl + \rho_i + \sigma_j + \delta_p + \varepsilon_{ijp}, \quad (2)$$

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 to 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 β_2 and an insignificant β_3; 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 β_3 should dominate that of β_2.

Table 10 reports the results of the effect of clans on firms' profitability. In column 1, we first include $Clan\ \ Suranme_{ip}$ 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 in a general manner. However, the pattern changes when 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.[19] Our key clan interaction term in model (1) maintains its

[19] With our new sample, we aim to show that the causal relationship between the intensity of clans and the pattern of industrial specialization should be established in different subsamples.

significance, ensuring a positive effect of an overall intensity of local clans on the industrial specialization. More interestingly and importantly, 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 tells 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.

[Table 10 is about here]

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 some business and entertainment activities. In particular, 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 substantial overhead expenses unrelated to 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.

Table 11 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—

an overall clan intensity and a specific clan intensity—into our model. The patterns are almost the same to those observed in Table 10. 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 of clans is not confined to firms with a clan relationship; instead, the intensity of clans could generate a comparative advantage for all firms in this region and enhance their operating performance.

Taken together, the results of Tables 10 and 11 imply that the intensity of local clans could significantly increase the operating income and reduce overhead expenses for firms that highly rely on contract enforcement. This general effect of clans on firms greatly boosts the growth of contract-intensive industries and shapes the pattern of industrial specialization across regions in China.

5. Conclusion

Nowadays, institutions are seen as channels through which history matters (Acemoglu et al., 2001; Tabellini, 2008; Nunn, 2009). Using data on genealogies across Chinese prefectures, this paper examines the role of clans, 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, perform several robustness checks, and tackle the endogeneity issues by utilizing a particular instrumental variable, the number of southward migrants during 1127–1130. 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, clans should improve the contracting environment for *all* local firms.

This paper builds on the ongoing literature that explores how informal institutions play a role in China's unconventional path of economic growth. We provide evidence that clan culture, as a distinctively historical legacy, 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 thereby provide practical implications for emerging economies. When governments develop a national strategy to stimulate regional economic growth, they should recognize the comparative advantage created by traditional culture and appropriately guide the agglomeration of industrial sectors. Future research could consider a close scrutiny of the effect of historical institutions on industry agglomeration along other dimensions, like Confucian or temple associations. It is also worthy of note 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.

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.

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. (2019). The struggle for existence: migration, competition and human capital accumulation in historic China, Working Paper.

Bai, Ying, Kung, J. K. (2015). Diffusing Knowledge while Spreading God's Message: Protestantism and Economic Prosperity in China, 1840-1920, *Journal of the European Economic Association*, vol.13, 669-698.

Bai, Y., Kung, J. K., Ma, C. (2017). Endogenous Neo-Confucianism: The Path of Development in Historical China, Working Paper.

Chandrasekhar, A. G., Kinnan, C., and Larreguy, H. (2018). Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field, *American Economic Journal: Applied Economics*, vol.10(4), 43-78.

Chen, L. (1948). The reason why Lin'an was the capital of Southern Song Dynasty (Nansong dingdu linan de yuanyin), *Thought and Times*, vol.47, 24-28.

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

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

Cohen, M. (1990). Lineage organization in North China, *Journal of Asian Studies*, vol.49, 509–534.

Dell, M. (2010). The persistent effects of Peru's mining Mita, *Econometrica*, vol. 78(6), 1863-903.

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

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.

Feng, E. (2011), *The institution of Chinese clans and compilation of genealogies* (Zhongguo zongzu zhidu yu pudie bianzuan), Tianjin: Tianjin Ancient Books Publishing House

Feng, E. (2013), *Clans and Ancestral Temples in Ancient China* (Zhongguo gudai de zongzu he citang). Beijing: The Commercial Press

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.

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

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

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

Jaimovich, E. (2019). Roadways, input sourcing, and patterns of specialization, *European Economic Review*, vol.120, no.103319

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

Jia, R., Lan, X., Miqueld, G. (2021). Doing business in China: Parental background and government intervention determine who owns business. *Journal of Development Economics*, vol. 151, 102670.

Jiang, Q. (2007). *The Name List of Scholars in Qing Dynasty* (Qingchao jinshi timinglu), Beijing: Zhonghua Press.

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

Levchenko, A. A. (2007). Institutional quality and international trade, *Review of Economic Studies*, vol.74, 791-819.

Li, H., Zhou, L. (2005). Political turnover and economic performance: The incentive role of personnel control in China, *Journal of Public Economics*, vol.89, 1743-1762.

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

MacLeod, W. B. (2007). Reputations, Relationships, and Contract Enforcement, *Journal of Economic Literature*, vol.45, 595-628.

Maddison, A. (2007). Chinese Economic Performance in the Long Run. OECD Publishing.

Naughton, B. (2007). *The Chinese Economy: Transitions and Growth*, Cambridge, MA.: MIT 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.

Nunn, N. (2009). The Importance of History for Economic Development, *Annual Review of Economics*, vol.1, 65-92

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

Pasternak, B. (1969). The Role of the Frontier in Chinese Lineage Development, *Journal of Asian Studies*, vol. 28, 551–561.

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.

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

Shiue, C.H. (2017). Human capital and fertility in Chinese clans before modern growth, *Journal of Economic Growth*, vol.22, 351–396.

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.

Szonyi, M. (2002). *Practicing Kinship. Lineage and Descent in Late Imperial China*, Stanford: Stanford University Press.

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

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

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

Wu, S. (2000). *History of Population in China (Zhongguo renkou shi)*, vol.3, Shanghai: Fudan University Press

Wu, X. (2009). *Geographic Distribution of Jinshi in Ming Dynasty*, Hong Kong: The Chinese University of Hong Kong Press

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.

Yi, J. (2019). Endogenous Altruism: Theory and Evidence from Chinese Twins. *Journal of Labor Economics*, vol. 37, 247–295.

Zhang, C. (2019). Family support or social support? The role of clan culture, *Journal of Population Economics*, vol. 32, 529–549.

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

Zheng, D., Kuroda, T. (2013) The impact of economic policy on industrial specialization and regional concentration of China's high-tech industries, *Annals of Regional Science*, vol.50, 771-790.

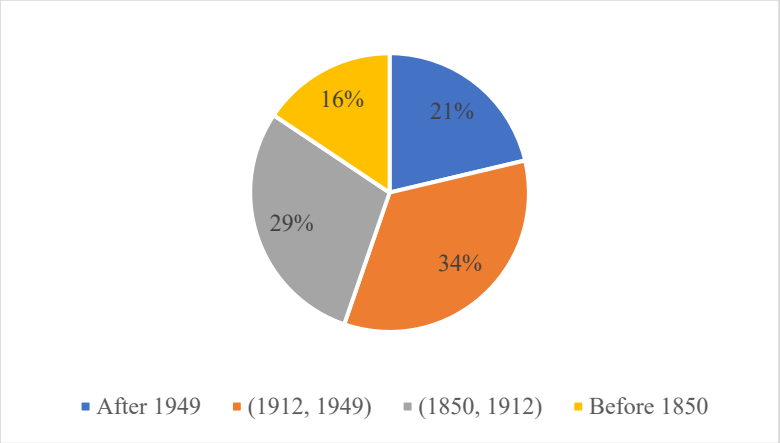


Fig. 1 Distribution of genealogies by the year of compilation

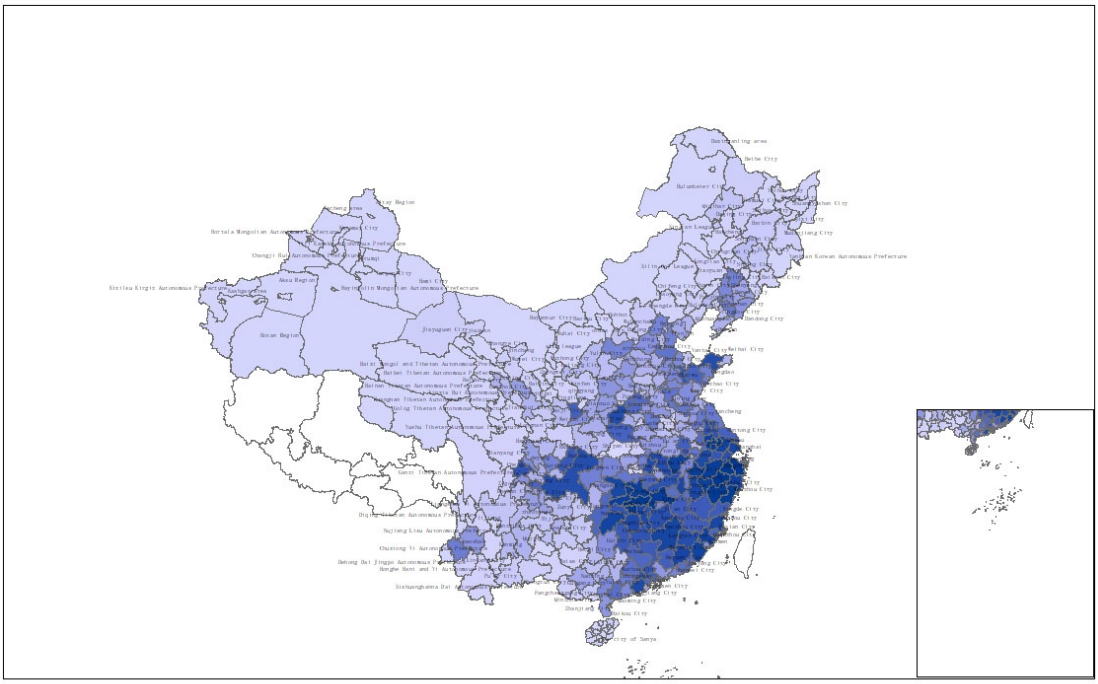
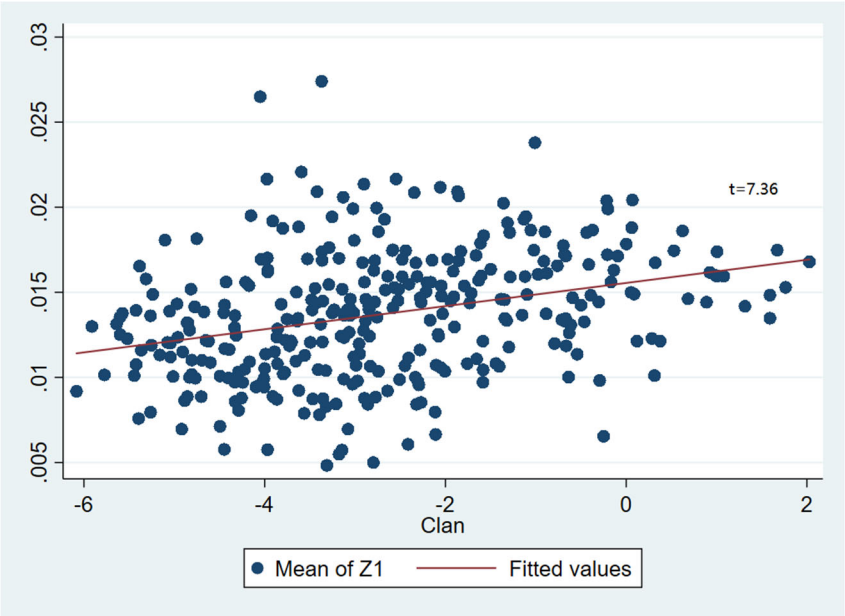
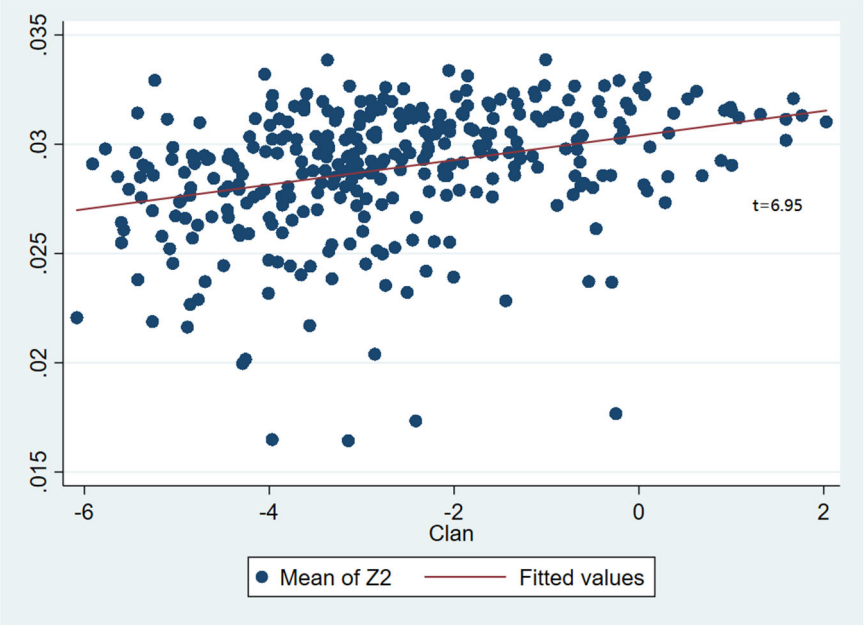


Fig. 2 The geography distribution of clans

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



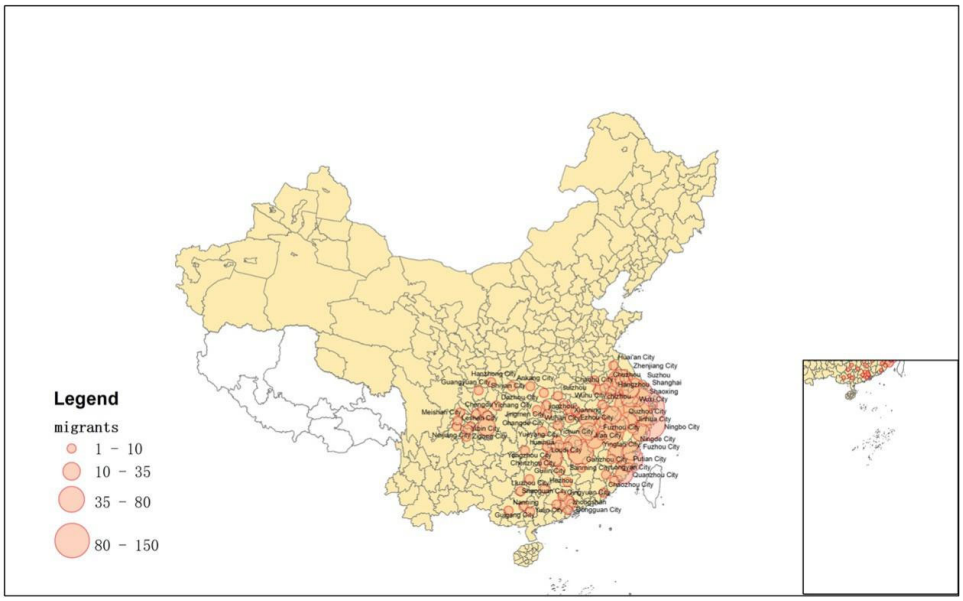
(A)



(B)

Fig. 3 Strength of the clan and the 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).



(A)

(B)

Fig. 4. 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.

Table 1 A simple comparison of industrial output

	(1)	(2)	(3)
Panel A: *z1*			
	Strong clan culture	Weak clan culture	Diff.
High intensity	0.0218	0.0130	0.0087***
Low intensity	0.0278	0.0327	-0.0049***
Panel B: *z2*			
	Strong clan culture	Weak clan culture	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 2 Clans and industrial specialization: OLS regressions

	z1			*z2*		
	Full sample		Cities with Clan>0	Full sample		Cities with Clan>0
	(1)	(2)	(3)	(4)	(5)	(6)
Z × *Clan*	1.7053***	1.2903***	1.0947***	2.0069***	1.3974***	1.2093***
	(0.1210)	(0.1304)	(0.1473)	(0.1629)	(0.1694)	(0.1913)
Z × *ln pop*		1.6382***	1.4133***		1.4400***	1.1424**
		(0.2595)	(0.3579)		(0.3637)	(0.4991)
Z × *ln GDPpc*		1.3656***	1.9265***		3.5389***	3.9944***
		(0.3299)	(0.3569)		(0.4526)	(0.4963)
agr × *ln AGRpc*		2.0197***	0.6360		2.2799***	1.0939*
		(0.5634)	(0.5845)		(0.5544)	(0.5763)
mine × *ln Minepc*		0.7118***	0.7194***		0.7684***	0.7540***
		(0.1286)	(0.1527)		(0.1318)	(0.1521)
Skill × *Edu*		392.6638***	422.7767***		428.0194***	462.7735***
		(66.6044)	(69.5260)		(70.8531)	(73.8614)
Prefecture FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9240	9240	7616	9240	9240	7616
adj. R-sq	0.607	0.614	0.579	0.605	0.615	0.582

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 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. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 3 Clans and contracting institutions

	legal cases	Industrial output	
	(1)	(2)	(3)
Clan	-0.0628***		
	(0.0166)		
Z1 × predicted_legal case		-13.1265***	
		(1.7986)	
Z2 × predicted_legal case			-19.4398***
			(2.4366)
Controls	Yes	Yes	Yes
Prefecture FE	No	Yes	Yes
Industry FE	No	Yes	Yes
N	238	6664	6664
adj. R-sq	0.581	0.598	0.601

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. In columns 2-4, the control variables are the same as those in Table 2. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 4 China's contract intensity

	(1)	(2)	(3)
China's Z × Clan	1.0956***	0.8885***	1.0157***
	(0.1375)	(0.1438)	(0.1625)
China's Z × ln pop		0.7934***	0.4504
		(0.3023)	(0.4051)
China's Z × ln GDPpc		0.6470*	0.8790**
		(0.3731)	(0.3992)
agr × ln AGRpc		1.9133***	0.2564
		(0.6165)	(0.6453)
mine × ln Minepc		0.7273***	0.7286***
		(0.1292)	(0.1501)
Skill × Edu		458.8136***	498.7735***
		(66.0419)	(68.4269)
Prefecture FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
N	9240	9240	7616
adj. R-sq	0.600	0.606	0.573

Notes: This table reports the estimation results by 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*. The control variables are the same as those in Table 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 5 Genealogies compiled before 1949, 1912, and 1850, and after 1978

	Full sample				Cities with Clan			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Z × *clan_before1949*	4.0634***				3.7535***			
	(0.4344)				(0.5527)			
Z × *clan_before1912*		3.5698***				2.8791***		
		(0.4085)				(0.4659)		
Z × *clan_before1850*			2.6239***				1.7704***	
			(0.4081)				(0.4343)	
Z × *clan_after 1978*				4.063***				1.150***
				(0.434)				(0.170)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9240	9240	9240	9,240	7616	7616	7616	7616
adj. R-sq	0.614	0.613	0.611	0.629	0.579	0.578	0.576	0.595

Notes: This table reports the estimation results by counting the number of genealogies before 1949, 1912, and 1850, and after 1978. Columns 1-4 presents the results of full sample and columns 5-8 report the estimations of subsample in which the prefecture has stored at least one genealogy in the given period. The control variables are the same as those in Table 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 6 Controlling for the success of Keju

	z1			*z2*		
	Full sample		Cities with *jinshi*>0	Full sample		Cities with *jinshi*>0
	(1)	(2)	(3)	(4)	(5)	(6)
Z × *jinshi*	0.6047***	-0.2562	-0.1577	0.8577***	0.0326	0.2628
	(0.1382)	(0.1710)	(0.2116)	(0.1816)	(0.2320)	(0.2889)
Z × *Clan*		1.4469***	1.3489***		1.3775***	1.2745***
		(0.1616)	(0.1680)		(0.2197)	(0.2292)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9240	9240	7476	9240	9240	7476
adj. R-sq	0.610	0.614	0.589	0.612	0.614	0.589

Notes: Jinshi refers to total number of *jinshi* during Ming and Qing dynasties normalized by 2007 population. The data are collected from *Distribution of Jinshi in Ming Dynasty* and *List of Jinshi in Qing Dynasty. 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 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 7 Role of clans before the market reform

	2007 output		1985 output	
	Z1	*Z2*	*Z1*	*Z2*
	(1)	(2)	(3)	(4)
Z × *Clan*	0.6918***	1.0075***	0.2343	0.1612
	(0.1963)	(0.3284)	(0.1433)	(0.2159)
Z × *ln pop*	1.6541***	0.0153	-0.6260**	-0.7747*
	(0.5661)	(0.7649)	(0.3018)	(0.4327)
Z × *ln GDPpc*	0.4930	1.4751	-1.3459***	-0.6799
	(0.6618)	(0.9481)	(0.4910)	(0.7188)
agr × *ln AGRpc*	2.2316***	2.1274***		
	(0.7560)	(0.7592)		
mine × *ln Minepc*	0.5257***	0.5938***		
	(0.1658)	(0.1735)		
Skill × *Edu*	298.1461***	307.2699***		
	(82.8815)	(82.9812)		
Prefecture FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
N	2268	2268	2268	2268
adj. R-sq	0.544	0.542	0.605	0.604

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 1-3 are the output of an industry in a prefecture at year 2007 and in columns 4-6 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 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 8 The relation between clan 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	0.8068***	0.7286***	0.4897***			
	(0.0585)	(0.0555)	(0.1020)			
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
adj. R-sq	0.497	0.467	0.393	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. We control for population, GDP per capita, land size, human capital, agricultural output per capita, and mineral output per capita. In columns 4-6, *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 clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 9 The relation between clan culture and industrial specialization: 2SLS regressions

	Full sample		Clan>0		Migration>0	
	z1	*z2*	*z1*	*z2*	*z1*	*z2*
	(1)	(2)	(3)	(4)	(5)	(6)
Z × *Clan*	1.3524***	1.3739***	1.2329***	1.2574***	1.0695**	1.6773**
	(0.1971)	(0.2763)	(0.2213)	(0.3104)	(0.5336)	(0.7309)
Z × *ln pop*	1.6051***	1.4528***	1.3933***	1.1348**	-0.4118	-1.4231*
	(0.2711)	(0.3824)	(0.3585)	(0.4990)	(0.5877)	(0.8196)
Z × *ln GDPpc*	1.3153***	3.5576***	1.8060***	3.9528***	-0.3079	0.8753
	(0.3516)	(0.4788)	(0.3843)	(0.5287)	(0.7861)	(1.0414)
agr × *ln AGRpc*	2.0138***	2.2844***	0.6301	1.0897*	2.2341**	2.2896***
	(0.5638)	(0.5565)	(0.5848)	(0.5772)	(0.8754)	(0.8611)
mine × *ln Minepc*	0.7080***	0.7687***	0.7134***	0.7535***	0.2050	0.1880
	(0.1291)	(0.1319)	(0.1525)	(0.1521)	(0.2011)	(0.1969)
Skill × *Edu*	393.1584***	428.0358***	423.7395***	462.7237***	504.4239***	506.4862***
	(66.6232)	(70.8752)	(69.5261)	(73.8249)	(112.4805)	(110.0361)
Kleibergen-Paap rk LM statistic	748.952	395.394	692.381	364.933	241.3	128.885
Kleibergen-Paap rk Wald F statistic	2320.192	1226.336	2205.164	1161.345	458.921	256.105
Prefecture FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9240	9240	7616	7616	2660	2660

Notes: This table presents two-stage-least-squared estimations for the relation between the strength of local clans and industrial specialization. The instrument for *Clan* is the number of migrants settled in the prefecture during 1127-1130 southward migration. *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; columns 5 and 6 use the sample in which the prefectures have received at least one migrant in the migration wave. For brevity, we do not report first stage estimation results here. The control variables are the same as those in Table 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 10 Firm level analysis: clan and operating profits

Dependent variable: Operating profit ratio							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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)
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
Surname FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	296631	296631	296631	296631	296631	296631	296631
adj. R-sq	0.370	0.370	0.370	0.370	0.370	0.370	0.370

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 scaled by operating income. *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 measured by the logarithm of total assets, the number of employees. All regressions include prefecture, industry, ownership, and surname fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 11 Firm level analysis: clan and overhead expenses

Dependent variable: Overhead expenses ratio

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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)
Contols	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
Surname FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	296533	296533	296533	296533	296533	296533	296533
adj. R-sq	0.212	0.212	0.212	0.212	0.212	0.212	0.212

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 overhead expenses scaled by operating income. *Clan Surnam*e 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 from 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 measured by the logarithm of total assets, the number of employees. All regressions include prefecture, industry, ownership, and surname fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. **p < 0.1; **p < 0.05; ***p < 0.01*.

Online Appendix is 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, $z_i^1 = \sum_k \delta_{ik} R_k$, where δ_{ik} is the share of the value of input k used in the production of industry i and R_k is the share of input k that is neither sold on the exchange nor reference-priced.	Nunn (2007)
Z2	Contract intensity, $z_i^2 = \sum_k \delta_{ik} R_k$, where δ_{ik} is the share of the value of input k used in the production of industry i and R_k is the share of input k that is not sold on the exchange.	Nunn (2007)
China's Z	Contract intensity of China	He et al. (2019)
ln pop	Ln (population 2007)	*China Statistics Yearbook for Regional Economy*
ln GDPpc	Ln (GDP per capita)	*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 the 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 1912)/ 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*
Legal case	Ln [1(+ total number of resolved civil and commercial cases over the period of 2003-2007) / 2007 population]	The yearbooks annually compiled by each prefecture
Migration	Ln (1+#migrants settled in a prefecture during 1127-1130 southward migration wave)	*History of Migration in China*
Migration_surname	Ln (1+# migrants of each surname settled in a prefecture during 1127-1130 southward migration wave)	*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 the entrepreneur)	*The General Catalog of Chinese Genealogies*
ln landsize	Ln (the area of the prefecture)	*China Statistical Yearbook for Regional Economy*
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 An alternative measure of clan strength from CFPS

	Z1	Z2
	(1)	(2)
Z × *Clan_Survey*	9.954***	5.704***
	(1.436)	(1.863)
Z × *ln pop*	1.752***	1.531***
	(0.315)	(0.454)
Z × *ln GDPpc*	2.803***	5.108***
	(0.331)	(0.469)
agr × *ln AGRpc*	0.991*	1.509***
	(0.582)	(0.578)
mine × *ln Minepc*	0.761***	0.796***
	(0.156)	(0.152)
Skill × *Edu*	402.9***	452.5***
	(68.75)	(74.90)
Prefecture FE	Yes	Yes
Industry FE	Yes	Yes
Observations	8,092	8,092
R-squared	0.609	0.609

Notes: This table reports the estimation results by using an alternative measure of clan strength. We focus on a survey question from CFPS: does your family have any genealogy? *Clan_survey* is a ratio variable, defined as the number of people answering yes scaled by the total number of people in this prefecture. The control variables are the same as those in Table 2. All regressions include prefecture and industry fixed effects. Robust standard errors clustered at the prefecture level are reported in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

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

$$U_j = \left[\int_{\omega\in\Omega_j} x_{ij}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}} \tag{B1}$$

where ω indexes varieties in the product set Ω_j, $x_{ij}(\omega)$ is the quantity of variety ω from region i demanded by the representative consumer in region j and $\sigma > 1$ captures the elasticity of substitution between varieties. Then, consumer optimization yields the following demand for variety ω:

$$x_{ij}(\omega) = \frac{\left(p_{ij}(\omega)\right)^{-\sigma}}{P_j^{1-\sigma}} Y_j \tag{B2}$$

where $p_{ij}(\omega)$ is the price of variety ω, $P_j = \left[\int_{\omega\in\Omega_j} p_{ij}(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{1-\sigma}}$ is an aggregate price index, and Y_j represents the total expenditure of region j. 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 (ϕ) according to the following production function:

$$Y = \phi\left(\frac{L}{\mu_L}\right)^{\mu_L}\left(\frac{K}{\mu_K}\right)^{\mu_K}\left(\frac{M}{\mu_M}\right)^{\mu_M} \tag{B3}$$

where μ_L, μ_K and μ_M are positive and $\mu_L + \mu_K + \mu_M = 1$. K and L respectively denote capital and labor inputs employed, and M denotes the intermediate inputs bundle. The intermediates bundle M is assembled by combining one bundle of continuum relationship-specific intermediate inputs, D, and another bundle of intermediate inputs without specific relationship, Z, according to the following CES aggregator:

$$M = \left(D^{\frac{\varsigma-1}{\varsigma}} + Z^{\frac{\varsigma-1}{\varsigma}} \right)^{\frac{\varsigma}{\varsigma-1}}, \tag{B4}$$

where $\varsigma > 1$ is the elasticity of substitution. These component input bundles themselves are also CES aggregates:

$$\begin{aligned} D &= \left(\int_{\Omega_d} d\,(l)^{\frac{\theta-1}{\theta}} dl \right)^{\frac{\theta}{\theta-1}} \\ Z &= \left(\int_{\Omega_z} z\,(h)^{\frac{\theta-1}{\theta}} dh \right)^{\frac{\theta}{\theta-1}} \end{aligned} \tag{B5}$$

where $\theta > 1$ is the elasticity of substitution. Ω_d deontes the set of inputs which is relationship-specific; Ω_z deontes the set of inputs which is not relationship-specific. One unit of intermediat input d or z requires x units of L and y units of K. We use r to denote the rental rate for capital and w to represent the wage payment for unit labor. The rental rate and wage payment are exogenously determined.

For the relationship-specific inputs d, after the production unit is formed, K can only recover a fraction ρ of the investment. The parameter ρ captures quality of contract enforcement. Better institutions thus correspond to higher values of ρ. 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 K 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 $p_d - wx - r\rho y$. We adopt the assumption that ex post the parties reach a Nash bargaining solution and each receive one half of the surplus. Thus, K will

only enter the d-good production if the individual rationality constraint, $\frac{1}{2}(p_d - wx - r\rho y) \geq r(1-\rho)y$, holds. This implies that the price of d satisfies $p_d = wx + r(2-\rho)y$.

The clan culture improves the producers' contracting environment in several ways that are explained in Section 2: a) the effective transmission of information along the clan network will increase the costs of opportunistic behaviors; b) strong collectivism among clan members and the disciplines executed by clans impede both outsiders and clan members to conduct opportunistic behaviors, that is said, facilitating the contract enforcement; c) the cultural norms may be internalized in the spiritual world of their members as well as local people in the long run, so that they instinctively discriminate against opportunistic behaviors. Therefore, ρ is an increasing function of the intensity of clan culture (κ) (i.e., $\rho'(\kappa) > 0$).[20] That is, the material cost to produce the relationship-specific inputs, p_d, is lower for the firms in the regions with the clan culture.

To produce the inputs z, which are not relationship-specific, K can recover its whole investment. Hence, the material cost to produce inputs without specific relationship satisfies $p_z = wx + ry$.

The firm chooses labor L, capital K, and the amounts of intermediate inputs $d(l)$ and $z(h)$, given the wage rate w, the rental rate r, and the prices of intermediate inputs $p(l)$ and $z(h)$. Given the above production function and the firm's productivity ϕ, the marginal cost of inputs when producing the final variety satisfies:

$$c(\phi) = \frac{1}{\phi} r^{\mu_K} w^{\mu_L} P_M^{\mu_M} \tag{B6}$$

where $P_M \equiv \left(P_D^{1-\varsigma} + P_Z^{1-\varsigma}\right)^{\frac{1}{1-\varsigma}}$ is the price index for the intermediate inputs bundle M. $P_D = \left(\int_{\Omega_d} p_d(l)^{1-\theta} dl\right)^{\frac{1}{1-\theta}}$ is the price index for the bundle of the relationship-specific intermediate inputs. $P_Z = \left(\int_{\Omega_z} p_z(l)^{1-\theta} dl\right)^{\frac{1}{1-\theta}}$ is the price index for the bundle of intermediate inputs, which are not relationship-specific.

[20] 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.

B3. Comparative Statics

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

$$\pi(\phi) = \max_{p(\phi)} \big(p(\phi) - c(\phi)\big) p(\phi)^{-\sigma} \sum_j P_j^{\sigma-1} Y_j \tag{B7}$$

which implies that the selling revenue ($r(\phi)$) are equal to:

$$r(\phi) = p(\phi)q(\phi) = \left(\frac{\sigma}{\sigma-1} c(\phi)\right)^{1-\sigma} \sum_j P_j^{\sigma-1} Y_j \tag{B8}$$

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

$$d\ln r(\phi) = -(\sigma-1)\mu_M d\ln P_M = (\sigma-1)\mu_M \frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma} + P_Z^{1-\varsigma}} \frac{ry\rho'(\kappa)}{wx + r(2-\rho)y} d\kappa \tag{B9}$$

where $\frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}}$ corresponds to the degree of dependence on contract enforcement, with a larger value indicating a higher degree of contract intensity.[21] The previous equation implies that:

$$\frac{d\ln r(\phi)}{d\kappa d \frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma} + P_Z^{1-\varsigma}}} = (\sigma-1)\mu_M \frac{ry\rho'(\kappa)}{wx + r(2-\rho)y} > 0 \tag{B10}$$

Hence, we have the following proposition:

[21] This is consistent with Nunn's measure: $\frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}} = \frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}} \times 1 + \frac{P_Z^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}} \times 0$, where $\frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}}$ denotes the expenditure share on the intermediates inputs whose production needs relationship-specific investments and $\frac{P_Z^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}}$ reflects the expenditure share on the non-relationship-specific intermediates inputs whose production does not need relationship-specific investments.

Proposition 1. *The firms in the sectors with higher contract intensity (i.e., higher* $\frac{P_D^{1-\varsigma}}{P_D^{1-\varsigma}+P_Z^{1-\varsigma}}$*) sell relatively more in regions with the clan culture.*