# Informal Institutions, Corporate Innovation, and Policy Innovation *

Jian Xie† Bernard Yeung‡ Qinhong Yu§

August 9, 2025

**Abstract**

Informal institutions can play a crucial role in fostering corporate and policy innovation, especially when formal institutions are weak. However, their intangible nature makes them difficult to quantify. In this paper, we proxy the strength of kinship-based informal institutions using surname homogeneity among business owners, specifically, the extent to which they share a limited number of surnames within the same county. Our analysis reveals that a one-standard-deviation increase in the strength of informal institutions is associated with a 21.1% increase in patent filings and an 18.9% increase in policy innovation. We find that kinship-related informal institutions foster corporate innovation by compensating for weak formal institutions, enhancing protection for intellectual property rights, facilitating access to finance, improving public service delivery, and promoting supply chain cooperation. We also suggest that kinship-related informal institutions encourage local governments to engage in policy experimentation, which relies on the collaboration of business owners. This experimentation process is easier to coordinate and monitor in counties dominated by a few kinship networks. Both informal institutions and policy innovation contribute to economic development and foster entrepreneurial market entries. However, the positive impact of informal institutions declines over time as formal institutions strengthen in China.

*Keywords: Informal Institutions, Surname Homogeneity, Corporate Innovation, Policy Innovation, Entrepreneurs, Economic Development, China*

*JEL codes: O3, O43, O17, E26, L26, Z13*

*We thank Ying Bai, Bishnupriya Gupta, Renliang Liu, Dennis Novy, Yancheng Qiu, Ao Wang, Mingang Zhang, Maggie Zhou, and participants at the 2024 SUFE-SDU-Warwick Economics Workshop, the 2024 IESR-Warwick Workshop, the seminar at the Li Anmin Institute of Economic Research, the Greater Bay Area Workshop on Applied Microeconomics and Public Policy, the CUHK Brownbag Seminar, the SUSTech Brownbag Seminar, and the 3rd Conference on Global Innovation and Intellectual Property, for their helpful comments. All errors are our own.

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

‡College of Business, Southern University of Science and Technology. Email: byeung@nus.edu.sg.

§College of Business, Southern University of Science and Technology. Email: 12431486@mail.sustech.edu.cn.

# 1 Introduction

Institutional factors are critical drivers of innovation and economic growth (North, 1973, 1981). However, in environments where formal institutions are lacking, sustaining innovation and economic growth needs informal institutions. As defined by North (1990), informal institutions are a set of rules, procedures, and norms that constrain individual behavior to maximize collective benefit. Some pioneering works have emphasized the role of informal institutions (Guiso et al., 2006; Greif and Tabellini, 2010) in conferring legitimacy, providing guidance, and incentivizing firms and individuals when formal institutions are absent or underdeveloped, thereby shaping economic outcomes. This paper investigates whether kinship-based informal institutions promote corporate innovation and foster policy experimentation, and whether they also contribute to broader economic development and entrepreneurial market entry.

Measuring the strength of informal institutions is inherently challenging due to their intangible nature and reliance on unwritten norms. In this study, we use surname homogeneity among business owners as a proxy for the strength of kinship-based informal institutions. A high degree of surname homogeneity among business owners within a location suggests the presence of strong kinship-based networks and shared group identities. The formation and significance of common-surname networks are well documented in the literature (e.g., Freedman, 1958; Peng, 2004; Angelucci et al., 2010; Tan et al., 2021; Huang et al., 2023). These kinship-based networks can foster trust, social cohesion, and mutual understanding while also facilitating more effective communication (McPherson et al., 2001). Moreover, norms rooted in ancestral ties can serve as informal enforcement mechanisms, with violations potentially resulting in social sanctions such as reputational damage or community exclusion. Accordingly, high surname homogeneity among business owners may serve as an indicator of stronger kinship-based informal institutions, all else being equal.

We propose and empirically examine the hypothesis that kinship-based informal institutions can foster corporate innovation and motivate local governments to pursue policy experimentation. Both forms of innovation, at the firm and government levels, are vital to economic development. The importance of corporate innovation is rooted in a rich theoretical tradition, including the groundbreaking works of Schumpeter (1934), Nelson and Winter (1985), and Aghion and Howitt (1992). Empirical evidence further reinforces this view, as documented in studies such as Corrado et al. (2009), Bloom et al. (2013), Hall et al. (2010), and Bloom and Van Reenen (2007), among many others. On the policy front, recent research, such as Wang and Yang (2021) and Luo et al. (2024), highlights the significance of local policy experimentation in China's development, with potential relevance for many other countries. We argue that both corporate and government innovations depend on a supportive environment, and that kinship-based informal institutions, among other factors, help create such an environment by facilitating experimentation and innovation.

Kinship-based informal institutions may foster corporate innovation through several channels.

First, they reinforce formal intellectual property protections by leveraging social norms and sanctions to deter infringement. Second, they ease financial constraints by providing network-based access to funding, bypassing the collateral and credit requirements of formal lenders. Third, in areas with weak public services, kinship networks offer quasi-public goods, such as infrastructure, security, and social support, thereby reducing operational uncertainty. Finally, they lower coordination costs (Davis, 2006), enhance resource mobilization, and enable large-scale collective action (Xu and Yao, 2015), effectively substituting for underdeveloped market mechanisms and improving supply chain efficiency. However, the link between homogeneity-based informal institutions and innovation is not necessarily positive. A substantial literature emphasizes the innovation-enhancing role of diversity (Kerr and Lincoln, 2010; Alesina et al., 2016; Fiszbein, 2022; Posch et al., 2023), which expands the production frontier by introducing varied skills, perspectives, and cognitive approaches. Our proxy for informal institutions, surname homogeneity among business owners, may reflect lower diversity, which could potentially dampen innovation. Moreover, stronger informal institutions can foster conservatism, social insularity, or entrenched power dynamics, all of which may inhibit innovation and broader development. Therefore, whether informal institutions foster corporate innovation remains an empirical question.

Kinship-based informal institutions may also encourage local governments to experiment with policy innovations aimed at promoting development, addressing social issues, or solving administrative challenges. In hierarchical political systems, policy experimentation is a defining feature, with successful innovations enhancing officials' career prospects and facilitating the diffusion of new policies. This dynamic is vividly documented in the case of China by Wang and Yang (2021) and Luo et al. (2024). Effective policy innovation typically requires support from the private sector, which is more attainable in counties where kinship ties foster trust among a few dominant families. In contrast to fragmented business environments, these cohesive networks facilitate consensus and implementation more easily (Easterly and Levine, 1997), thereby enhancing the appeal of experimentation to local leaders. We therefore hypothesize that kinship-based business networks are positively associated with the frequency of local policy innovations, offering an additional channel through which informal institutions can promote corporate innovation. Since both forms of innovation are crucial for economic development, we further examine whether kinship-based informal institutions can foster economic development and entrepreneurial market entry.

We examine these questions in the context of China for three key reasons. First, as the world's second-largest economy, China has undergone significant institutional transformation in recent decades, with marked regional and temporal variation in informal institutions. For example, China's ranking in institutional development on the WIPO Global Innovation Index improved from 79th in 2016 to 44th in 2024. Second, China's low level of foreign immigration, compared to large countries such as the United States, Germany, France, and U.K., results in a more ethnically homogeneous population. This makes regional surname homogeneity a cleaner proxy for informal institutions,

with less risk of confounding from ethnic diversity. Third, China's growth has been driven by bottom-up policy reforms that enhance productivity, particularly in areas where central government oversight is weak (Chen et al., 2024), thereby creating space for local policy experimentation. We can explore how local business environments, shaped by kinship-based informal institutions, may also facilitate such experimentation.

Using registration data for the universe of Chinese private enterprises, we measure the strength of informal institutions by calculating annual surname homogeneity among business owners at the county level, based on an entropy approach. Employing a panel regression with county and year fixed effects for the period 2001–2020, we find that stronger informal institutions are positively associated with patent filings by private firms, which we use as a proxy for corporate innovation. Specifically, a one-standard-deviation increase in the strength of informal institutions is associated with a 21.1% increase in innovation outcomes relative to the sample mean, an economically significant magnitude. We also find a positive association between the strength of informal institutions and county-level policy innovation, as measured by the number of policy experiments conducted at the county level. A one-standard-deviation increase in the strength of informal institutions is associated with an 18.9% increase in policy experiments, relative to the sample mean.

A key concern is the potential for omitted variable bias, given the complex factors that shape informal institutions. We address this in several ways. First, we exploit within-county variation in informal institutions and innovation by controlling for time-invariant county-level unobservables using county fixed effects. Second, our baseline and robustness checks include a broad set of controls to account for time-varying confounders. Third, we employ an instrumental variable based on rice suitability to predict the strength of informal institutions. The rationale is that rice cultivation requires extensive sharing of water resources and coordinated collective efforts to construct and maintain irrigation systems, endeavors historically facilitated by kinship-based institutions (Freedman, 1966; Talhelm et al., 2014; Jia and Kung, 2025). The instrumental variable estimates further reinforce our main findings.

We conduct extensive robustness checks to validate our regression results. First, to account for potential time lags between changes in informal institutions and innovation, we use lags of one to five years as independent variables. Second, we sequentially exclude three types of special samples: counties where ethnic minority populations exceed 10%, centrally administered municipalities, and special economic zones. Third, we test an alternative measure of surname homogeneity using the Herfindahl-Hirschman Index. Finally, we add controls for potential confounders, including economic development, industry structure and homogeneity, clan culture, agricultural diversity, and the presence of treaty ports. Our results, using either corporate innovation or policy innovations as dependent variables, remain robust across all specifications.

We further investigate the mechanisms through which informal institutions foster corporate innovation, focusing on whether they can substitute for formal institutions across four key dimensions:

intellectual property rights, financial development, public administration, and market development. Our results indicate that the positive impact of informal institutions diminishes as these formal institutions become more established. This implies that in weaker institutional environments, companies rely more on kinship networks to support innovation (Tsai, 2004; Mair et al., 2012; Doh et al., 2017). We also find that informal institutions are more likely to foster policy innovations when formal institutions are weak. These effects were strongest in the early 2000s, when China's formal institutions were less developed, and declined over time, underscoring the compensatory role of informal institutions.

Finally, we examine the relationship between informal institutions, policy innovation, economic development, and entrepreneurial market entry. Using nighttime light intensity as a proxy for economic development and the number of new firms per 1,000 people as a measure of entrepreneurial market entry, we find that both outcomes are positively and significantly associated with informal institutions and policy innovation. This suggests that informal institutions foster development and entrepreneurship both directly and indirectly by enabling policy innovation.

This paper makes four key contributions to the literature. First, we advance understanding of how informal institutions shape innovation as well as economic outcomes (e.g., North, 1990; Peng, 2004; Guiso et al., 2006; Davis, 2006; Peng et al., 2009; Greif and Tabellini, 2010; Angelucci et al., 2010; Helmke and Levitsky, 2012; Xu and Yao, 2015; Buonanno and Vanin, 2017; Minbaeva et al., 2023). Prior studies often rely on small-scale cross-sectional data or lack robust empirical analysis. We address this gap by constructing a large-scale panel dataset that captures informal institutions through surname homogeneity among private-sector business owners, drawing on the full universe of China's private business sector. This measure also connects to the literature on surname-based social networks (McPherson et al., 2001; Angelucci et al., 2010; Güell et al., 2015; Tan et al., 2021; Bai and Kung, 2022; Huang et al., 2023). We demonstrate that informal institutions can substitute for weak formal institutions in enhancing corporate innovation through kinship-based networks, which protect intellectual property rights, facilitate access to finance, provide public services, and secure market access. Moreover, we document a diminishing impact of informal institutions on innovation as formal institutions mature. We also highlight the potential role of informal institutions in incentivizing policy experimentation, which is crucial for economic outcomes.

Second, we contribute to the understanding of the factors driving the enhancement of China's corporate innovation. Previous research has attributed this improvement to various elements, such as intellectual property rights protection (Ang et al., 2014; Fang et al., 2017), international trade (Hu and Jefferson, 2009; Liu and Qiu, 2016; Chen et al., 2017; Liu and Ma, 2020; Liu et al., 2021), subsidies and policies (Ding and Li, 2015; Fang et al., 2018; Shao and Xiao, 2019; Wei et al., 2021; Zheng and Li, 2020), productivity growth (Fang et al., 2020; Köbig et al., 2022), and infrastructure development (Wang et al., 2018; Gao and Zheng, 2020; Hanley et al., 2022).[1] Our study highlights the

[1]For a more comprehensive analysis, please refer to Wei et al. (2017).

roles of informal institutions in China in fostering innovation and growth, providing insights into how developing countries can leverage various approaches to enhance their innovation capacity.

Third, we contribute to the understanding of China's policy innovation, an important yet still underexplored aspect of Chinese politics. Qian et al. (2006) provide a theoretical framework to explain how region-based organizational forms influence policy experimentation. Heilmann (2008a) and Heilmann (2008b) discuss the institutional framework, distinctive instruments, procedural mechanisms, and outcomes of policy experimentation in China. Luo et al. (2024) demonstrate that officials' incentives, including prospects for promotion and risks of punishment, shape both the nature and diffusion of policy innovations. Wang and Yang (2021) document that policy innovations in China are more likely to be introduced in regions with more favorable conditions. Similarly, Chen et al. (2024) find that policy innovations are more likely to occur in areas with higher GDP growth. We extend this perspective by highlighting the role of anticipated collaboration from the business sector as an additional motivating factor. Informal institutions facilitate such collaboration, thereby fostering policy innovations that are positively associated with both economic development and entrepreneurial market entry.

Finally, we contribute to advancing a balanced approach between homogeneity and diversity in economics. The relationship between diversity and innovation remains a topic of debate, with mixed empirical findings (Alesina et al., 2003; Ozgen, 2021). Some studies highlight diversity's positive impact on innovation (Kerr and Lincoln, 2010; Alesina et al., 2016; Fiszbein, 2022; Posch et al., 2023), while others argue that homogeneity fosters trust, reduces conflict, and supports effective policy-making, thereby promoting innovation and development (Easterly and Levine, 1997; Alesina et al., 2003; Alesina and La Ferrara, 2005; Habyarimana et al., 2007; Ashraf and Galor, 2011). Furthermore, the relationship between homogeneity and innovation may be non-linear. Ashraf and Galor (2013) find that societal genetic diversity affects development outcomes in a hump-shaped manner, and Alesina and La Ferrara (2005) suggest that diversity is beneficial in developed societies with strong institutions but detrimental in weaker institutional contexts. Using regional variation in surname homogeneity in China and its varied stages of development across regions and over time, we find that homogeneity supports innovation during the early stages of development, while diversity becomes advantageous as formal institutions strengthen, consistent with Alesina and La Ferrara (2005).

The rest of the paper is organized as follows. Section 2 introduces the concept of Chinese surnames and explains the relationships between surname homogeneity among business owners and informal institutions, corporate innovation, and policy innovation. Section 3 details our data, variables, and measurement methods. Section 4 presents our empirical approaches and results, including the baseline estimation, the long difference approach and other robustness checks, heterogeneity of institutional development analyses, and analysis across different periods. Section 5 presents the analysis of policy innovation. Section 6 reports additional results on the overall economic impact. Finally, Section 7 concludes the paper.

# 2 Conceptual Framework

## 2.1 Introduction to Chinese Surnames

Most Chinese surnames date back over two thousand years and have diverse origins, including geographical locations, occupations, tribal affiliations, official titles, natural features, conferred names by emperors, and phonetic transliterations of non-Han ethnic names. Ruofu and Yida (1996) compile a total of 11,969 Chinese surnames across all historical periods. Among them, the 504 surnames included in "*The Hundred Surnames*" (*Baijiaxing*, compiled during the Song Dynasty, 960-1279 AD) are significantly more common than others. These 504 surnames consist of 444 single-character surnames and 60 double-character surnames. Over the past two thousand years, Chinese surnames have typically been passed down through the patriarchal line, signifying family lineage and serving as a connection to ancestors and extended family or clan members. This tradition reinforces the importance of filial piety, family ties, and social identity. It is rare for Chinese individuals to change their surnames, except during periods of persecution or other life-threatening events.

A complex interplay of factors, including geographic features, historical migrations, socio-political events, cultural characteristics, and fertility rates, shapes the spatial distribution of Chinese surnames. In some regions of China, specific surnames dominate entire villages or areas, shaped by geographic isolation or strong patrilineal clan traditions that maintain genealogical records and emphasize shared ancestry. In contrast, regions with higher migration and more inclusive cultural norms tend to display greater surname diversity. Figure 1 illustrates the surname distribution of entrepreneurs in two example counties in 2000. As illustrated in Figure 1a, the surnames "Chen" and "Huang" dominate, accounting for a substantial proportion of the population. By comparison, Figure 1b demonstrates a comparatively more diverse distribution of surnames. The difference in surname homogeneity between the two counties may be attributed to Hanjiang's relatively rugged geographic features, stronger clan culture, and lower historical migration compared to Yingjiang. The distribution of surnames varies not only across counties but also over time within the same county.




(a) Hanjiang in Fujian Province (b) Yingjiang in Anhui Province

Figure 1: The Distribution of Entrepreneurs' Surnames in Two Counties, 2000

Individuals within a county or smaller administrative unit who share the same surname often

demonstrate higher levels of mutual support, trust, and understanding. At the first level, people with the same surname within a small region may belong to the same extended families or clans. These surname networks are essentially kinship-based, forming crucial relational bonds that hold significant importance for many Chinese individuals and play a critical role in influencing economic activities in China (Greif and Tabellini, 2010). At the second level, even when individuals sharing the same surname no longer maintain clear kinship ties, they may still trace their lineage to a common ancestor. These ancestral connections and shared genetic heritage can evoke emotional responses, fostering a strong sense of belonging and identity within the group (Tan et al., 2021).

Numerous studies have documented the impact of surname-based networks in promoting mutual support among individuals in China. For example, Xu and Yao (2015) find that village leaders from the two largest surname-based families are more effective at organizing community activities. Similarly, Huang et al. (2023) show that greater surname homogeneity among a firm's board members reduces agency costs and mitigates information asymmetry. Tan et al. (2021) also use the sharing of directors' surnames as a proxy for measuring social ties in organizations. Importantly, this phenomenon is not unique to China. For instance, Cruz et al. (2017) show that people are more likely to support candidates who share the same surname as them during elections in the Philippines. Buonanno and Vanin (2017) find that regions in Italy with more concentrated surnames are associated with lower crime rates, highlighting the role of surname-based networks in fostering social cohesion. Angelucci et al. (2010) demonstrate that surname-based networks influence secondary school enrollment in Mexico.

## 2.2 Surname Homogeneity among Business Owners and Informal Institutions

Surname homogeneity among business owners within a county reflects the extent to which dominant company owners share common surnames. High homogeneity suggests strong kinship-based networks or dominance by a few clans, whereas low homogeneity implies greater diversity, likely driven by migration or more inclusive cultural norms. We use surname homogeneity as a proxy for the strength of kinship-based informal institutions, all else equal.

Measuring informal institutions is challenging due to their intangible nature, cultural nuances, and reliance on unwritten norms, social behaviors, and interpersonal networks. North (1990) defines institutions as rules, procedures, and norms that constrain individual behavior to maximize collective benefit. Informal institutions, in particular, are internally enforced and socially sanctioned through norms, behavioral expectations, and standards of conduct. In this context, high surname homogeneity among dominant company owners in a county reflects strong kinship ties and shared identities that align with these principles. Such networks foster trust and mutual support, with norms rooted in ancestry functioning as informal extensions of enforcement mechanisms. Social sanctions–such as reputational loss or exclusion–reinforce adherence, while historical settlement

patterns and clan dominance help preserve these norms over time. Thus, surname homogeneity serves as a proxy for the strength and persistence of kinship-based informal institutions.

Some prior research defines informal institutions in similar ways. For example, Freedman (1958) argues that lineage groups organized around shared surnames function as informal institutions in regions where formal bureaucratic institutions are weak. Similarly, Greif and Tabellini (2010) regard kinship-based networks as a significant form of informal institution in China. Xu and Yao (2015) use the presence of village leaders from the two dominant surnames within a village as a proxy for measuring the strength of local informal institutions, emphasizing that these institutions, established by large kinship-based networks, consist of rules created and enforced by kinship members. Furthermore, Tsai (2004) highlights that kinship-based networks can operate as informal institutions, facilitating the pooling and redistribution of resources among their members. Section A.2 provides evidence demonstrating the strong correlation between surname homogeneity among company owners and measures of kinship-based networks.

### 2.3 Surname Homogeneity among Business Owners and Corporate Innovation

The relationship between surname homogeneity among business owners and corporate innovation in a region, however, could be complex, encompassing both negative and positive forces. On the one hand, high surname homogeneity might hinder innovation for several reasons. First, homogeneous societies often share traditions, beliefs, and norms. In such settings, individuals tend to amplify or reinforce preexisting beliefs, making them reluctant to generate new ideas or adopt alternative perspectives. Second, high surname homogeneity often reflects social insularity, with limited exposure to outside influences and fewer opportunities for collaboration, learning, and the adoption of external knowledge or complementary skills. Third, high surname concentration may lead to entrenched power structures, where a few dominant family clans hold significant control over decision-making. Such hierarchies can undermine merit-based competition, discouraging the emergence of innovative ideas that may challenge the status quo. Finally, homogeneous communities tend to preserve their traditions and are generally more conservative and risk-averse, making them less conducive to innovation, which thrives on uncertainty, experimentation, and disruption.

On the other hand, surname homogeneity among business owners can reflect the strength of kinship-based informal institutions, as discussed in Section 2.2. In regions with underdeveloped formal institutions, these kinship-based informal institutions can act as substitutes, fostering mutual trust and support among their members (Freedman, 1958; Tsai, 2004). They can promote corporate innovation through several mechanisms.

First, kinship-based informal institutions complement formal systems by safeguarding intellectual property through social enforcement. Ethical norms within these networks deter intellectual property theft and other forms of opportunistic behavior, as violators face social sanctions (Greif

et al., 1994). These institutions also help protect members from external exploitation and unfair competition by other groups.

Second, kinship-based arrangements can substitute for formal institutions in financing innovation. Business owners can raise capital through their kinship networks when formal financial institutions are unwilling or unable to provide financing. These networks are characterized by long-term relational ties and high levels of mutual trust, which significantly reduce information asymmetry and transaction costs in borrowing and lending. Moreover, these networks tend to offer more flexible terms and personalized support, enabling entrepreneurs to invest in R&D without the stringent requirements typically imposed by formal lenders.

Third, kinship-based informal institutions can provide some public services, such as maintaining public safety and infrastructure and offering social security. For example, Buonanno and Vanin (2017) find that surname-based networks help lower crime rates in Italy. This occurs because these informal institutions facilitate resource pooling, mutual aid, and risk-sharing among members, which collectively contribute to creating a safe and supportive environment for the community. When the government is inefficient and unable to provide adequate public services, the role of these informal networks becomes even more important. The higher quality of public services they provide helps create a more conducive environment for innovation.

Finally, kinship-based informal institutions can complement formal market mechanisms, as they are often associated with higher levels of social trust, lower transaction costs, and smoother collaboration. In regions with underdeveloped or inefficient factor and product markets, opportunistic behavior can undermine arm's-length transactional relationships. Individuals often turn to these informal institutions to organize production, facilitate trade, and coordinate economic activities. By doing so, they enhance the efficiency of supply chains and promote the flow of resources, which can lead to innovation. As argued by Lim et al. (2024), Chinese communities in Java promote local economic development through their commercial networks.

These factors suggest a multifaceted and complex relationship between surname homogeneity among business owners and corporate innovation. The positive effects of surname homogeneity often emerge where formal institutions are weak, as kinship-based informal institutions play a critical role in protecting intellectual property rights, facilitating access to finance, delivering public services, and supporting market activities. By contrast, its negative effects tend to arise in contexts where formal institutions are more established. As Alesina and La Ferrara (2005) explain, diversity can be advantageous at higher levels of development, as advanced economies and wealthier societies are better equipped to leverage the productivity benefits of skill complementarities. These societies possess institutional frameworks capable of managing the inherent conflicts associated with diversity, thereby enabling them to maximize their potential. This aligns with the findings of Ashraf and Galor (2011), who argue that cultural homogeneity positively influenced economic development during the agricultural era, a period characterized by weak formal institutions, but has negatively

affected income per capita during the industrialization era.

### 2.4 Surname Homogeneity among Business Owners and Policy Innovation

In China, many policy innovations—new policies designed to promote economic development, address social issues, or tackle administrative challenges—are not implemented nationwide at the outset. Instead, they are often piloted in selected regions known as experimentation sites (shidian or shiyanqu). Wang and Yang (2021) document that China's policy experimentation process exhibits positive selection bias, whereby policy innovations are more likely to be introduced in regions with favorable conditions.

We hypothesize that policy innovations are more likely to be implemented in counties with higher surname homogeneity among business owners for several reasons. First, more homogeneous groups are better able to reach consensus on implementing development-oriented policies. This stands in contrast to more fragmented societies, where collective action problems may hinder the provision of sound public policy, as Easterly and Levine (1997) argued. Second, as discussed previously, higher surname homogeneity among business owners reflects the presence of stronger kinship-based informal institutions. These institutions promote trust, reduce coordination costs, and facilitate cooperation both among business owners and between business owners and government officials. Such cooperation reduces the risk of resistance or unrest when new policies are piloted. Finally, these counties often possess well-organized local kinship-based networks that can mobilize resources rapidly and effectively, thereby increasing the probability of successful implementation. If policy innovations generally have positive effects, this could be another channel through which they promote innovation by local private firms.

## 3 Data and Variables

### 3.1 Surname Homogeneity among Business Owners

#### 3.1.1 Measurement

The data source for business owners' surnames is the registration records of the universe of Chinese enterprises from the State Administration for Industry and Commerce (SAIC). Following Bai et al. (2024), we identify a business owner's surname based on the legal representative's surname. Our analysis focuses on private firms, as legal representatives of state-owned or foreign-controlled firms are unlikely to be founders. From 2000 to 2019, approximately 49 million private firms were registered, accounting for about 86% of all firms. We focus on private business owners rather than the general population for two reasons: (1) surname homogeneity among business owners more accurately reflects the connectedness of key decision-makers, who are the primary drivers of

innovation; and (2) comprehensive surname data for the general population is unavailable. The 2005 mini-census includes surnames for only 2.5 million individuals, far fewer than in the firm dataset. Nevertheless, as shown in Figure A.2, surname homogeneity among private business owners is highly correlated with that of the general population within regions.

We compute surname homogeneity among business owners at the county level, which includes counties, county-level districts, and county-level cities, rather than higher administrative levels. This choice is based on the idea that durable surname-based connections are more likely to form within smaller, more socially cohesive areas, where interactions among business owners are more frequent and intense. These ties tend to weaken at broader geographic levels, such as cities or provinces. Moreover, the county level is the smallest administrative unit for which all key variables required in this study are available.

We measure the surname homogeneity among business owners using a modified entropy-based approach. Originally proposed by Shannon (1948) to quantify diversity in information theory, this approach has since been widely applied across various scientific disciplines. In the field of economics, it has been utilized by Buonanno and Vanin (2017) and Posch et al. (2023) to calculate surname diversity in their analyses. Given that our study focuses on surname homogeneity, the conceptual opposite of diversity, we adapt their formula by reversing its sign. Specifically, the surname homogeneity among business owners at the county level is calculated as follows:

$$SurnameHomogeneity_{c,t} = \sum_{k} p_{c,t,k} log_2(p_{c,t,k}) \tag{3.1}$$

where $p_{c,t,k}$ is the fraction of business owners with surname $k$ in county $c$ in year $t$. We consider business owners of all existing firms in year $t$. The functional form of $\log_2(p_{c,t,k})$ reflects the idea that rare surname groups contribute more "surprise" compared to more common surname groups. In general, fewer surnames and more concentrated fractions ($p_{c,t,k}$) result in higher homogeneity (or lower diversity). In an extreme case where all business owners in a county share the same surname in a given year, this measure equals 0, representing the highest possible value, as the measure is typically negative when $p_{c,t,k} < 1$. In contrast, a larger number of surnames and more evenly distributed fractions ($p_{c,t,k}$) lead to lower homogeneity (or higher diversity).

To assess whether this measure captures the strength of kinship-based informal institutions, we present a validation test in Section A.2 of the Appendix. Another commonly used measure of homogeneity is based on the Herfindahl-Hirschman Index (HHI). However, it has a much more skewed distribution compared to the entropy-based approach. Therefore, we use entropy as our baseline measure and use HHI as a robustness check.

### 3.2 Corporate Innovation

We measure corporate innovation using the number of patent applications filed by private firms in each county and year, normalized by the local population in 2000. Patent applications are widely used as a measure of innovation in the literature (e.g., Liu et al., 2021; Liu and Qiu, 2016; Aghion et al., 2009; Hashmi, 2013; Hu and Jefferson, 2009). Limitations notwithstanding, patent application data generally reflects innovative activities, as it captures the effectiveness of innovation efforts, including both observable and unobservable inputs (Liu and Qiu, 2016). Furthermore, other measures of innovation, such as R&D expenditure, are not available at the county level across China for an extended period. We use patent applications instead of granted patents due to the potential for a significant and uncertain time lag between innovation and patent approval. Our results remain robust when granted patents are used, as discussed in Section 4.2.2.

We focus exclusively on patent applications by private firms rather than all patent applications within a county, as our measure of surname homogeneity is based solely on business owners of private firms, and thus should correspond to patent applications by these firms. Furthermore, we focus on invention patents, excluding utility models and design patents, as the latter two reflect minimal real innovation. We focus on patent applications filed between 2001 and 2020, as China's accession to the WTO in 2001 significantly affects the country's innovation landscape. Data on Chinese patents are sourced from the Global Patent Database (IncoPat),[2] which provides comprehensive information on applicants as well as the dates of application.

### 3.3 Policy Innovation

Our data on policy innovation comes from PKULaw, which holds a comprehensive collection of government documents.[3] Following Wang and Yang (2021), we search for government documents issued between 2000 and 2020 that contain the terms "experimentation points" (shidian) or "experimental zones"(shiyanqu). These documents contain information regarding the experimentation with policy innovations. Since our analysis focuses on county-level outcomes, we restrict our sample to experimentation with policy innovations implemented at the county level, excluding those introduced at higher administrative levels. After removing duplicates and substantively identical documents, we obtain 1,067 government documents related to experimentation with policy innovations at the county level.

For each year, we calculate the number of policy innovations implemented in each county. These policy innovations may vary in their impact–some promoting economic development, while others may have adverse effects. Since our goal is to examine the relationship between surname homogeneity among business owners and policy innovations, and because policy innovations are not ex-

[2] https://www.incopat.com/
[3] https://www.pkulaw.com/

clusively for promoting economic gains, we choose to count all identified policy innovations rather than only those that favor economic development.

## 3.4 Control Variables

We include several sets of control variables in our analysis.

**1. Geographic Controls**

Our geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Average ruggedness and distance to the coast are computed for each county using GIS software. The provincial capital dummy identifies counties located in provincial capitals, while the special economic zone dummy indicates whether a county falls within a designated special economic zone. Provincial capitals typically have greater access to resources and attract more non-local businesses, while special economic zones are designed to promote corporate and policy experimentation. These geographic characteristics may influence corporate innovation, policy experimentation, and surname homogeneity. As they are time-invariant, we account for them by interacting each with year-fixed effects

**2. Transportation Controls and the Number of Private Firms**

Transportation controls consist of a high-speed railway station dummy and an airport dummy. The high-speed railway station dummy identifies counties with a high-speed railway station in a given year, while the airport dummy indicates whether the county is located in a city with an airport. We use city-level data for airports, as airports typically serve a broader region than a county. The influence of infrastructure on innovation is well-documented in the literature. Data on high-speed railway stations are sourced from China Railway, while airport data are hand-collected. We also control for the number of private firms, which is calculated using registration records from the universe of Chinese enterprises. This variable is relevant because it correlates with both surname homogeneity among business owners and local innovation.

**3. Demographic Controls**

Our demographic controls include population, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Population reflects a county's scale, while average years of schooling and urbanization rate capture education and economic development levels, respectively. The fraction of minorities and migration rate may be correlated with surname homogeneity. These variables are sourced from the 2000 census and are interacted with year dummies since they are time-invariant.

**4. Socioeconomic Controls**

Finally, we control for several socioeconomic variables that might influence corporate and policy innovation, including marketization index, bank density, number of universities, and number of elite universities (Project 211 universities). Marketization index, which varies at the city level, is sourced

from the China Market Index Database.[4] It reflects the relationship between the government and the market, as well as the development of the market. Bank density, defined as the number of banks normalized by population, serves as a measure of financial development. The data is sourced from the China Banking and Insurance Regulatory Commission.[5] The number of universities and the number of elite universities are used to measure educational development. The latter is basically time-invariant and is interacted with year dummies.

## 3.5 Variables in the Heterogeneity of Institutional Development Analyses

To deepen our understanding of the relationship between innovation and informal institutions, proxied by surname homogeneity among business owners, we use four sets of institutional development measures as moderating variables.

**1. Intellectual Property Rights Protection**

We measure the level of intellectual property rights protection using an existing variable: the intellectual property rights protection index. The index is a subcomponent of the broader marketization index from the China Market Index Database.[6] It is only available at the provincial level. However, the advantage of using provincial-level intellectual property rights protection is that it is less likely to be influenced by a county's surname homogeneity.

**2. Financial Development**

We use two proxies to measure financial development. The first proxy is bank density, as mentioned in Section 3.4, which varies at the county level. The second proxy is the ratio of total loans and deposits from formal financial institutions to GDP. The data is sourced from China's Provincial Statistical Yearbooks and varies at the provincial level.

**3. Quality of Public Administration**

The quality of public administration is proxied by public service expenditure, normalized by GDP in 2010 at the provincial level. The data on public service expenditure is sourced from the China Statistical Yearbooks and is available only from 2007.

**4. Development of Market Forces**

This measure is based on the indices of the factor and product market development. These indices are only available at the provincial level and are subcomponents of the broader marketization index. In contrast to the broader marketization index, which captures the relationship between the market and the government, these indices specifically focus on the factor and product markets.

[4] https://cmi.ssap.com.cn/
[5] https://www.cbirc.gov.cn/en/view/pages/index/index.html
[6] https://cmi.ssap.com.cn/

### 3.6 Other Variables

In Section 4.2.2, we conduct robustness checks by controlling for additional variables. We measure the strength of clan culture based on genealogies. Genealogy data are sourced from the Shanghai Library Chinese Genealogy Knowledge Service Platform.[7] We measure agricultural diversity based on the caloric content of five crops. Data on the caloric content of each crop are from the Food and Agriculture Organization (FAO) of the United Nations.[8] In Section 6, we measure economic development using average nighttime light intensity. The data originates from DMSP/OLS (Defense Meteorological Satellite Program/Operational Linescan System) and NPP-VIIRS (National Polar-orbiting Partnership/Visible Infrared Imaging Radiometer Suite). Data on entrepreneurial market entries are based on the registration information of the University of Chinese Firms. The summary statistics for all the variables used in this paper are presented in Table A.1.

## 4 Empirical Strategy and Results

### 4.1 Baseline Estimation

To investigate the relationship between surname homogeneity among business owners and corporate innovation at the county level, we employ the following econometric specification:

$$Y_{c,t+1} = \alpha + \beta SurnameHomogeneity_{c,t} + X'_{c,t}\gamma + \mu_c + \delta_t + \varepsilon_{c,t} \tag{4.1}$$

where $Y_{c,t+1}$ is the number of patent filings by private firms in county $c$ in year $t+1$, normalized by 1000 local population;[9] It covers the period from 2001 to 2020. $SurnameHomogeneity_{c,t}$, as in Equation 3.1, is the surname homogeneity among business owners in county $c$ in year $t$, using data from all existing private firms in year $t$; $X'_{c,t}$ indicates control variables. There are four sets of controls, as shown in Section 3.4: geographic controls, transportation controls and the number of private firms, demographic controls, and socioeconomic controls. Time-invariant controls, including geographic variables, demographic controls, and the number of elite universities, are interacted with year dummies to account for differential trends or the time-varying effects of these fixed characteristics. $\mu_c$ represents the county fixed effects; $\delta_t$ represents the year fixed effects; $\varepsilon_{c,t}$ is the error term. We winsorize the outcome variable at the 5th and 95th percentiles to mitigate the influence of outliers.[10] Tibet and Qinghai are excluded from the estimation due to the presence of numerous omitted variables. We use two-way clustering of standard errors at the county and year levels to account for

[7] https://jiapu.library.sh.cn/#/

[8] https://data.apps.fao.org/

[9] Our results remain robust when patent filings are normalized by the number of private firms in the county, as shown in Table A.4, and when using unnormalized counts, as shown in Table A.5.

[10] Our results remain robust when the outcome variable is winsorized at the 1st and 99th percentiles, as shown in Table A.6.

potential correlation of the error terms within the same county across years and within the same year across different counties.

Table 1 presents the baseline estimates. Column (1) begins with a specification that includes only county and year fixed effects. In column (2), we add geographic controls, and in column (3), we incorporate transportation controls and the number of private firms. Column (4) further includes demographic variables, while column (5) introduces socioeconomic variables, comprising the full set of controls. Column (5) represents our preferred specification, corresponding to Equation 4.1. The estimates indicate that surname homogeneity is positively correlated with patent filings, with results statistically significant at the 1% level across all five columns. According to our preferred specification in column (5), a one-unit increase in surname homogeneity results in a 0.028 increase in patent filings per 1,000 people. In other words, a one-standard-deviation increase in surname homogeneity is associated with a 21.1% rise in patent filings,[11] relative to the sample period's mean value. This represents an economically meaningful magnitude.

Table 1: Baseline Results

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.106*** | 0.057*** | 0.041*** | 0.029*** | 0.028*** |
| | (0.012) | (0.006) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.627 | 0.657 | 0.685 | 0.703 | 0.719 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

We address several important concerns:

The first major concern pertains to omitted variables that could be correlated with both surname homogeneity and corporate innovation. We address this issue in three ways. First, we include county fixed effects in our specification, which allows us to subsume time-invariant county-level omitted variables and focus exclusively on within-county changes in innovation output driven by within-county changes in surname homogeneity. This approach aligns with that of Xu and Yao

[11] $(0.028 \times 0.587)/0.078 = 21.1\%$

(2015), leveraging the strengths of a panel data structure. By doing so, we effectively eliminate all time-invariant county-level unobservables. Second, we include four sets of control variables in our specification, which helps mitigate concerns about time-varying county-level confounders. In Section 4.2.2, we further include additional potential controls, and the results remain robust. Third, in Section 4.3, we employ an instrumental variable based on rice suitability to provide additional evidence addressing concerns about omitted variable bias.

The second major concern relates to the time lag between the innovation process and patent application. The innovation process typically precedes the patent application. However, under China's first-to-file patent system, applicants are incentivized to file as quickly as possible to secure their rights to an invention. Therefore, in our baseline analysis, we set the outcome variable at time $t+1$ and the explanatory variable at time $t$. Considering that the time lag could be longer, we conduct robustness checks, as detailed in Section 4.2.2. Our results remain consistent when the outcome variable is delayed by 2, 3, 4, or 5 years.

The third major concern is serial dependence in surname homogeneity, as changes within a county are typically small from year to year. Each year, we calculate surname homogeneity based on the set of existing private firms. While some firms may exit and others enter between year $t$ and $t+1$, substantial overlap likely remains, leading to serial dependence in the measure. We address this concern in three ways. First, we use Newey-West standard errors to account for serial dependence. The results remain robust to this adjustment, as shown in Table A.7. Second, we provide more intuitive evidence through a long-difference approach, which effectively eliminates serial dependence, albeit at the cost of reducing the available sample size. We discuss the robustness check in Section 4.2.1. Third, as suggestive evidence, we address serial dependence by employing sparse sampling. Specifically, we restrict the sample to observations from the years 2000, 2003, 2006, 2009, 2012, 2015, and 2018, and re-estimate our baseline regression. This approach mitigates the serial dependence of the independent variable. The results, reported in Table A.8, are similar to our baseline findings.[12]

The last major concern is that patents may not accurately capture real innovation. For example, Wei et al. (2021) highlight that China's patent subsidy policy (the InnoCom program), introduced in 2008, promoted patent trading among firms and may have artificially inflated the number of patent applications, thereby reducing their effectiveness as a measure of corporate innovation intensity. To address this concern, we restrict our sample to the period between 2001 and 2008; that is, prior to the implementation of the InnoCom program, as shown in Table A.10. Our baseline results remain intact.

One might be interested in the relationship between time-varying controls and patent filings (normalized per 1,000 people). In Table A.2, we present the coefficients of time-varying controls

[12]To ensure that the results are not sensitive to the choice of gap length, we also restrict the sample to observations from the years 2000, 2004, 2008, 2012, and 2016. The corresponding results are reported in Table A.9.

on patent filings (normalized per 1,000 people) in column (1), based on our preferred specification as shown in column (5) of Table 1. In column (2) of Table A.2, we show the relationship between time-varying controls and surname homogeneity among business owners.

## 4.2 Robustness Checks

This sub-section reports important robustness checks on our baseline results.

### 4.2.1 Long Difference Approach

We adopt a long-difference approach to address serial correlation in surname homogeneity among business owners. We examine the relationship between the change in surname homogeneity and the change in patent filings over a 20-year period. By taking differences within a county, we can effectively eliminate time-invariant confounders. Specifically, we estimate the following specification:

$$\Delta Y_{c,t+1} = \alpha + \beta \Delta SurnameHomogeneity_{c,t} + X'_{c,t}\gamma + \varepsilon_{c,t} \tag{4.2}$$

Here, $\Delta Y_{c,t+1}$ represents the change in innovation in county $c$ over the period from 2001 to 2020, and $\Delta SurnameHomogeneity_{c,t}$ represents the change in surname homogeneity of entrepreneurs in county $c$ over the period from 2000 to 2019. As in our baseline estimation, the outcome variable is delayed by one year relative to the explanatory variable to account for the time lag between innovation and patent applications. We include the same set of controls as in Equation 4.1, using their values in 2000 for all time-varying controls, ensuring they are pre-determined.

Table 2 presents the above robustness check results as we progressively include more controls across columns. A significantly positive correlation is observed in all columns. Specifically, column (5), which includes the full set of controls, represents our preferred specification. According to this specification, a one-unit increase in the change of surname homogeneity among business owners corresponds to a 0.033-unit increase in the change of innovation output. This represents a 17.6% increase relative to the mean value of the outcome variable. These findings suggest that over a 20-year period, within a county, an increase in surname homogeneity among business owners is associated with a rise in patent filings per capita. This result is consistent with our baseline estimation in Table 1.

### 4.2.2 Other Robustness Checks

#### a. Patent Filings over Longer Periods

The baseline result is the regression relationship between the number of patent filings in year $t+1$ and surname homogeneity among business owners in year $t$. Patent filings, however, could

Table 2: Long Difference Approach

| Dependent variable | Change in patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Change in Surname homogeneity | 0.115*** | 0.069*** | 0.043*** | 0.032*** | 0.033*** |
| | (0.006) | (0.006) | (0.005) | (0.005) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Observations | 2,678 | 2,678 | 2,678 | 2,678 | 2,622 |
| $R^2$ | 0.146 | 0.229 | 0.348 | 0.397 | 0.424 |
| Mean dependent vars. | 0.186 | 0.186 | 0.186 | 0.186 | 0.188 |

Notes: County-level OLS regressions. The outcome variable is the change in patent filings per 1,000 people over the period from 2001 to 2020. The explanatory variable is the change in surname homogeneity among business owners over the period from 2000 to 2019. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

be affected by surname homogeneity in the longer past. To account for potential time lags, we consider the cumulative number of patent filings over longer periods: from $t+1$ to $t+2$, $t+1$ to $t+3$, $t+1$ to $t+4$, and $t+1$ to $t+5$. For this analysis, we replicate the baseline model specified in Equation 4.1, replacing the outcome variable with the cumulative number of patent filings over these extended periods. The results, presented in Tables A.11, A.12, A.13, and A.14, correspond to the four respective timeframes. All these results consistently confirm our baseline findings.

**b. Exclude Special Samples**

We exclude three types of special samples that could potentially affect our results. The first type consists of counties with a significant presence of minority populations. According to the 2005 population census, over 90% of China's population is ethnic Han. However, some counties have a high proportion of minorities, resulting in a distinct cultural and socioeconomic environment. Therefore, we exclude counties with minority populations exceeding 10%. The second type includes municipalities directly administered by the central government–Beijing, Shanghai, Tianjin, and Chongqing. These municipalities function as provinces, with an administrative level higher than other cities, potentially influencing policy implementation and economic activities. The third type comprises special economic zones (Shenzhen, Zhuhai, Shantou, Xiamen, and Hainan), whose distinct institutional arrangements may foster greater innovation through alternative approaches.

The estimates excluding these three types of special samples are presented in Table A.15, Table A.16 and Table A.17, respectively. As shown, our results remain robust across all these tables. Notably, the magnitudes in Table A.15 are larger compared to our baseline estimation, suggesting that minority populations may have influenced our baseline results.

**c. Another Measure of Surname Homogeneity: HHI**

In our baseline estimation, we quantify surname homogeneity using an entropy-based measure, which has a well-established theoretical foundation in information theory. To ensure the robustness of our results, we also employ an alternative and widely used measure of concentration, the Herfindahl-Hirschman Index (HHI) as below:

$$SurnameHomogeneity_{c,t} = \sum_{k} p_{c,t,k}^2 \tag{4.3}$$

where $p_{c,t,k}$ is the fraction of entrepreneurs with surname $k$ in county $c$ in year $t$. A higher value of this measure indicates greater surname concentration. However, this measure is highly skewed and generates several outliers. To mitigate the influence of these outliers, we winsorize the data at the 5% and 95% levels. The results, presented in Table A.18, demonstrate that this measure of surname homogeneity continues to predict corporate innovation positively.

**d. Including More Controls**

Our baseline estimation controls for four sets of variables. However, concerns about potential omitted variable bias may still arise. To address this issue and ensure robustness, we incorporate additional variables that could confound our estimation.

First, higher levels of innovation may be driven by overall economic development. Based on our baseline estimation, we further control for a measure of economic development, nighttime light intensity, in Table A.19.[13] Second, we account for industry homogeneity, which could affect innovation outcomes. The method for measuring industry homogeneity is analogous to that used for surname homogeneity, but instead of the fraction of specific surnames, we use the fraction of firms sorted by specific industries. The results are shown in Table A.20. Third, we include industry structure as a control, measured by the fraction of firms of different sizes.[14] High fractions of smaller firms could correlate with both surname homogeneity and innovation. The results are reported in Table A.21. Fourth, we include the dynamic effects of clan culture, proxied by the number of genealogies per 1,000 people, interacted with year dummies.[15] This control is included to address the possibility that our results are driven by clan culture rather than surname homogeneity. The corresponding results are displayed in Table A.22. Fifth, we control for agricultural diversity interacted with year dummies, as it may influence innovation (Fiszbein, 2022).[16] The results are displayed in Table A.23. Finally, we include a treaty port dummy interacted with year dummies, as treaty ports may have influenced cultural characteristics and innovation through exposure to foreign cultures. The results

[13] Even though we use the lagged value of nighttime light intensity, it may be a bad control, as innovation could also influence economic development.

[14] (1) firms with registered capital smaller than 100,000 CNY, (2) firms with capital between 100,000 and 1,000,000 CNY, (3) firms with capital between 1,000,000 and 5,000,000 CNY, and (4) firms with capital between 5,000,000 and 10,000,000 CNY.

[15] The number of genealogies in each county, normalized by 1,000 local residents, serves as the measure of clan culture. Since it is time-invariant, this variable is interacted with year dummies.

[16] We measure agricultural diversity by considering the caloric contribution of five key crops: rice, wheat, maize, sweet potato, and white potato. This diversity is quantified using an entropy-based functional form, analogous to our measure of surname homogeneity, as outlined in Equation 3.1.

are presented in Table A.24. As evident from the results across all these tables, our findings remain robust, indicating that the estimation is unlikely to be significantly affected by omitted variable bias.

**e. Granted Patents**

Our baseline regression follows the common practice of using the number of patent applications filed as a proxy for corporate innovation, as the granting of a patent involves uncertain time lags. However, it is unclear whether an application will ultimately be successful. Additionally, some applications may be filed for strategic purposes, which may not reflect genuine innovation. To address these concerns, we replace the original outcome variable with the number of patents granted (normalized by 1,000 people) and reestimate our baseline model as specified in Equation 4.1. The results, presented in Table A.25, indicate that this alternative measure of innovation does not alter our findings. The positive correlation observed with both measures of innovation strengthens our confidence in our findings.

## 4.3 Instrumental Variable Estimation

Within a county, changes in surname homogeneity can be influenced by various factors, such as historical agricultural practices, migration, and cultural developments. While we have controlled for a comprehensive set of covariates, the possibility of omitted variable bias remains. To address this concern, we employ an instrumental variable strategy that leverages exogenous variation in rice suitability.

Rice suitability may serve as a valid predictor of the strength of informal institutions because rice cultivation requires substantial shared water usage and intensive collective coordination to develop and maintain irrigation systems, efforts historically facilitated by kinship-based institutions (Freedman, 1966; Talhelm et al., 2014; Jia and Kung, 2025). Consequently, counties with higher rice suitability tend to exhibit greater surname concentration, as the agricultural advantages of rice cultivation favored large clan-based farming structures, which in turn help sustain stronger informal institutions. In modern society, dominant lineages can leverage accumulated resources, such as land, capital, and kinship networks, to enter and expand within markets. Their scale advantages may progressively crowd out smaller competitors, thereby amplifying surname concentration and strengthening informal institutions. As a result, the homogenizing effect of rice suitability may intensify over time. Since our measure of informal institutions–surname homogeneity among business owners–varies over time, we interact rice suitability with year to capture its dynamic influence. Specifically, the first-stage regression is specified as follows:

$$SurnameHomogeneity_{c,t} = \alpha + \beta ln(RiceSuitability_c + 1) \times Year_t + X'_{c,t}\gamma + \mu_c + \delta_t + \varepsilon_{c,t} \quad (4.4)$$

where $Rice\ Suitability_c$ represents the rice suitability for county $c$. We use the functional form

$ln(Rice\ Suitability_c + 1)$ due to the highly skewed distribution of the variable and the presence of zero values.[17] Control variables and fixed effects are consistent with those used in our baseline Equation 4.1.

Table 3 presents the instrumental variable (IV) estimates using the interaction between rice suitability and year as the instrument. The first-stage results show that the instrument has strong predictive power: the coefficient on the instrument is statistically significant at the 1% level across all specifications. In Column (5), our preferred specification with the full set of controls, the Kleibergen-Paap F-statistic is 60.934, well above the commonly used threshold of 10, indicating that weak instrument concerns are minimal. In the second stage, surname homogeneity consistently exhibits positive and statistically significant effects on corporate innovation across all columns. This consistency reinforces the positive direction and robustness of our baseline findings.

The IV estimates are larger than the OLS estimates. One possible explanation is the presence of a local average treatment effect (LATE). The IV strategy identifies the effect for complier counties, those whose informal institutions evolved in response to their suitability for rice cultivation. These counties are likely to be ones where informal institutions are more functional or economically salient. In such contexts, increases in surname homogeneity may reflect genuinely effective kinship-based institutions that reduce transaction costs and enhance mutual trust and support. As a result, these counties likely experienced particularly large innovation gains from stronger informal institutions. A second explanation relates to measurement error. Surname homogeneity may not fully capture the strength of informal institutions, especially if it has weaker institutional implications in certain counties. In such cases, attenuation bias in the OLS estimates could lead to an underestimation of the true effect.

## 4.4 Heterogeneity of Institutional Development Analyses

To provide corroborative evidence that surname homogeneity among business owners is associated with corporate innovation—based on the premise that kinship-based informal institutions can supplement weak formal institutions—we examine four substitution channels, each guided by theoretical considerations.

### 4.4.1 Intellectual Property Rights Protection

First, kinship-based informal institutions can supplement formal institutions by providing protection for intellectual property rights. When intellectual property rights protection is weak, strong kinship-based connections among entrepreneurs can mitigate peer thefts of new ideas and technologies and help to safeguard each other's intellectual property against outsiders' stealing. In contrast, when formal institutions provide robust intellectual property rights protection, individuals rely less

[17] Our results remain robust when the variable is used in levels without log transformation.

Table 3: IV: Interaction Between Rice Suitability and Year

| Second-stage dependent variable | Patents per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.578*** | 0.846*** | 0.884*** | 1.152*** | 1.185*** |
| | (0.097) | (0.158) | (0.181) | (0.209) | (0.220) |
| First-stage dependent variable | Surname homogeneity | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| ln(Rice suitability+1) $\times$Year | 0.002*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,455 | 53,335 | 53,335 | 53,335 | 52,295 |
| $R^2$ | 0.829 | 0.867 | 0.870 | 0.898 | 0.899 |
| KP F-statistics | 88.860 | 63.231 | 53.286 | 62.763 | 60.934 |

Notes: County-level 2SLS regressions. The outcome variable in the second stage is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable in the second stage is the surname homogeneity among business owners. IV is the interaction term between rice suitability and year. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. index. We use two-way clustering of standard errors at the county and year level. * p<0.1; ** p<0.05; *** p<0.01.

on kinship-based informal institutions. Furthermore, formal intellectual property rights protection may encourage more entrepreneurs outside the established networks to innovate. Consequently, the role of kinship-based informal institutions is expected to diminish as the level of intellectual property rights protection offered by formal institutions increases.

We explore this by examining the regression relationship between corporate innovation, surname homogeneity among business owners, and the proxy for intellectual property rights protection, as well as the interaction between the latter two terms. Specifically, we estimate the following equation:

$$\begin{aligned} Y_{c,t+1} = \alpha + \beta_1 SurnameHomogeneity_{c,t} \times IPR_{p,t} + \beta_2 SurnameHomogeneity_{c,t} \\ + \beta_3 IPR_{p,t} + X'_{c,t}\gamma + \mu_c + \delta_t + \varepsilon_{c,t} \end{aligned} \tag{4.5}$$

where $IPR_{p,t}$ represents the measure of intellectual property rights protection in province $p$, year $t$. Other notations are consistent with those in our baseline estimation in Equation 4.1.

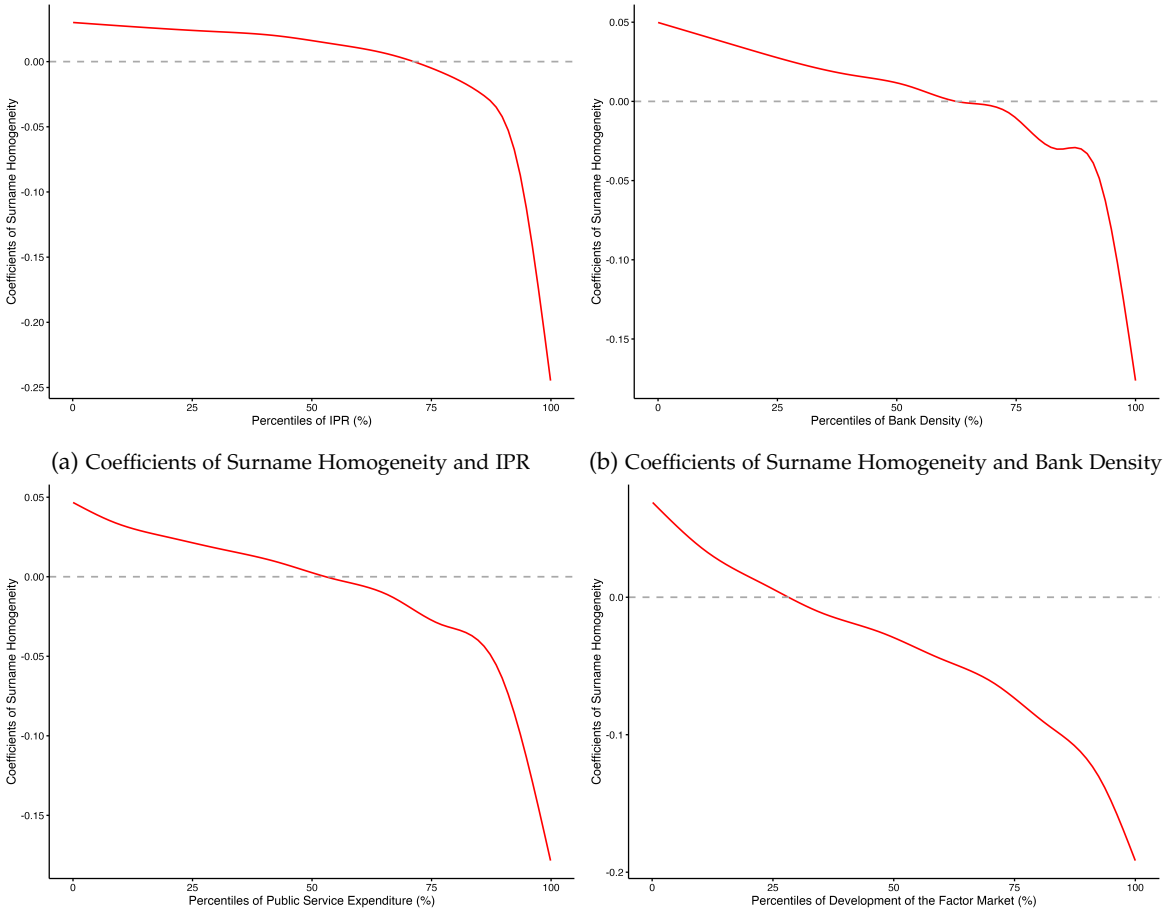
The results are presented in Table 4. Surname homogeneity, IPR, and their interaction term are jointly significant; e.g., the F-statistic of the coefficients in column (5) is 26.4. The coefficients of the interaction terms are significantly negative. The patterns of regression coefficients for the interaction term and surname homogeneity in all columns in Table 4 suggest that surname homogeneity is generally associated with higher innovation; however, this effect weakens as available intellectual property rights (IPR) protection strengthens and may become negative. We plot the overall coefficients of surname homogeneity with respect to IPR protection based on the estimates from column (5). As shown in Figure 2a, when intellectual property rights protection reaches its 71.7% percentile, the coefficient of surname homogeneity on corporate innovation turns from positive to negative. The results indicate that surname homogeneity fosters innovation more strongly when intellectual property rights protection is weaker. When intellectual property rights protection is strong enough, surname homogeneity among business owners becomes negatively correlated with corporate innovation.[18]

[18]IPR itself is positively correlated with corporate innovation when surname homogeneity is not too high. For example, in column (5) of Table 4, given the mean value of surname homogeneity (-5.699), the coefficient of IPR should be 0.006 (-0.028+(-0.006)*(-5.699). Generally, the coefficient on IPR is positive when surname homogeneity is below -4.7, which applies to approximately 94.7% of the observations. However, for the remaining 5.3% of observations, where surname homogeneity exceeds -4.7, IPR may hinder innovation.

Table 4: Heterogeneity Analysis: Intellectual Property Rights Protection

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × IPR | -0.010*** | -0.009*** | -0.007*** | -0.006*** | -0.006*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| Surname homogeneity | 0.087*** | 0.054*** | 0.041*** | 0.032*** | 0.031*** |
| | (0.013) | (0.007) | (0.006) | (0.006) | (0.006) |
| IPR | -0.055*** | -0.048*** | -0.036*** | -0.029*** | -0.028*** |
| | (0.009) | (0.007) | (0.005) | (0.004) | (0.004) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 51,048 | 50,934 | 50,934 | 50,934 | 49,870 |
| $R^2$ | 0.670 | 0.694 | 0.709 | 0.726 | 0.738 |
| Mean dependent vars. | 0.072 | 0.072 | 0.072 | 0.072 | 0.072 |

Notes: County-level OLS regressions. "IPR" indicates the intellectual property rights protection index. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2019 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.




(a) Coefficients of Surname Homogeneity and IPR

(b) Coefficients of Surname Homogeneity and Bank Density

(c) Coefficients of Surname Homogeneity and Expenditure

(d) Coefficients of Surname Homogeneity and Factor

Figure 2: Coefficients of Surname Homogeneity with Respect to the Strength of Formal Institutions

Note: These four figures show the relationships between the coefficients of surname homogeneity and IPR protection, bank density, public service expenditure, and factor market development, respectively.

#### 4.4.2 Financial Development

Second, kinship-based informal institutions can provide financial support to their members to fuel innovation. As discussed in Chen et al. (2022), kinship networks and financial markets act as competing substitutes in China. The R&D process requires significant capital, and entrepreneurs can raise funds from their kinship networks when access to formal financial institutions is limited. However, the role of kinship networks in financing innovation is likely to diminish as financial development progresses.

We measure the level of financial development using two proxies. The first proxy is bank density, defined as the number of banks normalized by the local population. The second proxy is the ratio of total loans and deposits from formal financial institutions to GDP. These two variables capture the availability of formal financial services. The first variable varies at the county level, while the second variable varies at the province level. We employ a specification similar to Equation 4.5, but replace the measures of intellectual property rights protection with those of financial development.

Table 5 and Table 6 present the results using the two financial development proxies. In both tables, the three key independent variables are jointly significant. For example, for column (5) of Table 5, the F-statistic for the joint significance of surname homogeneity, bank density, and their interaction is 50.03. Also, the interaction between surname homogeneity and financial development is negative and significant in all columns, indicating that the positive relationship between entrepreneurial surname homogeneity and corporate innovation weakens with financial development and may eventually become negative. We plot, in Figure 2b, the overall coefficients of surname homogeneity with respect to bank density based on the estimates from column (5) of Table 5. The turning point occurs at a bank density of 0.149 (0.054 / 0.363). For 64.2% of the observations (where bank density is below 0.149), surname homogeneity positively correlates with innovation; for the remaining 35.8%, the correlation is negative.[19]

[19] Bank density itself predicts innovation positively: at the mean level of surname homogeneity (-5.699), its coefficient is 0.433 (-1.636 + (-0.363) $\times(-5.699)$).

Table 5: Heterogeneity Analysis: Bank Density

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Bank density | -0.529*** | -0.517*** | -0.424*** | -0.371*** | -0.363*** |
| | (0.073) | (0.068) | (0.060) | (0.058) | (0.058) |
| Surname homogeneity | 0.127*** | 0.083*** | 0.064*** | 0.054*** | 0.054*** |
| | (0.015) | (0.009) | (0.008) | (0.008) | (0.008) |
| Bank density | -2.434*** | -2.421*** | -1.947*** | -1.675*** | -1.636*** |
| | (0.439) | (0.413) | (0.356) | (0.340) | (0.334) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,635 | 53,515 | 53,515 | 53,515 | 52,495 |
| $R^2$ | 0.673 | 0.696 | 0.712 | 0.720 | 0.723 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. "Bank density" is the number of banks normalized by the local population. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table 6: Heterogeneity Analysis: Financial Development

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Financial | -0.084*** | -0.099*** | -0.073*** | -0.062*** | -0.055*** |
| | (0.009) | (0.010) | (0.009) | (0.008) | (0.007) |
| Surname homogeneity | 0.294*** | 0.270*** | 0.200*** | 0.168*** | 0.151*** |
| | (0.034) | (0.027) | (0.024) | (0.021) | (0.019) |
| Financial | -0.499*** | -0.596*** | -0.441*** | -0.388*** | -0.351*** |
| | (0.062) | (0.060) | (0.055) | (0.052) | (0.045) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.642 | 0.677 | 0.695 | 0.710 | 0.725 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. "Financial" indicates financial development, measured by the ratio of total loans and deposits from formal financial institutions to GDP. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

#### 4.4.3 Quality of Public Administration

Third, kinship-based informal institutions can provide public services and reduce the risk of expropriation by local governments (Easterly and Levine, 1997; Zhang, 2020). If this is the case, it could be another potential channel through which surname homogeneity facilitates innovation, as better public services, such as maintaining safe environments and infrastructure, benefit innovation. Moreover, the effectiveness of this channel likely depends on the quality of public administration. Specifically, when the quality of public administration is lower, surname homogeneity should play a more critical role in promoting innovation.

We measure public administration quality using the annual public service expenditure normalized by GDP in 2010 at the provincial level.[20] We also employ a specification similar to Equation 4.5, but replace the measure of intellectual property rights protection with that of public administration quality.

Table 7 reports the regression results. The coefficients for surname homogeneity, public service expenditure, and their interaction term are jointly significant, as indicated by an F-statistic of 22.86 based on the estimation in column (5). The coefficients of the interaction terms are negative, suggesting that the positive influence of surname homogeneity in fostering corporate innovation diminishes as public service provision improves. In such cases, entrepreneurs increasingly rely on formal public institutions for the provision of public services rather than support from informal kinship-based networks. We plot the overall coefficients of surname homogeneity with respect to public service expenditure based on the estimates from column (5) of Table 7. As shown in Figure 2c, the turning point occurs when public service expenditure equals 0.03 (calculated as 0.072/2.368). In 52.8% of observations where public service expenditure is below the threshold, surname homogeneity correlates positively with innovation; in the remaining 47.2%, where public service expenditure exceeds the threshold, the correlation turns negative.

[20] The data on public service expenditure is available only from 2007.

Table 7: Heterogeneity Analysis: Public Service Expenditure

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Expenditure | -2.716*** | -3.676*** | -3.154*** | -2.618*** | -2.368*** |
| | (0.487) | (0.506) | (0.423) | (0.392) | (0.355) |
| Surname homogeneity | 0.208*** | 0.144*** | 0.119*** | 0.072*** | 0.072*** |
| | (0.023) | (0.019) | (0.016) | (0.021) | (0.020) |
| Expenditure | -17.014*** | -21.253*** | -18.151*** | -14.378*** | -12.767*** |
| | (2.946) | (2.986) | (2.469) | (2.296) | (2.036) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 34,931 | 34,853 | 34,853 | 34,853 | 34,125 |
| $R^2$ | 0.763 | 0.782 | 0.786 | 0.792 | 0.798 |
| Mean dependent vars. | 0.115 | 0.115 | 0.115 | 0.115 | 0.116 |

Notes: County-level OLS regressions. "Expenditure" indicates the annual public service expenditure normalized by GDP in 2010. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2008 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

#### 4.4.4 Development of the Market Forces

Fourth, kinship-based informal institutions can reduce coordination costs and facilitate the organization of production. When markets are underdeveloped, these informal institutions may help secure production factors and facilitate product sales, thereby improving supply chain efficiency. However, this function will likely reduce as markets become more established and operate effectively.

We explore this possibility using the interaction term between surname homogeneity and the strength of market forces. Specifically, we consider two indices: the index of development of the factor market and the index of development of the product market. Our estimation follows a specification similar to that in Equation 4.5.

The results are presented in Table 8 and Table 9, respectively. In all columns of both tables, the coefficients for surname homogeneity, the index of market development, and their interaction terms are jointly significant. For example, based on the regression in column (5) of Table 8, the F-statistic is 24.55. Column (5) of Table 8, which represents our preferred specification, shows that the coefficients of the interaction terms are significant and negative, suggesting that the positive role of surname homogeneity diminishes as market mechanisms improve. Based on the estimates in column (5) of Table 8, Figure 2d illustrates this negative relationship. The turning point occurs when the index of factor market development reaches 4.79 (computed as $0.091/0.019$). In 27.6% of observations where the index is below the threshold, surname homogeneity is positively associated with innovation; in the remaining 72.4%, the association becomes negative.

Table 8: Heterogeneity Analysis: Development of the Factor Market

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Factor market | -0.028*** | -0.028*** | -0.023*** | -0.021*** | -0.019*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Surname homogeneity | 0.172*** | 0.134*** | 0.106*** | 0.098*** | 0.091*** |
| | (0.028) | (0.020) | (0.018) | (0.015) | (0.015) |
| Factor market | -0.156*** | -0.160*** | -0.127*** | -0.121*** | -0.109*** |
| | (0.019) | (0.019) | (0.017) | (0.016) | (0.015) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.663 | 0.690 | 0.704 | 0.717 | 0.731 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. "Factor market" indicates the index of development of the factor market. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table 9: Heterogeneity Analysis: Development of the Product Market

| Dependent variable | Patents filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Product market | -0.004 | -0.001 | -0.004 | -0.001 | -0.001 |
| | (0.005) | (0.005) | (0.004) | (0.003) | (0.002) |
| Surname homogeneity | 0.131*** | 0.066* | 0.065** | 0.038 | 0.033* |
| | (0.028) | (0.033) | (0.025) | (0.023) | (0.018) |
| Product market | -0.019 | -0.004 | -0.017 | -0.000 | 0.003 |
| | (0.030) | (0.031) | (0.022) | (0.019) | (0.015) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.628 | 0.658 | 0.685 | 0.704 | 0.720 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. "Product market" indicates the index of development of the product market. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

## 4.5 Change in the Correlation over Time

Another way to examine the role of surname homogeneity among business owners in shifting from supporting to hindering innovation as formal institutions develop is to investigate how this change unfolds as China's overall institutional framework improves over time.

Between 2001 and 2020, China experienced significant advancements in formal institutions. From Figure A.3a to Figure A.3f, we show the changes in the six time-variant proxies for formal institutions in our heterogeneity analyses, reflecting the formal institutions' improvement over the past 20 years. Additionally, the Global Innovation Index published by WIPO reflects China's institutional development over time: China ranked 79th in institutional development in 2016, improved to 62nd in 2020, and reached 44th in 2024. These trends highlight the development of China's formal institutions alongside its economic growth.

Given that the positive correlation between surname homogeneity among business owners and corporate innovation is tied to the role of informal institutions in fostering innovation, we expect that this correlation diminishes over time as formal institutions strengthen.

To test this hypothesis, we examine how the correlation between surname homogeneity among business owners and innovation evolves over time using the following econometric specification:

$$Y_{c,t+1} = \alpha + \sum_{2000}^{2019} \beta_t SurnameHomogeneity_{c,t} \times Year_t + X'_{c,t}\gamma + \mu_c + \delta_t + \varepsilon_{c,t} \tag{4.6}$$

where $Year_t$ denotes year dummies, and $\beta_t$ represents the corresponding year-specific coefficients. Other notations are as previously defined.

This specification allows the correlation between surname homogeneity among business owners and innovation to vary flexibly by year. We plot the year-specific coefficients in Figure 3, which correspond to those reported in column (5) of Table A.3. As shown in the figure, the correlation is positive in the early 2000s, gradually declines over time, and turns negative after 2010. This pattern suggests that the role of kinship-based informal institutions in promoting corporate innovation diminishes over time as formal institutions improve, to the extent that it switches to play a marginally significant negative role since 2011.

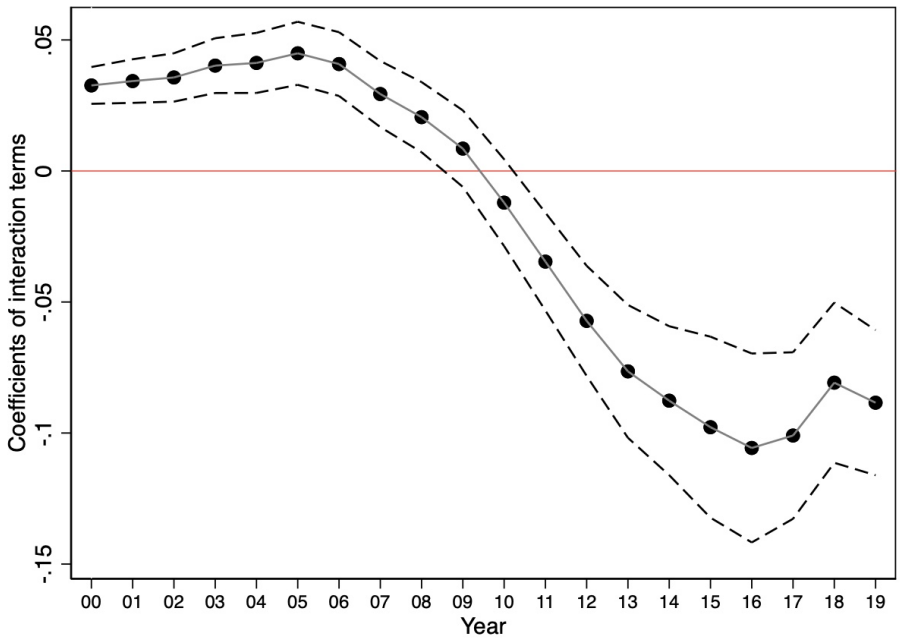



Figure 3: The Correlation over Time

Notes: This figure shows the correlation between surname homogeneity among business owners and innovation over time. We plot the coefficients and the 95% confidence intervals, as reported in column (5) of Table A.3.

# 5 Policy Innovation

## 5.1 Baseline Estimation

We now turn to whether surname homogeneity among business owners in the private sector could be conducive to policy innovation. This is an important topic for emerging economies, which often have to experiment with policy innovation to foster development toward a market-based economy.

Easterly and Levine (1997) argue that ethnic fragmentation in Sub-Saharan Africa hinders the development of effective public policy by undermining consensus on growth-promoting measures. In contrast, regions with high surname homogeneity among business owners–often characterized by kinship-based informal institutions–are more conducive to policy experimentation and consensus-driven economic development. When a local government advances a policy innovation, it requires cooperation from the private sector. Such cooperation is easier to negotiate and monitor when there are only a few dominant families involved, compared to a fragmented business landscape with many unrelated owners. The former context entails lower transaction costs related to negotiation, enforcement, and coordination, as well as greater capacity to reward cooperation and punish non-cooperation. As a result, innovative policies are likely to be more effective in more homogeneous business communities. The conjecture is consistent with Wang and Yang (2021) and Chen et al. (2024) who point out that China's policy experimentation has a positive selection bias in that policy experiments are implemented in locations with favorable conditions. Typically, the central government chooses experimentation sites at the provincial level, and then the provincial governments further delegate the experimentation to specific favored prefectural cities or counties within their jurisdictions (See also Qian and Xu (1993) and Qian et al. (2006)).

To test this hypothesis, we estimate Equation 4.1, replacing the original outcome variable with the number of policy innovations implemented at the county level.[21] In this analysis, we use the raw number of policy innovations without normalization. This choice is motivated by the fact that policy innovations are far less frequent than patent filings and are less directly tied to local population size. Nonetheless, our results remain robust when the outcome is normalized by the local population, as shown in Table A.26. Control variables and other settings remain the same as in our baseline estimation. The results are presented in Table 10. As shown, counties with higher surname homogeneity among business owners are associated with a greater number of policy innovations, which is consistent with our hypothesis. Specifically, a one-standard-deviation increase in surname homogeneity leads to an approximately 18.9% increase in the number of policy innovations, relative to the sample mean–an economically meaningful magnitude.

[21] To maintain consistency with our baseline estimation, the outcome variable is winsorized at the 5% and 95% percentiles. However, the results are also robust to alternative winsorization thresholds, such as the 1% and 99% percentiles, as reported in Table A.27.

Table 10: Surname Homogeneity among Business Owners and Policy Innovation

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.107*** | 0.066*** | 0.056*** | 0.050*** | 0.047*** |
| | (0.010) | (0.011) | (0.011) | (0.011) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.242 | 0.253 | 0.255 | 0.261 | 0.265 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

To address concerns regarding serial dependence in surname homogeneity, we employ three approaches similar to those used in Section 4.1. First, we apply Newey-West standard errors to account for potential autocorrelation; the results are reported in Table A.28. Second, we implement a long-difference specification, analogous to that in Section 4.2.1, using changes in the number of policy innovations as the outcome variable. The results are presented in Table A.29. Third, we adopt a sparse sampling strategy by restricting the sample to observations from the years 2000, 2003, 2006, 2009, 2012, 2015, and 2018. The results are shown in Table A.30.[22] Across all three strategies, the results remain consistent with our baseline findings on policy innovation, suggesting that serial dependence in surname homogeneity does not pose a serious concern.

## 5.2 Robustness Checks

We conduct a series of robustness checks, analogous to those presented in Section 4.2.2, to ensure that our findings remain consistent across different specifications when the outcome variable is the number of policy innovations.

**a. Policy Innovations over Longer Periods**

There may be time lags between the design of a policy and the official announcement of its experimentation. To account for potential time lags, we examine the cumulative number of policy

[22]To ensure that our findings are not sensitive to the choice of interval length, we also restrict the sample to the years 2000, 2004, 2008, 2012, and 2016. The corresponding results are reported in Table A.31.

innovations over extended time horizons, following the same approach as in the case of patents. Specifically, we consider the periods spanning t+1 to t+2, t+1 to t+3, t+1 to t+4, and t+1 to t+5. We modify the baseline model specified in Equation 4.1 by replacing the outcome variable with the cumulative number of policy innovations over each of these extended periods. The results, presented in Tables A.32, A.33, A.34, and A.35, correspond to these four respective timeframes. Consistent with our baseline findings, all results strongly corroborate our main conclusions.

**b. Exclude Special Samples**

As we discuss in 4.2.2, three types of special samples could affect the results, thus we exclude: (1) counties with minority populations exceeding 10% (due to distinct socio-cultural environments), (2) the four centrally-administered municipalities (Beijing, Shanghai, Tianjin, Chongqing) given their unique provincial-level status and potential impact on policy implementation, and (3) special economic zones (Shenzhen, Zhuhai, Shantou, Xiamen, and Hainan) because of the relative different institutional arrangements. The estimates excluding these three types of special samples are presented in Table A.36, Table A.37 and Table A.38, respectively. As shown, our results remain robust across all these tables.

**c. Another Measure of Surname Homogeneity: HHI**

To ensure that our results are robust to alternative functional forms used to measure homogeneity, we reestimate the model using the HHI, that is formula 4.3, as the explanatory variable.[23] The results, presented in Table A.39, closely mirror our baseline estimates.

**d. Including More Controls**

To further mitigate potential omitted variable bias beyond our baseline controls, we conduct an analogous set of robustness checks on policy innovation outcomes, incorporating the same suite of additional control variables used in our patent analysis. Specifically, we sequentially include controls for: (1) Economic development (nighttime light intensity); (2) Industry homogeneity (measured analogously to surname homogeneity); (3) Industry structure (fractions of firms by size categories); (4) Dynamic clan culture effects (genealogies per 1000 people $\times$ year dummies); (5) Agricultural diversity $\times$ year dummies (entropy-based crop diversity); and (6) Treaty port status $\times$ year dummies.

The results, shown in Tables A.40–A.45 (mirroring the structure of our patent robustness checks), consistently demonstrate that our basic finding linking surname homogeneity to policy innovation remains robust. This indicates that our estimation for policy innovation is unlikely to be significantly biased by these potential confounders.

[23] The HHI is winsorized at the 5th and 95th percentiles, the same as before.

## 5.3 Instrumental Variable Estimation

We conduct the same IV analysis, using a first-stage specification identical to Equation 4.4. In the second stage, we replace the outcome variable with the number of policy innovations. Table A.46 reports the results. The first-stage estimates once again confirm the strong predictive power of the instrument: the interaction term $ln(Rice\ Suitability + 1) \times Year$ is statistically significant at the 1% level across all specifications. In our preferred specification, shown in Column (5) with the full set of controls, the Kleibergen-Paap F-statistic is 45.254, comfortably exceeding the conventional weak-instrument threshold of 10. In the second stage, surname homogeneity continues to exhibit consistently positive and statistically significant effects on policy innovation. These findings reinforce the baseline pattern, suggesting that surname homogeneity promotes the experimentation of policy innovations.

## 5.4 Heterogeneity of Institutional Development Analyses

As demonstrated in Section 5.1, stronger informal institutions foster experimentation with policy innovation. One plausible explanation is that informal institutions facilitate cooperation between local governments and business communities, making it easier to coordinate and implement policy experiments. We hypothesize, however, that the role of informal institutions on policy innovation is contingent upon the strength of local formal institutions. Specifically, when formal institutions, such as legal enforcement, bureaucratic capacity, and regulatory infrastructure, are sufficiently developed, local governments can rely on these mechanisms to support policy experimentation. In such contexts, the contribution of informal institutions may diminish, as formal channels provide sufficient support for policy implementation. Conversely, in settings where formal institutions are weak, local governments may depend more heavily on cohesive and cooperative business communities to cooperate with and sustain policy innovations. This suggests a potential substitutability between informal and formal institutions in facilitating government-led policy experimentation.

We, therefore, conduct a heterogeneity analysis parallel to that presented in Section 4.4, replacing the outcome variable with the number of policy innovations at the county level. To capture the strength of formal institutions, we use the same set of indicators as in Section 4.4, including the intellectual property rights protection index, bank density, financial development, public service expenditure, an index of factor market development, and an index of product market development. The results of these six heterogeneity analyses are reported in Table A.47, Table A.48, Table A.49, Table A.50, Table A.51, and Table A.52, respectively. Across all specifications, we consistently find that the interaction term between surname homogeneity and the level of formal institutional development is negative. This pattern suggests that the positive role of informal institutions in policy innovation becomes weaker as formal institutions become more developed, indicating that informal institutions can partially compensate for the limitations of weak formal institutions in facilitating

policy experimentation.

## 5.5 Change in the Correlation over Time

Similar to Section 4.5, we examine how the correlation between informal institutions and policy innovation evolves over time. We estimate the same regression as in Equation 4.6, replacing the outcome variable with the number of policy innovations and allowing the correlation between informal institutions and policy innovation to vary flexibly by year. We plot the year-specific coefficients in Figure 4, according to the estimates in column (5) of Table A.53. As shown in the figure, the correlation is generally positive in the early 2000s but turns negative after 2010. This pattern suggests that the role of kinship-based informal institutions in promoting policy innovation diminished after 2010. However, the correlation begins to rise again after 2016, and becomes positive and significant in 2019. The ending pattern may be due to complex changes in the economic and political environment or shifts in international relations; this observation awaits further research investigation.

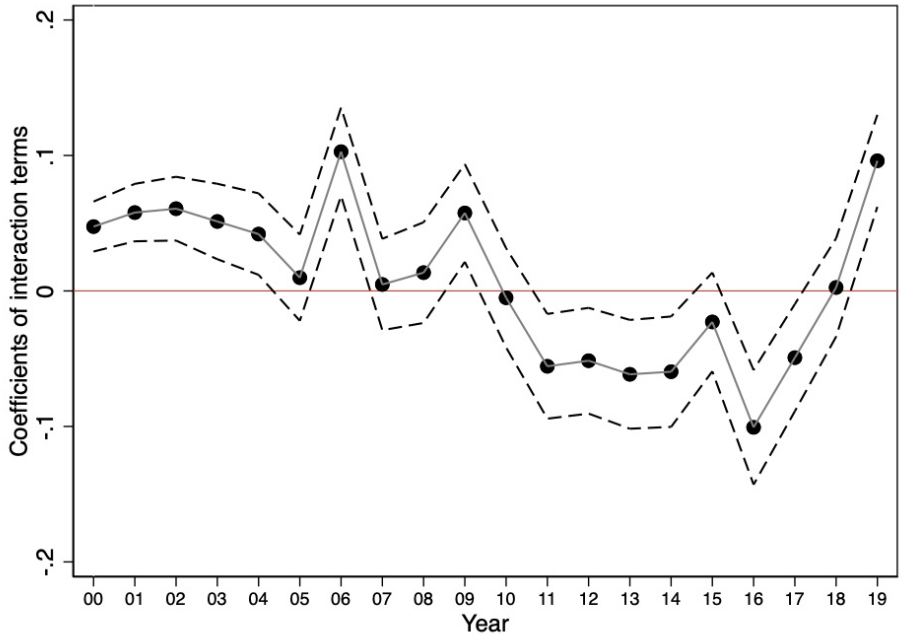



Figure 4: The Correlation over Time

Notes: This figure shows the correlation between surname homogeneity among business owners and policy innovation over time. We plot the coefficients and the 95% confidence intervals, as reported in column (5) of Table A.53.

## 5.6 Policy Innovation and Corporate Innovation

We also examine how experimentation with policy innovation and surname homogeneity collaborate to promote corporate innovation. The results are presented in Table 11. They indicate that both policy innovation and surname homogeneity among entrepreneurs promote corporate innovation.

The two variables are jointly significant. Our results indicate that (i) surname homogeneity among entrepreneurs promotes corporate innovation on its own and by fostering more experimentation with policy innovation; (ii) policy innovation positively supports corporate innovation on its own.

Table 11: Policy Innovation and Corporate Innovation

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.115*** | 0.067*** | 0.049*** | 0.034*** | 0.029*** |
| | (0.014) | (0.008) | (0.006) | (0.006) | (0.006) |
| Policy innovation | 0.034*** | 0.028*** | 0.022*** | 0.022*** | 0.018*** |
| | (0.008) | (0.007) | (0.005) | (0.005) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,919 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.633 | 0.661 | 0.688 | 0.705 | 0.720 |
| Mean dependent vars. | 0.079 | 0.078 | 0.078 | 0.078 | 0.079 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variables are surname homogeneity among business owners and the number of policy innovations. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

# 6 Overall Economic Impact

Our basic premise is that surname homogeneity among business owners signals strong kinship-based informal institutions, which foster innovation in private firms and encourage experimentation with policy innovations. These kinship-based informal institutions enhance economic performance by fostering trust, reducing coordination and transaction costs, and promoting entrepreneurial market entry through improved information sharing, risk mitigation, and stronger social capital. Furthermore, decision-making cadres anticipate ground-level support and cooperation based on these features and initiate policy innovations that enhance economic performance. This conjecture is consistent with the findings of Chen et al. (2024).

To test the conjecture, we employ Equation 4.1 from our baseline specification, including surname homogeneity and policy innovation as explanatory variables, but replace the outcome variable with economic development and entrepreneurial market entries measures, respectively. Specifically, we measure economic development using nighttime light intensity and entrepreneurial market entries

using the number of newly registered private firms (normalized by 1,000 of the local population). These variables are measured at the county level and vary across years.[24] Other settings remain the same as in our baseline estimation.

Table 12 presents the results when the outcome variable is economic development, and Table 13 shows the results for entrepreneurial market entries. Across these tables, we observe that both economic development and entrepreneurial market entries are positively and significantly explained by surname homogeneity among business owners. These findings confirm that kinship-based informal institutions can be conducive to a variety of economic outcomes, complementing our analysis of their impact on corporate and policy innovation. We also find that policy innovation, by itself, is positively correlated with economic development and entrepreneurial market entries, suggesting its direct positive role in contributing to broader economic activity.

Table 12: Economic Development

| Dependent variable | Nighttime light | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 137.746*** | 72.550*** | 63.753*** | 41.220*** | 36.905*** |
| | (19.838) | (17.313) | (16.309) | (13.303) | (12.447) |
| Policy innovation | 21.583*** | 17.033** | 14.967** | 13.223** | 10.229** |
| | (6.913) | (5.965) | (5.384) | (5.260) | (4.722) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,619 | 52,259 | 52,259 | 52,259 | 51,639 |
| $R^2$ | 0.983 | 0.984 | 0.984 | 0.985 | 0.985 |
| Mean dependent vars. | 943.632 | 941.065 | 941.065 | 941.065 | 947.993 |

Notes: County-level OLS regressions. The outcome variable is nighttime light. Time spans from 2001 to 2020 for the outcome variable. The explanatory variables are surname homogeneity among business owners and the number of policy innovations. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

[24] They are also winsorized at the 5% and 95% percentiles to maintain consistency with the baseline estimation.

Table 13: Entrepreneurship

| Dependent variable | New firm per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.708*** | 0.322*** | 0.247*** | 0.085 | 0.044 |
| | (0.088) | (0.062) | (0.058) | (0.063) | (0.055) |
| Policy innovation | 0.142*** | 0.092*** | 0.069** | 0.071** | 0.045* |
| | (0.043) | (0.031) | (0.027) | (0.026) | (0.023) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,879 | 52,759 | 52,759 | 52,759 | 51,719 |
| $R^2$ | 0.820 | 0.835 | 0.838 | 0.857 | 0.865 |
| Mean dependent vars. | 1.660 | 1.658 | 1.658 | 1.658 | 1.661 |

Notes: County-level OLS regressions. The outcome variable is the number of newly established private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variables are surname homogeneity among business owners and the number of policy innovations. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

# 7 Conclusion

Leveraging a comprehensive dataset encompassing the universe of Chinese private firms, we compute surname homogeneity among business owners at the county level and use it as a proxy for kinship-based informal institutions. Our analysis reveals that the proxy positively predicts corporate patent filings by private firms and policy innovation experiments by governments. These relationships hold under county fixed effects, a wide range of control variables, and an instrumental variable approach based on rice suitability. Furthermore, experiments in policy innovation also significantly explain corporate innovation.

We investigate the mechanisms through which kinship-based informal institutions facilitate both corporate innovation and government experimentation with policy innovation. In contexts where formal institutions are weak, informal networks can substitute by protecting intellectual property, improving access to finance, enabling public service delivery, and supporting supply chain coordination. Similarly, when formal institutions are underdeveloped, governments often rely on kinship-based networks to mobilize business communities and foster cooperation in policy experimentation. However, as formal institutions strengthen, the marginal role of kinship-based informal institutions tends to decline, and may even become a hindrance to innovation if they resist change or entrench existing interests. Our empirical findings support this conditional effect.

Finally, we also find that both surname homogeneity among business owners and experimentation with policy innovation contribute to better economic development and more entrepreneurial market entries. These results indicate the holistic relationship between informal institutions and policy innovation, as well as a wide range of indicators of economic development.

Our findings, though based on the context of China, have significant implications for other developing countries. We demonstrate how kinship-based informal institutions promote innovation across four key dimensions, thereby offering a deeper understanding of how firms innovate in the absence of robust formal institutions. Additionally, we reveal a dynamic interplay between informal and formal institutions: while the significance of informal institutions diminishes as formal institutions strengthen, their role is particularly crucial during the earlier stages of economic development, including stimulating policy innovation. This underscores the potential of informal networks in promoting innovation and economic progress when formal structures are still evolving.

# References


**Aghion, Philippe and Peter Howitt**, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2).

**_ , Richard Blundell, Rachel Griffith, Peter Howitt, and Susanne Prantl**, "The effects of entry on incumbent innovation and productivity," *The Review of Economics and Statistics*, 2009, *91* (1), 20–32.

**Alesina, Alberto and Eliana La Ferrara**, "Ethnic diversity and economic performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

**_ , Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**, "Fractionalization," *Journal of Economic Growth*, 2003, *8*, 155–194.

**_ , Johann Harnoss, and Hillel Rapoport**, "Birthplace diversity and economic prosperity," *Journal of Economic Growth*, 2016, *21*, 101–138.

**Ang, James S, Yingmei Cheng, and Chaopeng Wu**, "Does enforcement of intellectual property rights matter in China? Evidence from financing and investment choices in the high-tech industry," *Review of Economics and Statistics*, 2014, *96* (2), 332–348.

**Angelucci, Manuela, Giacomo De Giorgi, Marcos A Rangel, and Imran Rasul**, "Family networks and school enrolment: Evidence from a randomized social experiment," *Journal of Public Economics*, 2010, *94* (3-4), 197–221.

**Ashraf, Quamrul and Oded Galor**, "Cultural diversity, geographical isolation, and the origin of the wealth of nations," Technical Report, National Bureau of Economic Research 2011.

**_ and _** , "The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development," *American Economic Review*, 2013, *103* (1), 1–46.

**Bai, Chong-En, Ruixue Jia, Hongbin Li, and Xin Wang**, "Entrepreneurial reluctance: Talent and firm creation in China," *The Economic Journal*, 2024, p. ueae094.

**Bai, Ying and James Kai sing Kung**, "Surname distance and technology diffusion: the case of the adoption of maize in late imperial China," *Journal of Economic Growth*, 2022, *27* (4), 569–607.

**Bloom, Nicholas and John Van Reenen**, "Measuring and explaining management practices across firms and countries," *The quarterly journal of Economics*, 2007, *122* (4), 1351–1408.

**_ , Mark Schankerman, and John Van Reenen**, "Identifying technology spillovers and product market rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

**Buonanno, Paolo and Paolo Vanin**, "Social closure, surnames and crime," *Journal of Economic Behavior & Organization*, 2017, *137*, 160–175.

**Chen, Heng, Bingjing Li, and Xiaodong Zhu**, "Bottom-Up Institutional Change and Growth in China," 2024.

**Chen, Zhiwu, Chicheng Ma, and Andrew J Sinclair**, "Banking on the Confucian clan: why China developed financial markets so late," *The Economic Journal*, 2022, *132* (644), 1378–1413.

**Chen, Zhiyuan, Jie Zhang, and Wenping Zheng**, "Import and innovation: Evidence from Chinese firms," *European Economic Review*, 2017, *94*, 205–220.

**Corrado, Carol, Charles Hulten, and Daniel Sichel**, "Intangible capital and US economic growth," *Review of income and wealth*, 2009, *55* (3), 661–685.

**Cruz, Cesi, Julien Labonne, and Pablo Querubin**, "Politician family networks and electoral outcomes: Evidence from the Philippines," *American Economic Review*, 2017, *107* (10), 3006–3037.

**Davis, Lewis S**, "Growing apart: The division of labor and the breakdown of informal institutions," *Journal of Comparative Economics*, 2006, *34* (1), 75–91.

**Ding, Xuedong and Jun Li**, *Incentives for innovation in China: Building an innovative economy*, Routledge, 2015.

**Doh, Jonathan, Suzana Rodrigues, Ayse Saka-Helmhout, and Mona Makhija**, "International business responses to institutional voids," 2017.

**Easterly, William and Ross Levine**, "Africa's growth tragedy: policies and ethnic divisions," *The Quarterly Journal of Economics*, 1997, pp. 1203–1250.

**Fang, Jing, Hui He, and Nan Li**, "China's rising IQ (Innovation Quotient) and growth: Firm-level evidence," *Journal of Development Economics*, 2020, *147*, 102561.

**Fang, Lily H, Josh Lerner, and Chaopeng Wu**, "Intellectual property rights protection, ownership, and innovation: Evidence from China," *The Review of Financial Studies*, 2017, *30* (7), 2446–2477.

**Fang, Lily, Josh Lerner, Chaopeng Wu, and Qi Zhang**, "Corruption, government subsidies, and innovation: Evidence from China," Technical Report, National Bureau of Economic Research 2018.

**Fiszbein, Martin**, "Agricultural diversity, structural change, and long-run development: Evidence from the United States," *American Economic Journal: Macroeconomics*, 2022, *14* (2), 1–43.

**Freedman, Maurice**, "Lineage Organisation in Southeast China: Fukien and Kwangtung," *London: University of*, 1958.

_ , *Chinese Lineage and Society: Fukien and Kwangtung*, Athlone Press, 1966.

**Gao, Yanyan and Jianghuai Zheng**, "The impact of high-speed rail on innovation: An empirical test of the companion innovation hypothesis of transportation improvement with China's manufacturing firms," *World Development*, 2020, *127*, 104838.

**Greif, Avner and Guido Tabellini**, "Cultural and institutional bifurcation: China and Europe compared," *American Economic Review*, 2010, *100* (2), 135–140.

**_ , Paul Milgrom, and Barry R Weingast**, "Coordination, commitment, and enforcement: The case of the merchant guild," *Journal of Political Economy*, 1994, *102* (4), 745–776.

**Güell, Maia, José V Rodríguez Mora, and Christopher I Telmer**, "The informational content of surnames, the evolution of intergenerational mobility, and assortative mating," *The Review of Economic Studies*, 2015, *82* (2), 693–735.

**Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, "Does culture affect economic outcomes?," *Journal of Economic Perspectives*, 2006, *20* (2), 23–48.

**Habyarimana, James, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein**, "Why does ethnic diversity undermine public goods provision?," *American Political Science Review*, 2007, *101* (4), 709–725.

**Hall, Bronwyn H, Jacques Mairesse, and Pierre Mohnen**, "Measuring the Returns to R&D," in "Handbook of the Economics of Innovation," Vol. 2, Elsevier, 2010, pp. 1033–1082.

**Hanley, Douglas, Jiancheng Li, and Mingqin Wu**, "High-speed railways and collaborative innovation," *Regional Science and Urban Economics*, 2022, *93*, 103717.

**Hashmi, Aamir Rafique**, "Competition and innovation: The inverted-U relationship revisited," *Review of Economics and Statistics*, 2013, *95* (5), 1653–1668.

**Heilmann, Sebastian**, "From local experiments to national policy: the origins of China's distinctive policy process," *The China Journal*, 2008, (59), 1–30.

_ , "Policy experimentation in China's economic rise," *Studies in comparative international development*, 2008, *43* (1), 1–26.

**Helmke, Gretchen and Steven Levitsky**, *Informal institutions and comparative politics: A research agenda*, Edward Elgar Publishing, 2012.

**Hu, Albert Guangzhou and Gary H Jefferson**, "A great wall of patents: What is behind China's recent patent explosion?," *Journal of Development Economics*, 2009, *90* (1), 57–68.

**Huang, Baibing, Endong Yang, and Yang Zhang**, "Board surname sharing and investment efficiency: Evidence from Chinese state-owned enterprises," *Corporate Governance: An International Review*, 2023, *31* (4), 597–624.

**Jia, Ruixue and James Kai sing Kung**, "The Culture and Institutions of Confucianism," 2025.

**Kerr, William R and William F Lincoln**, "The supply side of innovation: H-1B visa reforms and US ethnic invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

**Köbig, Michael, Kjetil Storesletten, Zheng Song, and Fabrizio Zilibotti**, "From imitation to innovation: Where is all that Chinese R&D going?," *Econometrica*, 2022, *90* (4), 1615–1654.

**Lim, Gedeon J, Quoc-Anh Do, and Sebastian Ellingsen**, "Communities of Commerce: The Legacy of Chinese Immigration on Java, Indonesia," *Indonesia (November 15, 2024)*, 2024.

**Liu, Qing and Hong Ma**, "Trade policy uncertainty and innovation: Firm level evidence from China's WTO accession," *Journal of International Economics*, 2020, *127*, 103387.

_ **and Larry D Qiu**, "Intermediate input imports and innovations: Evidence from Chinese firms' patent filings," *Journal of International Economics*, 2016, *103*, 166–183.

_ **, Ruosi Lu, Yi Lu, and Tuan Anh Luong**, "Import competition and firm innovation: Evidence from China," *Journal of Development Economics*, 2021, *151*, 102650.

**Luo, Kaicheng, Shaoda Wang, and David Yang**, "On the Centralization of Policy Making in China," 2024.

**Mair, Johanna, Ignasi Marti, and Marc J Ventresca**, "Building inclusive markets in rural Bangladesh: How intermediaries work institutional voids," *Academy of Management Journal*, 2012, *55* (4), 819–850.

**McPherson, Miller, Lynn Smith-Lovin, and James M Cook**, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, 2001, *27* (1), 415–444.

**Minbaeva, Dana, Alena Ledeneva, Maral Muratbekova-Touron, and Sven Horak**, "Explaining the persistence of informal institutions: The role of informal networks," *Academy of Management Review*, 2023, *48* (3), 556–574.

**Nelson, Richard R and Sidney G Winter**, *An evolutionary theory of economic change*, harvard university press, 1985.

**North, Douglass C**, *The rise of the western world: A new economic history*, Cambridge University Press, 1973.

_ , "Structure and change in economic history," *W&W Norton*, 1981.

_ , "Institutions, institutional change and economic performance," *Cambridge University*, 1990.

**Ozgen, Ceren**, "The economics of diversity: Innovation, productivity and the labour market," *Journal of Economic Surveys*, 2021, *35* (4), 1168–1216.

**Peng, Mike W, Sunny Li Sun, Brian Pinkham, and Hao Chen**, "The institution-based view as a third leg for a strategy tripod," *Academy of Management Perspectives*, 2009, *23* (3), 63–81.

**Peng, Yusheng**, "Kinship networks and entrepreneurs in China's transitional economy," *American Journal of Sociology*, 2004, *109* (5), 1045–1074.

**Posch, Max, Jonathan Schulz, and Joseph Henrich**, "Surname diversity, social ties and innovation," *Social Ties and Innovation (August 3, 2023)*, 2023.

**Qian, Yingyi and Chenggang Xu**, "The M-form hierarchy and China's economic reform," *European Economic Review*, 1993, *37* (2-3), 541–548.

**_ , Gerard Roland, and Chenggang Xu**, "Coordination and experimentation in M-form and U-form organizations," *Journal of Political Economy*, 2006, *114* (2), 366–402.

**Ruofu, Du and Yuan Yida**, "Encyclopedia of Chinese Surnames," 1996.

**Schumpeter, Joseph A**, "The theory of economic development, translated by Redvers Opie," *Harvard: Economic Studies*, 1934, *46* (1600), 0404.

**Shannon, Claude Elwood**, "A mathematical theory of communication," *The Bell System Technical Journal*, 1948, *27* (3), 379–423.

**Shao, Yuchen and Chengrui Xiao**, "Corporate tax policy and heterogeneous firm innovation: Evidence from a developing country," *Journal of Comparative Economics*, 2019, *47* (2), 470–486.

**Talhelm, Thomas, Xuemin Zhang, Shigehiro Oishi, Chen Shimin, Dongyuan Duan, Xuezhao Lan, and Shinobu Kitayama**, "Large-scale psychological differences within China explained by rice versus wheat agriculture," *Science*, 2014, *344* (6184), 603–608.

**Tan, Youchao, Jason Xiao, Cheng Colin Zeng, and Hong Zou**, "What's in a name? The valuation effect of directors'sharing of surnames," *Journal of Banking & Finance*, 2021, *122*, 105991.

**Tsai, Kellee S**, "Imperfect substitutes: the local political economy of informal finance and microfinance in rural China and India," *World Development*, 2004, *32* (9), 1487–1507.

**Wang, Shaoda and David Y Yang**, "Policy experimentation in china: The political economy of policy learning," Technical Report, National Bureau of Economic Research 2021.

**Wang, Xu, Zhuan Xie, Xiaobo Zhang, and Yiping Huang**, "Roads to innovation: Firm-level evidence from People's Republic of China (PRC)," *China Economic Review*, 2018, *49*, 154–170.

**Wei, Shang-Jin, Jianhuan Xu, Ge Yin, and Xiaobo Zhang**, "Bureaucratic bean counting and patent subsidies: A welfare evaluation of China's pro-innovation program," Technical Report, Technical Report 2021. Working Paper 2021.

**_ , Zhuan Xie, and Xiaobo Zhang**, "From "made in China" to "innovated in China": Necessity, prospect, and challenges," *Journal of Economic Perspectives*, 2017, *31* (1), 49–70.

**Xie, Jian and Song Yuan**, "The cultural origins of family firms," *Journal of Comparative Economics*, 2025.

**Xu, Yiqing and Yang Yao**, "Informal institutions, collective action, and public investment in rural China," *American Political Science Review*, 2015, *109* (2), 371–391.

**Zhang, Chuanchuan**, "Clans, entrepreneurship, and development of the private sector in China," *Journal of Comparative Economics*, 2020, *48* (1), 100–123.

**Zheng, Shilin and Zhaochen Li**, "Pilot governance and the rise of China's innovation," *China Economic Review*, 2020, *63*, 101521.

# Appendices for Online Publication Only

Jian Xie, Bernard Yeung, and Qinhong Yu

August 9, 2025

## A Appendices

### A.1 Summary Statistics

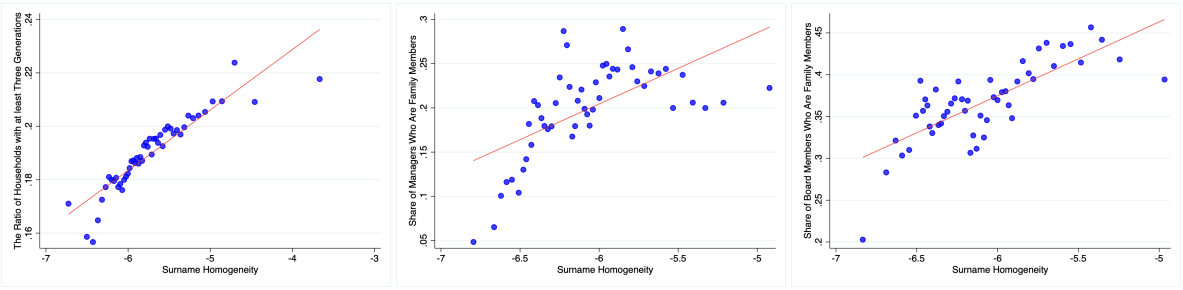
Table A.1: Summary Statistics

| | Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Patent filings per 1000 people | 54,100 | 0.077 | 0.166 | 0.000 | 0.663 |
| Surname homogeneity | 56,970 | -5.699 | 0.587 | -7.867 | 0.000 |
| Ruggedness | 56,679 | 2.484 | 2.478 | 0.017 | 17.180 |
| Distance to coast | 56,805 | 569,223.880 | 578,025.046 | 154.065 | 3648153.000 |
| Province capital | 56,805 | 0.105 | 0.306 | 0.000 | 1.000 |
| Special economic zone | 56,805 | 0.016 | 0.125 | 0.000 | 1.000 |
| High-speed railway station | 57,372 | 0.086 | 0.281 | 0.000 | 1.000 |
| Airport | 57,372 | 0.481 | 0.500 | 0.000 | 1.000 |
| Number of private firms | 57,372 | 4,517.843 | 13,626.536 | 0.000 | 519,106.000 |
| Population | 56,805 | 462,731.441 | 355,062.483 | 8,815.000 | 6445777.000 |
| Average schooling years | 56,805 | 7.555 | 1.262 | 1.849 | 11.848 |
| Urbanization rate | 56,805 | 27.383 | 24.540 | 2.260 | 98.910 |
| Fraction of minorities | 56,805 | 13.463 | 25.550 | 0.000 | 99.278 |
| Immigration rate | 56,805 | -0.000 | 0.127 | -0.567 | 0.904 |
| Marketization index | 56,007 | 9.142 | 3.492 | 1.217 | 19.874 |
| Banks density | 56,700 | 0.143 | 0.114 | 0.000 | 2.182 |
| Number of universities | 57,372 | 1.081 | 2.714 | 0.000 | 36.000 |
| Number of 211 university | 57,372 | 0.047 | 0.442 | 0.000 | 18.000 |
| Intellectual property rights protection | 51,908 | 5.442 | 8.086 | 0.040 | 53.510 |
| Financial development | 57,372 | 2.665 | 0.773 | 1.413 | 7.578 |
| Public service expenditure | 38,248 | 0.034 | 0.019 | 0.008 | 0.125 |
| Development of the factor market | 57,372 | 6.718 | 3.073 | 0.801 | 16.713 |
| Development of the product market | 57,372 | 7.905 | 1.715 | -0.247 | 10.609 |
| Policy innovation | 56,028 | 0.138 | 0.345 | 0.000 | 1.000 |
| Nighttime light | 56,406 | 920.838 | 1,198.420 | 22.038 | 4,424.641 |
| New established firms per 1000 people | 54,060 | 1.636 | 1.970 | 0.101 | 7.518 |
| Rice suitability | 56,973 | 1,449.100 | 1,889.868 | 0.000 | 7,078.907 |
| Patent filings per firm | 54,256 | 0.007 | 0.010 | 0.000 | 0.038 |
| HHI | 56,970 | 0.038 | 0.010 | 0.024 | 0.063 |
| Industry homogeneity | 56,805 | -5.052 | 0.379 | -5.550 | -4.146 |
| Genealogy density | 57,372 | 0.026 | 0.119 | 0.000 | 2.422 |
| Agricultural diversity | 56,805 | 1.521 | 0.491 | 0.000 | 2.232 |
| Treaty ports | 57,372 | 0.162 | 0.369 | 0.000 | 1.000 |
| Granted patent per 1000 people | 54,080 | 0.020 | 0.041 | 0.000 | 0.160 |
| Policy innovation per 100,000 people | 55,566 | 0.027 | 0.072 | 0.000 | 0.263 |

## A.2 Validation Test

We posit that surname homogeneity among business owners within a county could indicate the strength of kinship-based informal institutions, as discussed in Section 2.2. To validate this measure, we use three other county-level variables that potentially reflect the strength of kinship-based informal institutions. The first variable is the ratio of extended families, measured as the number of households with at least three generations divided by the total number of households. The prevalence of extended families, where multiple generations co-reside, typically indicates strong kinship ties, contrasting with nuclear families, which emphasize smaller and more independent household units. The second variable is the share of managers who are family members, and the third variable is the share of board members who are family members. These two variables measure family members' participation in the firm and have been shown to be strongly correlated with local kinship culture (Xie and Yuan, 2025).

Figure A.1 illustrates a strong positive relationship between surname homogeneity among business owners and all these three variables. This suggests that regions with higher surname homogeneity tend to have stronger kinship-based informal institutions.




(a) Extended Families (b) Family Managers (c) Family Board Members

Figure A.1: Validation Test for the Measure of Surname Homogeneity

Notes: Data on the number of households with at least three generations are from the 2000 Population Census. Data on the share of managers who are family members and the share of board members who are family members come from the China Private Enterprise Survey 2010.

## A.3 Figures

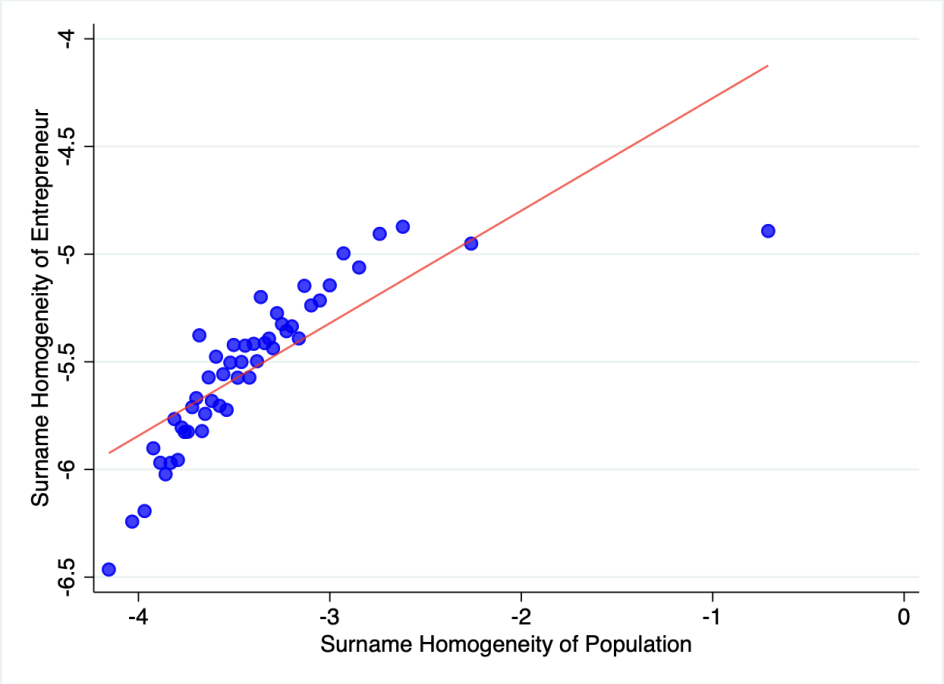
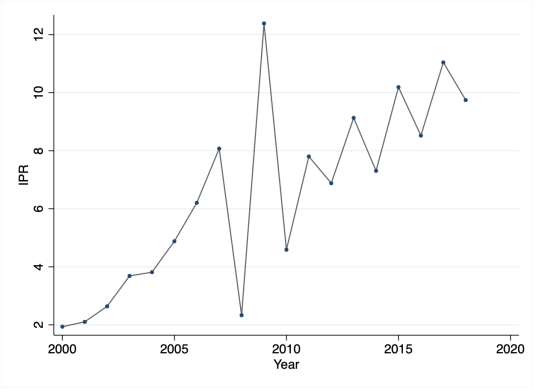
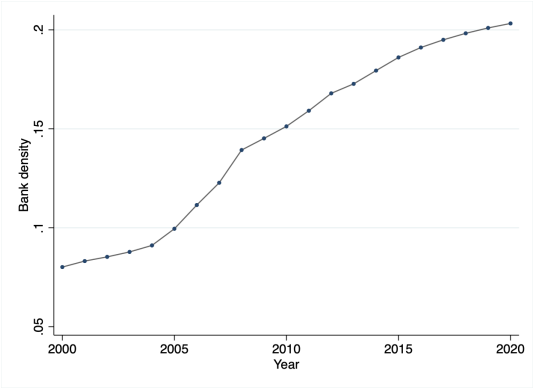



Figure A.2: Surname Homogeneity of the Population and surname homogeneity among business owners in 2005

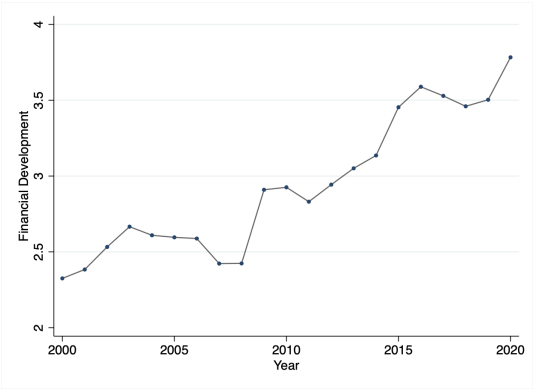
Notes: This figure shows the relationship between surname homogeneity of the population and surname homogeneity among business owners in 2005. Data on surname homogeneity of the population is from the 2005 mini-census.



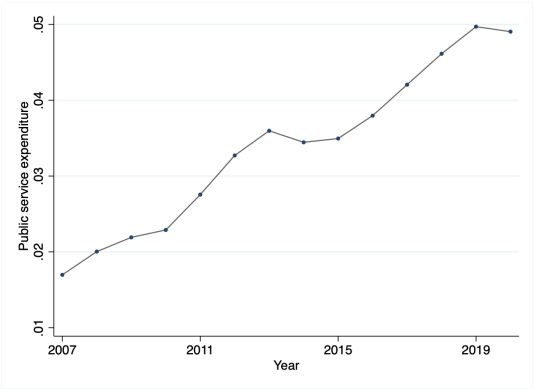

(a) Intellectual Property Rights Protection



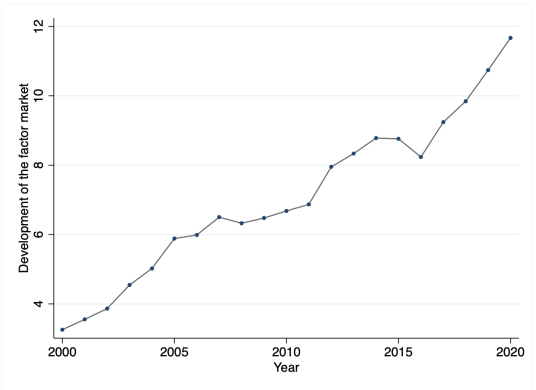

(b) Bank Density



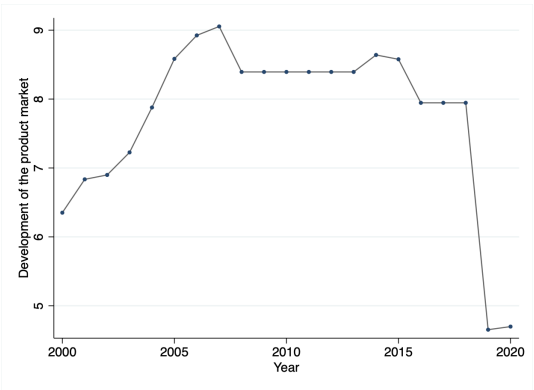

(c) Financial Development




(d) Public Service Expenditure




(e) Factor Market




(f) Product Market

Note: These six figures show the changes in seven measures of the strength of formal institutions over time in China, including the intellectual property rights protection index, bank density, financial development, public service expenditure, the index of factor market development, and the index of product market development. We compute their average values for each year.

## A.4 Coefficients of Controls

Table A.2: Coefficients of Controls

| Dependent variable | Patent filings per 1000 people | Surname Homogeneity |
|---|---|---|
| | (1) | (2) |
| Surname homogeneity | 0.028*** | |
| | (0.006) | |
| Firm number | 0.000*** | -0.000** |
| | (0.000) | (0.000) |
| High speed train station | 0.043*** | 0.030*** |
| | (0.009) | (0.010) |
| Airport | -0.029*** | -0.061*** |
| | (0.006) | (0.018) |
| University | 0.028*** | 0.022** |
| | (0.005) | (0.009) |
| Bank density | 0.527*** | 0.033 |
| | (0.049) | (0.111) |
| Market index | -0.006*** | -0.002 |
| | (0.002) | (0.004) |
| Other controls | Yes | Yes |
| County FE | Yes | Yes |
| Year FE | Yes | Yes |
| Observations | 52,495 | 52,495 |
| $R^2$ | 0.719 | 0.897 |
| Mean dependent vars. | 0.078 | -5.731 |

Notes: County-level OLS regressions. The outcome variable in column (1) is the number of patent filings by private firms in a county, normalized by 1000 local population. The outcome variable in column (2) is surname homogeneity among business owners. Time-variant controls include the number of private firms, a high-speed railway station dummy, an airport dummy, the number of universities, bank density, and the marketization index. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

## A.5 The Correlation over Time

Table A.3: The Correlation over Time

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Year 2000 | 0.093*** | 0.051*** | 0.040*** | 0.035*** | 0.033*** |
| | (0.011) | (0.008) | (0.007) | (0.006) | (0.006) |
| Surname homogeneity × Year 2001 | 0.101*** | 0.053*** | 0.042*** | 0.036*** | 0.034*** |
| | (0.012) | (0.009) | (0.008) | (0.007) | (0.006) |
| Surname homogeneity × Year 2002 | 0.107*** | 0.055*** | 0.043*** | 0.037*** | 0.036*** |
| | (0.013) | (0.009) | (0.008) | (0.007) | (0.007) |
| Surname homogeneity × Year 2003 | 0.119*** | 0.060*** | 0.047*** | 0.041*** | 0.040*** |
| | (0.015) | (0.010) | (0.009) | (0.008) | (0.008) |
| Surname homogeneity × Year 2004 | 0.126*** | 0.061*** | 0.048*** | 0.041*** | 0.041*** |
| | (0.016) | (0.011) | (0.010) | (0.009) | (0.008) |
| Surname homogeneity × Year 2005 | 0.130*** | 0.061*** | 0.048*** | 0.042*** | 0.045*** |
| | (0.017) | (0.012) | (0.010) | (0.009) | (0.009) |
| Surname homogeneity × Year 2006 | 0.126*** | 0.054*** | 0.042*** | 0.037*** | 0.041*** |
| | (0.018) | (0.013) | (0.011) | (0.009) | (0.009) |
| Surname homogeneity × Year 2007 | 0.114*** | 0.041*** | 0.029** | 0.027*** | 0.029*** |
| | (0.019) | (0.013) | (0.011) | (0.009) | (0.009) |
| Surname homogeneity × Year 2008 | 0.096*** | 0.024* | 0.015 | 0.017* | 0.021** |
| | (0.019) | (0.013) | (0.011) | (0.009) | (0.009) |
| Surname homogeneity × Year 2009 | 0.078*** | 0.005 | -0.002 | 0.004 | 0.009 |
| | (0.020) | (0.014) | (0.011) | (0.010) | (0.010) |
| Surname homogeneity × Year 2010 | 0.045** | -0.026* | -0.030** | -0.019* | -0.012 |
| | (0.021) | (0.015) | (0.012) | (0.011) | (0.011) |
| Surname homogeneity × Year 2011 | 0.011 | -0.058*** | -0.059*** | -0.044*** | -0.035*** |
| | (0.021) | (0.016) | (0.013) | (0.012) | (0.012) |
| Surname homogeneity × Year 2012 | -0.018 | -0.084*** | -0.083*** | -0.067*** | -0.057*** |
| | (0.022) | (0.017) | (0.014) | (0.013) | (0.013) |
| Surname homogeneity × Year 2013 | -0.040* | -0.105*** | -0.101*** | -0.088*** | -0.076*** |
| | (0.022) | (0.018) | (0.015) | (0.015) | (0.014) |
| Surname homogeneity × Year 2014 | -0.059** | -0.128*** | -0.117*** | -0.100*** | -0.088*** |
| | (0.023) | (0.020) | (0.016) | (0.016) | (0.016) |
| Surname homogeneity × Year 2015 | -0.068*** | -0.139*** | -0.123*** | -0.111*** | -0.098*** |
| | (0.023) | (0.022) | (0.018) | (0.018) | (0.018) |
| Surname homogeneity × Year 2016 | -0.073*** | -0.151*** | -0.127*** | -0.120*** | -0.106*** |
| | (0.023) | (0.023) | (0.018) | (0.019) | (0.018) |
| Surname homogeneity × Year 2017 | -0.069*** | -0.152*** | -0.121*** | -0.116*** | -0.101*** |
| | (0.023) | (0.022) | (0.017) | (0.017) | (0.017) |
| Surname homogeneity × Year 2018 | -0.067*** | -0.144*** | -0.108*** | -0.096*** | -0.081*** |
| | (0.022) | (0.023) | (0.018) | (0.017) | (0.016) |
| Surname homogeneity × Year 2019 | -0.067*** | -0.150*** | -0.107*** | -0.105*** | -0.088*** |
| | (0.022) | (0.020) | (0.015) | (0.015) | (0.015) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.667 | 0.697 | 0.710 | 0.720 | 0.732 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The explanatory variables are the interactions between surname homogeneity among business owners and year dummies. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

## A.6 Tables of Robustness Checks: Corporate Innovation

Table A.4: Baseline Results: Normalized by Number of Private Firms

| Dependent variable | Patent filings per firm | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.004*** | 0.003*** | 0.002*** | 0.002*** | 0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 54,256 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.573 | 0.582 | 0.586 | 0.591 | 0.597 |
| Mean dependent vars. | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by number of private firms. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.5: Baseline Results: Number of Patent Filings (Unnormalized)

| Dependent variable | Patent filings | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 55.702*** | 30.976*** | 22.830*** | 12.868*** | 11.645*** |
| | (5.828) | (3.431) | (2.862) | (2.968) | (2.795) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 54,256 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.637 | 0.671 | 0.698 | 0.715 | 0.731 |
| Mean dependent vars. | 36.754 | 36.763 | 36.763 | 36.763 | 37.011 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.6: Baseline Results: Winsorized at 1% and 99%

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.204*** | 0.093*** | 0.041*** | 0.045*** | 0.043*** |
| | (0.028) | (0.013) | (0.010) | (0.013) | (0.012) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.537 | 0.570 | 0.638 | 0.659 | 0.679 |
| Mean dependent vars. | 0.123 | 0.123 | 0.123 | 0.123 | 0.125 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.7: Baseline Results: Newey-West Standard Errors

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.106*** | 0.057*** | 0.041*** | 0.029*** | 0.028*** |
| | (0.003) | (0.003) | (0.002) | (0.003) | (0.002) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.627 | 0.657 | 0.685 | 0.703 | 0.719 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. Newey-West standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.8: Restricting the Sample to the Years 2000, 2003, 2006, 2009, 2012, 2015, and 2018

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.097*** | 0.054*** | 0.038*** | 0.028** | 0.026** |
| | (0.018) | (0.009) | (0.007) | (0.008) | (0.008) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 18,806 | 18,764 | 18,764 | 18,764 | 18,372 |
| $R^2$ | 0.617 | 0.650 | 0.683 | 0.704 | 0.721 |
| Mean dependent vars. | 0.072 | 0.072 | 0.072 | 0.072 | 0.072 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. We restrict the Sample to the Years 2000, 2003, 2006, 2009, 2012, 2015, and 2018. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.9: Restricting the Sample to the Years 2000, 2004, 2008, 2012, and 2016

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.092** | 0.055** | 0.038** | 0.033** | 0.030** |
| | (0.024) | (0.013) | (0.010) | (0.010) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 13,432 | 13,402 | 13,402 | 13,402 | 13,122 |
| $R^2$ | 0.606 | 0.637 | 0.673 | 0.690 | 0.708 |
| Mean dependent vars. | 0.067 | 0.067 | 0.067 | 0.067 | 0.067 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. We restrict the sample to the years 2000, 2004, 2008, 2012, and 2016. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.10: The Correlation Across Different Periods: 2001-2008

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.017*** | 0.010*** | 0.002 | 0.003* | 0.003* |
| | (0.004) | (0.003) | (0.002) | (0.002) | (0.001) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 21,491 | 21,443 | 21,443 | 21,443 | 20,995 |
| $R^2$ | 0.667 | 0.677 | 0.735 | 0.744 | 0.748 |
| Mean dependent vars. | 0.009 | 0.009 | 0.009 | 0.009 | 0.010 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2008 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.11: Baseline Results: T+1 to T+2

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.211*** | 0.117*** | 0.085*** | 0.063*** | 0.059*** |
| | (0.024) | (0.013) | (0.011) | (0.012) | (0.011) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 51,048 | 50,934 | 50,934 | 50,934 | 49,870 |
| $R^2$ | 0.653 | 0.682 | 0.710 | 0.727 | 0.743 |
| Mean dependent vars. | 0.153 | 0.153 | 0.153 | 0.153 | 0.154 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of patent filings by private firms in a county in the following two years, normalized by 1000 local population Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.12: Baseline Results: T+1 to T+3

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.316*** | 0.179*** | 0.132*** | 0.103*** | 0.096*** |
| | (0.014) | (0.020) | (0.017) | (0.017) | (0.016) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 48,361 | 48,253 | 48,253 | 48,253 | 47,245 |
| $R^2$ | 0.673 | 0.701 | 0.729 | 0.745 | 0.760 |
| Mean dependent vars. | 0.228 | 0.227 | 0.227 | 0.227 | 0.229 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of patent filings by private firms in a county in the following theree years, normalized by 1000 local population Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.13: Baseline Results: T+1 to T+4

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.419*** | 0.243*** | 0.181*** | 0.148*** | 0.137*** |
| | (0.049) | (0.019) | (0.024) | (0.022) | (0.022) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 45,674 | 45,572 | 45,572 | 45,572 | 44,620 |
| $R^2$ | 0.692 | 0.719 | 0.747 | 0.762 | 0.776 |
| Mean dependent vars. | 0.299 | 0.299 | 0.299 | 0.299 | 0.301 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of patent filings by private firms in a county in the following four years, normalized by 1000 local population Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.14: Baseline Results: T+1 to T+5

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.517*** | 0.307*** | 0.227*** | 0.194*** | 0.177*** |
| | (0.062) | (0.036) | (0.032) | (0.028) | (0.027) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 42,987 | 42,891 | 42,891 | 42,891 | 41,995 |
| $R^2$ | 0.711 | 0.737 | 0.766 | 0.780 | 0.793 |
| Mean dependent vars. | 0.368 | 0.367 | 0.367 | 0.367 | 0.370 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of patent filings by private firms in a county in the following five years, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.15: Excluding Cities with Minority Rates Above 10%

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.154*** | 0.110*** | 0.083*** | 0.061*** | 0.046*** |
| | (0.008) | (0.014) | (0.011) | (0.009) | (0.007) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 40,700 | 40,580 | 40,580 | 40,580 | 39,860 |
| $R^2$ | 0.643 | 0.679 | 0.700 | 0.723 | 0.740 |
| Mean dependent vars. | 0.092 | 0.091 | 0.091 | 0.091 | 0.092 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of patent filings by private firms in a county, normalized by 1000 local population Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.16: Excluding Municipalities Directly under the Central Government

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.099*** | 0.049*** | 0.036*** | 0.027*** | 0.027*** |
| | (0.012) | (0.006) | (0.005) | (0.006) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,015 | 51,915 | 51,915 | 51,915 | 51,275 |
| $R^2$ | 0.610 | 0.645 | 0.672 | 0.689 | 0.704 |
| Mean dependent vars. | 0.072 | 0.072 | 0.072 | 0.072 | 0.073 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.17: Excluding Special Economic Zones

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.105*** | 0.056*** | 0.039*** | 0.030*** | 0.028*** |
| | (0.012) | (0.006) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,875 | 52,755 | 52,755 | 52,755 | 51,935 |
| $R^2$ | 0.622 | 0.652 | 0.684 | 0.701 | 0.716 |
| Mean dependent vars. | 0.076 | 0.076 | 0.076 | 0.076 | 0.076 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.18: Another Measure of Surname Homogeneity: HHI

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| HHI | 4.293*** | 1.928*** | 1.325*** | 0.747*** | 0.783*** |
| | (0.448) | (0.207) | (0.260) | (0.260) | (0.218) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.615 | 0.654 | 0.684 | 0.702 | 0.718 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners, measured based on the Herfindahl-Hirschman Index (HHI). Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.19: Control for Economic Development

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.080*** | 0.046*** | 0.033*** | 0.024*** | 0.023*** |
| | (0.013) | (0.007) | (0.005) | (0.006) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Nighttime light | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,115 | 52,995 | 52,995 | 52,995 | 52,355 |
| $R^2$ | 0.656 | 0.680 | 0.702 | 0.715 | 0.729 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.20: Control for Industry Homogeneity

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.048*** | 0.013*** | 0.010** | 0.010* | 0.013** |
| | (0.006) | (0.004) | (0.004) | (0.005) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Industry homogeneity | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,555 | 53,435 | 53,435 | 53,435 | 52,315 |
| $R^2$ | 0.655 | 0.676 | 0.695 | 0.708 | 0.722 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.21: Control for Industry Structure

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.104*** | 0.057*** | 0.041*** | 0.028*** | 0.027*** |
| | (0.011) | (0.006) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Industry structure | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,533 | 53,413 | 53,413 | 53,413 | 52,293 |
| $R^2$ | 0.631 | 0.658 | 0.686 | 0.703 | 0.719 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. We control the fraction of firm size: (1) smaller than 100,000; (2) between 100,000 to 1,000,000; (3) between 1,000,000 to 5,000,000; (4) between 5,000,000 to 10,000,000. The size is measured by the registered capital in CNY. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.22: Control for Clan Culture

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.105*** | 0.056*** | 0.040*** | 0.029*** | 0.028*** |
| | (0.012) | (0.006) | (0.005) | (0.006) | (0.005) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Clan culture | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.631 | 0.661 | 0.688 | 0.707 | 0.722 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.23: Control for Agricultural Diversity

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.099*** | 0.052*** | 0.037*** | 0.022*** | 0.019*** |
| | (0.010) | (0.006) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Agricultural diversity | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.631 | 0.660 | 0.687 | 0.706 | 0.724 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.24: Control for Treaty Port Dummy

| Dependent variable | Patent filings per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.099*** | 0.053*** | 0.039*** | 0.029*** | 0.028*** |
| | (0.011) | (0.006) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Treaty port dummy | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.641 | 0.666 | 0.690 | 0.706 | 0.721 |
| Mean dependent vars. | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |

Notes: County-level OLS regressions. The outcome variable is the number of patent filings by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.25: Number of Granted Patents as the Outcome Variable

| Dependent variable | Granted Patents per 1000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.022*** | 0.012*** | 0.010*** | 0.007*** | 0.006*** |
| | (0.003) | (0.002) | (0.002) | (0.002) | (0.001) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,735 | 53,615 | 53,615 | 53,615 | 52,495 |
| $R^2$ | 0.663 | 0.689 | 0.701 | 0.726 | 0.739 |
| Mean dependent vars. | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |

Notes: County-level OLS regressions. The outcome variable is the number of granted patents by private firms in a county, normalized by 1000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

## A.7 Tables of Robustness Checks: Policy Innovation

Table A.26: Number of Policy Innovations Normalized by Local Population

| Dependent variable | Number of policy innovations per 100,000 people | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.025*** | 0.015*** | 0.012*** | 0.014*** | 0.013*** |
| | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,919 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.232 | 0.243 | 0.245 | 0.252 | 0.256 |
| Mean dependent vars. | 0.036 | 0.036 | 0.036 | 0.036 | 0.036 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations in a county, normalized by 100,000 local population. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. rsities. We use two-way clustering of standard errors at the county and year levels except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.27: Policy Innovation: Winsorized at 1% and 99%

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.182*** | 0.108*** | 0.080*** | 0.061*** | 0.057*** |
| | (0.017) | (0.018) | (0.016) | (0.018) | (0.017) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.299 | 0.313 | 0.320 | 0.327 | 0.333 |
| Mean dependent vars. | 0.195 | 0.196 | 0.196 | 0.196 | 0.198 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.28: Policy Innovation: Newey-West Standard Errors

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.107*** | 0.066*** | 0.056*** | 0.050*** | 0.047*** |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.006) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.242 | 0.253 | 0.255 | 0.261 | 0.265 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. Newey-West standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.29: Policy Innovation: Long Difference Approach

| Dependent variable | Change in number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Change in Surname homogeneity | 0.105*** | 0.054*** | 0.033*** | 0.033*** | 0.032*** |
| | (0.011) | (0.011) | (0.011) | (0.011) | (0.011) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Observations | 2,639 | 2,639 | 2,639 | 2,639 | 2,587 |
| $R^2$ | 0.037 | 0.104 | 0.128 | 0.132 | 0.141 |
| Mean dependent vars. | 0.160 | 0.160 | 0.160 | 0.160 | 0.161 |

Notes: County-level OLS regressions. The outcome variable is the change in number of policy innovations over the period from 2001 to 2020. The explanatory variable is the change in surname homogeneity among business owners over the period from 2000 to 2019. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. Robust standard errors are in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Table A.30: Restricting the Sample to the Years 2000, 2003, 2006, 2009, 2012, 2015, and 2018

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.090*** | 0.057*** | 0.044*** | 0.035*** | 0.032*** |
| | (0.007) | (0.010) | (0.010) | (0.006) | (0.008) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 18,674 | 18,479 | 18,479 | 18,479 | 18,115 |
| $R^2$ | 0.288 | 0.296 | 0.300 | 0.305 | 0.308 |
| Mean dependent vars. | 0.146 | 0.147 | 0.147 | 0.147 | 0.148 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. We restrict the Sample to the Years 2000, 2003, 2006, 2009, 2012, 2015, and 2018. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levelsall columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.31: Restricting the Sample to the Years 2000, 2004, 2008, 2012, and 2016

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.117*** | 0.079*** | 0.070*** | 0.056*** | 0.053*** |
| | (0.015) | (0.009) | (0.011) | (0.010) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 13,338 | 13,199 | 13,199 | 13,199 | 12,939 |
| $R^2$ | 0.362 | 0.372 | 0.373 | 0.379 | 0.384 |
| Mean dependent vars. | 0.146 | 0.147 | 0.147 | 0.147 | 0.148 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. We restrict the sample to the years 2000, 2004, 2008, 2012, and 2016. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails.. * p<0.1; ** p<0.05; *** p<0.01.

Table A.32: Policy Innovation: T+1 to T+2

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.212*** | 0.134*** | 0.115*** | 0.103*** | 0.097*** |
| | (0.016) | (0.020) | (0.020) | (0.019) | (0.019) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 50,689 | 50,159 | 50,159 | 50,159 | 49,171 |
| $R^2$ | 0.384 | 0.395 | 0.398 | 0.404 | 0.408 |
| Mean dependent vars. | 0.295 | 0.297 | 0.297 | 0.297 | 0.299 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of policy innovations implemented at the county level or below in the following two years. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns. * p<0.1; ** p<0.05; *** p<0.01.

Table A.33: Policy Innovation: T+1 to T+3

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.310*** | 0.195*** | 0.167*** | 0.151*** | 0.142*** |
| | (0.018) | (0.019) | (0.027) | (0.026) | (0.026) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 48,021 | 47,519 | 47,519 | 47,519 | 46,583 |
| $R^2$ | 0.484 | 0.495 | 0.499 | 0.504 | 0.509 |
| Mean dependent vars. | 0.453 | 0.455 | 0.455 | 0.455 | 0.458 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of policy innovations implemented at the county level or below in the following three years. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1) and (2). In column (1) and (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.34: Policy Innovation: T+1 to T+4

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.389*** | 0.245*** | 0.208*** | 0.190*** | 0.179*** |
| | (0.024) | (0.025) | (0.031) | (0.033) | (0.032) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 45,353 | 44,879 | 44,879 | 44,879 | 43,995 |
| $R^2$ | 0.561 | 0.571 | 0.575 | 0.579 | 0.584 |
| Mean dependent vars. | 0.615 | 0.618 | 0.618 | 0.618 | 0.623 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of policy innovations implemented at the county level or below in the following four years. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.35: Policy Innovation: T+1 to T+5

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.451*** | 0.280*** | 0.232*** | 0.219*** | 0.207*** |
| | (0.029) | (0.038) | (0.036) | (0.040) | (0.038) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 42,685 | 42,239 | 42,239 | 42,239 | 41,407 |
| $R^2$ | 0.626 | 0.634 | 0.639 | 0.642 | 0.647 |
| Mean dependent vars. | 0.778 | 0.782 | 0.782 | 0.782 | 0.788 |

Notes: County-level OLS regressions. The outcome variable is the cumulative number of policy innovations implemented at the county level or below in the following five years. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.36: Policy Innovation: Excluding Cities with Minority Rates Above 10%

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.147*** | 0.109*** | 0.094*** | 0.076*** | 0.064*** |
| | (0.015) | (0.014) | (0.013) | (0.015) | (0.014) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 40,560 | 40,440 | 40,440 | 40,440 | 39,800 |
| $R^2$ | 0.248 | 0.262 | 0.264 | 0.272 | 0.276 |
| Mean dependent vars. | 0.163 | 0.163 | 0.163 | 0.163 | 0.164 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns. * p<0.1; ** p<0.05; *** p<0.01.

Table A.37: Policy Innovation: Excluding Municipalities Directly under the Central Government

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.101*** | 0.053*** | 0.047*** | 0.045*** | 0.043*** |
| | (0.009) | (0.006) | (0.011) | (0.011) | (0.011) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 51,199 | 51,099 | 51,099 | 51,099 | 50,539 |
| $R^2$ | 0.219 | 0.232 | 0.233 | 0.240 | 0.242 |
| Mean dependent vars. | 0.136 | 0.136 | 0.136 | 0.136 | 0.137 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.38: Policy Innovation: Excluding Special Economic Zones

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.109*** | 0.067*** | 0.055*** | 0.051*** | 0.047*** |
| | (0.010) | (0.006) | (0.010) | (0.011) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,079 | 51,959 | 51,959 | 51,959 | 51,219 |
| $R^2$ | 0.238 | 0.248 | 0.251 | 0.257 | 0.260 |
| Mean dependent vars. | 0.144 | 0.144 | 0.144 | 0.144 | 0.144 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.39: Another Measure of Surname Homogeneity: HHI

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| HHI | 4.074*** | 2.212*** | 1.854*** | 1.355*** | 1.325*** |
| | (0.292) | (0.466) | (0.416) | (0.437) | (0.428) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.240 | 0.252 | 0.255 | 0.261 | 0.264 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners, measured based on the Herfindahl-Hirschman Index (HHI). Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels in all columns except column (1). In column (1), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.40: Policy Innovation: Control for Economic Development

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.099*** | 0.061*** | 0.052*** | 0.049*** | 0.044*** |
| | (0.009) | (0.011) | (0.010) | (0.010) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Nighttime light | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,619 | 52,259 | 52,259 | 52,259 | 51,639 |
| $R^2$ | 0.243 | 0.254 | 0.256 | 0.262 | 0.265 |
| Mean dependent vars. | 0.146 | 0.146 | 0.146 | 0.146 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.41: Policy Innovation: Control for Industry Homogeneity

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.069*** | 0.037*** | 0.036*** | 0.033*** | 0.036*** |
| | (0.009) | (0.010) | (0.010) | (0.009) | (0.009) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Industry homogeneity | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,177 | 52,619 | 52,619 | 52,619 | 51,579 |
| $R^2$ | 0.244 | 0.254 | 0.256 | 0.262 | 0.265 |
| Mean dependent vars. | 0.145 | 0.146 | 0.146 | 0.146 | 0.147 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.42: Policy Innovation: Control for Industry Structure

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.105*** | 0.064*** | 0.054*** | 0.047*** | 0.043*** |
| | (0.010) | (0.011) | (0.011) | (0.011) | (0.011) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Industry structure | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,160 | 52,602 | 52,602 | 52,602 | 51,562 |
| $R^2$ | 0.242 | 0.253 | 0.255 | 0.261 | 0.265 |
| Mean dependent vars. | 0.145 | 0.146 | 0.146 | 0.146 | 0.147 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. We control the fraction of firm size: (1) smaller than 100,000; (2) between 100,000 to 1,000,000; (3) between 1,000,000 to 5,000,000; (4) between 5,000,000 to 10,000,000. The size is measured by the registered capital in CNY. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.43: Policy Innovation: Control for Clan Culture

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.105*** | 0.065*** | 0.055*** | 0.049*** | 0.046*** |
| | (0.010) | (0.011) | (0.010) | (0.011) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Clan culture | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.245 | 0.255 | 0.257 | 0.263 | 0.267 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.44: Policy Innovation: Control for Agricultural Diversity

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.100*** | 0.061*** | 0.052*** | 0.045*** | 0.042*** |
| | (0.009) | (0.006) | (0.010) | (0.010) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Agricultural diversity | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,919 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.244 | 0.255 | 0.257 | 0.263 | 0.267 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels except column (2). In column (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

Table A.45: Policy Innovation: Control for Treaty Port Dummy

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.101*** | 0.063*** | 0.055*** | 0.050*** | 0.048*** |
| | (0.006) | (0.006) | (0.011) | (0.010) | (0.010) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| Treaty port dummy | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.245 | 0.255 | 0.257 | 0.263 | 0.266 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels except column (1) and (2). In column (1) and (2), we report standard errors clustered only at the county level because two-way clustering fails. * p<0.1; ** p<0.05; *** p<0.01.

## A.8 Policy Innovation: IV Results

Table A.46: IV: Interaction Between Rice Suitability and Year

| Second-stage dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity | 0.569*** | 0.891*** | 0.998*** | 0.997*** | 1.031*** |
| | (0.108) | (0.206) | (0.257) | (0.245) | (0.254) |
| First-stage dependent variable | Surname homogeneity | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| ln(Rice suitability) $\times$Year | 0.002*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,177 | 52,619 | 52,619 | 52,619 | 51,579 |
| $R^2$ | 0.844 | 0.877 | 0.879 | 0.901 | 0.902 |
| KP F-statistics | 76.372 | 47.882 | 36.505 | 46.572 | 45.254 |

Notes: County-level 2SLS regressions. The outcome variable in the second stage is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable. The explanatory variable in the second stage is the surname homogeneity among business owners. IV is the interaction term between rice suitability and year. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year level. * p<0.1; ** p<0.05; *** p<0.01.

## A.9 Policy Innovation: Heterogeneity of Institutional Development Analyses

Table A.47: Heterogeneity in Policy Innovation: Intellectual Property Rights Protection

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × IPR | -0.008*** | -0.007*** | -0.006*** | -0.006*** | -0.006*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Surname homogeneity | 0.099*** | 0.069*** | 0.061*** | 0.058*** | 0.056*** |
| | (0.011) | (0.010) | (0.009) | (0.009) | (0.009) |
| IPR | -0.045*** | -0.039*** | -0.032*** | -0.032*** | -0.032*** |
| | (0.011) | (0.009) | (0.009) | (0.009) | (0.009) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 50,689 | 50,159 | 50,159 | 50,159 | 49,171 |
| $R^2$ | 0.249 | 0.258 | 0.259 | 0.265 | 0.269 |
| Mean dependent vars. | 0.143 | 0.144 | 0.144 | 0.144 | 0.145 |

Notes: County-level OLS regressions. "IPR" indicates the intellectual property rights protection index. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2019 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.48: Heterogeneity in Policy Innovation: Bank Density

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Bank density | -0.207*** | -0.199*** | -0.136** | -0.201** | -0.195** |
| | (0.067) | (0.063) | (0.059) | (0.078) | (0.079) |
| Surname homogeneity | 0.113*** | 0.073*** | 0.062*** | 0.062*** | 0.062*** |
| | (0.011) | (0.012) | (0.012) | (0.014) | (0.014) |
| Bank density | -0.782* | -0.814* | -0.491 | -0.787* | -0.758 |
| | (0.420) | (0.390) | (0.366) | (0.450) | (0.453) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 52,899 | 52,779 | 52,779 | 52,779 | 51,759 |
| $R^2$ | 0.245 | 0.255 | 0.257 | 0.263 | 0.265 |
| Mean dependent vars. | 0.146 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. "Bank density" is the number of banks normalized by the local population. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.49: Heterogeneity in Policy Innovation: Financial Development

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Financial | -0.054*** | -0.064*** | -0.049*** | -0.048*** | -0.043*** |
| | (0.010) | (0.010) | (0.010) | (0.011) | (0.011) |
| Surname homogeneity | 0.226*** | 0.202*** | 0.162*** | 0.155*** | 0.141*** |
| | (0.027) | (0.028) | (0.027) | (0.031) | (0.030) |
| Financial | -0.286*** | -0.349*** | -0.261*** | -0.259*** | -0.232*** |
| | (0.055) | (0.060) | (0.060) | (0.065) | (0.064) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.244 | 0.255 | 0.256 | 0.262 | 0.266 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. "Financial" indicates financial development, measured by the ratio of total loans and deposits from formal financial institutions to GDP. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.50: Heterogeneity in Policy Innovation: Public Service Expenditure

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Expenditure | -0.783 | -0.818 | -0.707 | -0.310 | -0.023 |
| | (0.832) | (1.095) | (1.057) | (1.053) | (1.017) |
| Surname homogeneity | 0.089 | 0.073 | 0.064 | 0.048 | 0.036 |
| | (0.051) | (0.057) | (0.055) | (0.055) | (0.052) |
| Expenditure | -5.486 | -5.511 | -4.872 | -2.599 | -0.837 |
| | (4.588) | (6.006) | (5.782) | (5.773) | (5.505) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 34,684 | 34,320 | 34,320 | 34,320 | 33,644 |
| $R^2$ | 0.271 | 0.277 | 0.277 | 0.282 | 0.284 |
| Mean dependent vars. | 0.194 | 0.195 | 0.195 | 0.195 | 0.196 |

Notes: County-level OLS regressions. "Expenditure" indicates the annual public service expenditure normalized by GDP in 2010. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2008 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.51: Heterogeneity in Policy Innovation: Development of the Factor Market

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Factor market | -0.020*** | -0.018*** | -0.014*** | -0.013*** | -0.012*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Surname homogeneity | 0.158*** | 0.114*** | 0.098*** | 0.090*** | 0.084*** |
| | (0.015) | (0.015) | (0.015) | (0.016) | (0.016) |
| Factor market | -0.122*** | -0.111*** | -0.092*** | -0.084*** | -0.074*** |
| | (0.018) | (0.021) | (0.019) | (0.020) | (0.021) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.246 | 0.256 | 0.257 | 0.263 | 0.266 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. "Factor market" indicates the index of development of the factor market. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

Table A.52: Heterogeneity in Policy Innovation: Development of the Product Market

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity × Product market | -0.005 | -0.000 | -0.002 | 0.000 | -0.001 |
| | (0.003) | (0.004) | (0.004) | (0.004) | (0.005) |
| Surname homogeneity | 0.136*** | 0.067** | 0.068** | 0.049* | 0.053* |
| | (0.021) | (0.030) | (0.027) | (0.028) | (0.029) |
| Product market | -0.030* | -0.001 | -0.011 | 0.001 | -0.004 |
| | (0.017) | (0.024) | (0.024) | (0.026) | (0.028) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.242 | 0.253 | 0.255 | 0.261 | 0.265 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. "Product market" indicates the index of development of the product market. The outcome variable is the number of policy innovations implemented at the county level or below. Time spans from 2001 to 2020 for the outcome variable Time spans from 2001 to 2020 for the outcome variable. The explanatory variable is the surname homogeneity among business owners. Geographic controls include average ruggedness, distance to the coast, a provincial capital dummy, and a special economic zone dummy. Transportation controls and firm number consist of a high-speed railway station dummy, an airport dummy, and the number of private firms. Demographic controls include population size, average years of schooling, urbanization rate, fraction of minorities, and migration rate. Socioeconomic controls include the marketization index, bank density, number of universities, and number of elite universities. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.

## A.10 Policy Innovation: The Correlation over Time

Table A.53: The Correlation over Time

| Dependent variable | Number of policy innovations | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surname homogeneity× Year 2000 | 0.100*** | 0.061*** | 0.054*** | 0.051*** | 0.048*** |
| | (0.007) | (0.008) | (0.008) | (0.009) | (0.009) |
| Surname homogeneity× Year 2001 | 0.110*** | 0.068*** | 0.060*** | 0.061*** | 0.058*** |
| | (0.008) | (0.008) | (0.009) | (0.010) | (0.010) |
| Surname homogeneity× Year 2002 | 0.125*** | 0.080*** | 0.072*** | 0.064*** | 0.061*** |
| | (0.009) | (0.009) | (0.010) | (0.011) | (0.011) |
| Surname homogeneity× Year 2003 | 0.102*** | 0.056*** | 0.048*** | 0.053*** | 0.051*** |
| | (0.010) | (0.011) | (0.012) | (0.013) | (0.013) |
| Surname homogeneity× Year 2004 | 0.098*** | 0.030** | 0.022 | 0.042*** | 0.042*** |
| | (0.011) | (0.012) | (0.013) | (0.014) | (0.014) |
| Surname homogeneity× Year 2005 | 0.053*** | 0.002 | -0.007 | 0.015 | 0.010 |
| | (0.012) | (0.013) | (0.013) | (0.015) | (0.015) |
| Surname homogeneity× Year 2006 | 0.119*** | 0.098*** | 0.090*** | 0.104*** | 0.103*** |
| | (0.013) | (0.013) | (0.014) | (0.015) | (0.016) |
| Surname homogeneity× Year 2007 | 0.047*** | 0.017 | 0.010 | 0.009 | 0.005 |
| | (0.013) | (0.014) | (0.014) | (0.016) | (0.016) |
| Surname homogeneity× Year 2008 | 0.099*** | 0.037** | 0.030* | 0.022 | 0.014 |
| | (0.014) | (0.015) | (0.016) | (0.017) | (0.018) |
| Surname homogeneity× Year 2009 | 0.108*** | 0.055*** | 0.050*** | 0.054*** | 0.058*** |
| | (0.015) | (0.015) | (0.015) | (0.017) | (0.017) |
| Surname homogeneity× Year 2010 | 0.018 | -0.017 | -0.020 | -0.003 | -0.005 |
| | (0.015) | (0.016) | (0.016) | (0.017) | (0.018) |
| Surname homogeneity× Year 2011 | -0.015 | -0.069*** | -0.071*** | -0.057*** | -0.056*** |
| | (0.015) | (0.017) | (0.017) | (0.018) | (0.018) |
| Surname homogeneity× Year 2012 | -0.039** | -0.083*** | -0.082*** | -0.054*** | -0.051** |
| | (0.016) | (0.018) | (0.017) | (0.018) | (0.019) |
| Surname homogeneity× Year 2013 | 0.034** | -0.051*** | -0.048** | -0.063*** | -0.062*** |
| | (0.014) | (0.017) | (0.017) | (0.019) | (0.019) |
| Surname homogeneity× Year 2014 | -0.008 | -0.080*** | -0.073*** | -0.072*** | -0.060*** |
| | (0.016) | (0.019) | (0.018) | (0.020) | (0.019) |
| Surname homogeneity× Year 2015 | 0.013 | -0.044** | -0.033* | -0.034* | -0.023 |
| | (0.014) | (0.017) | (0.017) | (0.017) | (0.017) |
| Surname homogeneity× Year 2016 | -0.046*** | -0.102*** | -0.087*** | -0.111*** | -0.101*** |
| | (0.015) | (0.019) | (0.018) | (0.020) | (0.020) |
| Surname homogeneity× Year 2017 | -0.027 | -0.087*** | -0.067*** | -0.061*** | -0.049** |
| | (0.016) | (0.019) | (0.018) | (0.019) | (0.019) |
| Surname homogeneity× Year 2018 | -0.011 | -0.044** | -0.022 | -0.009 | 0.003 |
| | (0.015) | (0.019) | (0.018) | (0.017) | (0.017) |
| Surname homogeneity× Year 2019 | 0.078*** | 0.044** | 0.071*** | 0.082*** | 0.096*** |
| | (0.015) | (0.016) | (0.017) | (0.016) | (0.016) |
| Geographic controls | No | Yes | Yes | Yes | Yes |
| Transportation & firm num | No | No | Yes | Yes | Yes |
| Demographic controls | No | No | No | Yes | Yes |
| Socioeconomic controls | No | No | No | No | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 53,357 | 52,799 | 52,799 | 52,799 | 51,759 |
| $R^2$ | 0.247 | 0.258 | 0.259 | 0.264 | 0.268 |
| Mean dependent vars. | 0.145 | 0.145 | 0.145 | 0.145 | 0.146 |

Notes: County-level OLS regressions. The explanatory variables are the interactions between surname homogeneity among business owners and year dummies. We use two-way clustering of standard errors at the county and year levels. * p<0.1; ** p<0.05; *** p<0.01.