Why do Chinese SMEs avoid formal employment? Political connections and unequal access to formalization benefits

Ruochen Dai, Yue Feng, Chengfang Liu, Langrui Li, Lina Zhang, Ketong Zhu

PII: S1043-951X(25)00296-2

DOI: https://doi.org/10.1016/j.chieco.2025.102638

Reference: CHIECO 102638

To appear in: *China Economic Review*

Received date: 31 January 2025

Revised date: 21 November 2025

Accepted date: 16 December 2025

Please cite this article as: R. Dai, Y. Feng, C. Liu, et al., Why do Chinese SMEs avoid formal employment? Political connections and unequal access to formalization benefits, *China Economic Review* (2024), https://doi.org/10.1016/j.chieco.2025.102638

Why Do Chinese SMEs Avoid Formal Employment? Political Connections and Unequal Access to Formalization Benefits*

Ruochen Dai[a], Yue Feng[b], Chengfang Liu[c], Langrui Li[d], Lina Zhang[e], Ketong Zhu[f,•]

[a] Central University of Finance and Economics, Beijing 102206, China; Email: r.dai@cufe.edu.cn

[b] China University of Petroleum, Beijing 102249, China; Email: yue_feng_cup_2025@163.com

[c] China Center for Agricultural Policy, School of Advanced Agricultural Sciences, Peking University, Beijing 100871, China; Email: cfliu.ccap@pku.edu.cn

[d] School of Economics, Zhejiang University, Hangzhou 310058, China; Email: langruili@zju.edu.cn

[e] Fordhum University and Peking University, Bronx, NY 10458, United States; Email: lzhang114@fordham.edu

[f] School of Advanced Agricultural Sciences, Peking University, Beijing 100871, China; Email: zhuketong_2001@stu.pku.edu.cn

* Corresponding author is Ketong Zhu. We are grateful to the editors of this special issue and the anonymous referees for their insightful suggestions. We thank Zijun Cheng, Bing Lu, Shuanghe Shang, Wenkai Sun, Zhongwen Zhang, Xiaobo Zhang and participants in 24th China Economics Annual Conference for their valuable feedback. We gratefully acknowledge the financial support from the National Natural Science Foundation of China (#72342031, #72203252, and #42293273), and the Program for Innovation Research at the Central University of Finance and Economics. We thank the data support from the Center for Enterprise Research of Peking University. The authors are listed alphabetically according to the last names. Authors equally share the first authorship. All errors are ours.

• Corresponding author.
E-mail address: zhuketong_2001@stu.pku.edu.cn (K. Zhu)

Why Do Chinese SMEs Avoid Formal Employment? Political Connections and Unequal Access to Formalization Benefits

Abstract

Employment formalization among small and medium-sized enterprises (SMEs) remains low in developing countries, which weakens job security and social protection. Despite policy efforts, many SMEs evade social insurance obligations, raising concerns about labor rights and economic stability. Using data on 2,664 SMEs in the 2018 Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC), we document that 48.4% of SMEs provide no social insurance, while 74.6% cover only some employees. After controlling for firm characteristics as well as city and industry fixed effects, SMEs led by politically connected entrepreneurs have a 7.1-percentage-point higher probability of providing social insurance than those without such connections. This correlation likely stems from connected entrepreneurs' preferential access to formalization benefits—including formal credit and government procurement contracts—which collectively generate a political-connection premium that incentivizes compliance with labor regulations. These findings suggest that the uneven distribution of such benefits (concentrated among politically connected firms) contributes to the overall low formalization rates in China's SME sector, as most entrepreneurs lack access to these incentives. Moreover, the political-connection premium diminishes in regions with stronger financial inclusion and lower government intervention, suggesting that an improved business environment encourages broader formalization. These findings underscore the importance of policies that decouple formalization benefits from political connections to foster more inclusive labor protection and economic stability.

Keywords: SMEs; Employment Formalization; Political Connection; Business Environment

JEL codes: L26, and O17

I. Introduction

Small and medium-sized enterprises (SMEs) play a critical role in fostering employment and the economy, especially in developing countries (Aga et al., 2015; Xu et al., 2023). Statistics from the World Bank suggest SMEs account for 90 percent of businesses and provide more than 50 percent of employment worldwide, and they contribute to 40 percent of the GDP in emerging economies(Faye & Goldblum, 2022).

Despite their contributions to the economy and employment, the employment

formalization of SMEs still faces challenges. According to the ILO, a substantial number of workers in SMEs worldwide are employed informally as their working arrangements, whether de facto or de jure, are not covered by national labor legislation, income taxation or entitlements to social protection and other employment benefits (e.g., advance notice of dismissal, severance pay, paid annual or sick leave) (Kok & Berrios, 2019).

Employment formalization is a key goal of the Chinese central government, yet, as in many other countries, it has encountered unusually intense debate and particular market resistance. China passed the *Social Insurance Law* in 2010, requiring employers to make social insurance contributions for their employees as follows: pension, health insurance, work injury insurance, unemployment insurance, and maternity insurance. However, coverage remains far from universal. Statistics from the Ministry of Human Resources and Social Security show that by the end of 2023, out of the 740.41 million workers nationwide, only 243.73 million (or 32.9 percent) were enrolled in unemployment insurance (MOHRSS, 2024). From the employer's perspective, using first-hand data from 2,664 SMEs surveyed in China's 2018 Enterprise Survey for Innovation and Entrepreneurship (ESIEC), we find that nearly half (48.4%) of these firms offer no social insurance at all, while three-quarters (74.6%) don't provide it to all employees.

This situation raises a natural question: why does the employment formalization process remain sluggish in SMEs? Previous studies on the firms' formalization decisions typically adopt a cost-benefit analysis framework (e.g. Ulyssea, 2018, 2020), assuming that firms act as rational agents who weigh the costs of formalizing against its benefits. Hiring formal employees with social insurance constitutes an important aspect of firms' formalization decisions, which have been commonly discussed in the existing literature (e.g., Perry et al., 2007). While compliance with social insurance obligations undeniably increases labor costs, formalization also brings key benefits, such as enhanced corporate credibility, improved access to bank loans or capital market

financing, and greater business reputation, all of which facilitate business growth(Feng, 2013).[1] Accordingly, existing policy discussions on promoting formalization have largely focused on either lowering its costs or increasing its benefits. This decision-making process hinges on firms' cost-benefit trade-offs: while remaining informal avoids formalization costs, it also forfeits associated benefits (e.g., access to credit, government contracts) and carries potential compliance risks; firms opt for formalization only when perceived benefits outweigh costs.

Both academic literature and policy efforts have predominantly focused on addressing cost-side barriers to formalization—such as stringent entry regulations and high formalization costs—as the primary obstacles to firm formalization, yet evidence shows these approaches have yielded limited success. Policy efforts worldwide (e.g. Doing Business Project by the World Bank) have focused on reducing formalization costs to encourage informal firm formalization (World Bank, 2013, 2020), yet evidence shows these reforms have limited impacts on actual formalization (Bruhn & McKenzie, 2014; Floridi et al., 2020; Rocha et al., 2018; Ulyssea, 2020). Such policies may boost registrations but often fail to improve firm performance (Benhassine et al., 2018; De Mel et al., 2013). China's 2013 business registration reform similarly increased registrations (Barwick et al., 2025), while leaving most formalized firms still withholding social insurance - revealing persistent informality.

These findings point to critical yet underexplored factors on the benefits side that discourage SMEs from formalizing. In developing countries without profound formal institutions, even formalized SMEs often struggle to realize the benefits of formalization (e.g. Sri Lanka: de Mel et al., 2013; Brazil: De Andrade et al., 2016). This

[1] For example, in China, the National Enterprise Credit Information Publicity System (https://www.gsxt.gov.cn/index.html) discloses information about registered businesses, including their basic registration details, records of administrative penalties or violations, and annual reports that specify the number of employees covered by each of the five mandatory social insurance programs. When firms apply for bank loans, banks typically evaluate their creditworthiness based on social insurance contributions and other related factors. Moreover, businesses seeking government contracts must undergo compliance checks, with social insurance coverage serving as an important criterion. Additionally, platforms such as Alipay and Qichacha allow individuals and other business entities to query a business's credit information, with social insurance participation serving as a strong indicator of its financial stability and credibility.

limited value of formalization is further exacerbated by its uneven distribution: access to critical benefits—such as bank loans (Faccio, 2006; Khwaja & Mian, 2005; Li et al., 2008; M. Yu & Pan, 2008) and government contracts (Brogaard et al., 2020; Dicko, 2016; Goldman et al., 2013) —is disproportionately captured by politically connected entrepreneurs. As a result, unconnected entrepreneurs rationally perceive low returns to formalization, perpetuating their informal status. Since the majority of entrepreneurs lack political connections, this mechanism may help explain why the formalization process remains sluggish in developing countries. Analysis based on the ESIEC data also reveals substantial regional heterogeneity in SME formalization: formal employee shares average 57.6% in developed provinces like Shanghai, but drop to 21.4%-30.8% in less developed regions such as Gansu, Henan, and Liaoning.

To formalize our theoretical expectations, we extend Ulyssea's (2020) cost-benefit framework by focusing on how political connections and the business environment shape firms' formalization decisions through the access to formalization benefit. Specifically, politically connected entrepreneurs can secure low capital costs (e.g., via preferential access to formal credit) when operating formal firms, giving them higher net benefits from formalization. In contrast, unconnected entrepreneurs' formalized firms still face higher capital costs, which initially discourages formalization. However, when the business environment improves (e.g., through digital finance platforms or inclusive financial policies), unconnected firms also gain access to affordable capital, narrowing the cost gap between the two groups.

This generates a selection mechanism: connected entrepreneurs, with inherently lower capital costs, have a lower formalization cutoff (the threshold where expected benefits outweigh costs) than unconnected entrepreneurs, making them more likely to formalize and hire formal employees (Proposition 1). Meanwhile, as the business environment improves and unconnected firms' capital costs decline, the additional advantage of political connections diminishes, reducing the formalization gap between connected and unconnected groups (Proposition 2).

We tested these propositions using data from the 2018 ESIEC survey. Regression results controlling for firm characteristics, city, and industry fixed effects show that SMEs led by politically connected entrepreneurs are 7.1 percentage points more likely to provide social insurance for employees—a proxy for formalization—than unconnected counterparts (supporting Proposition 1). Further regional heterogeneity analysis reveals that while politically connected entrepreneurs exhibit a higher likelihood of formalization than unconnected ones across all regions, this formalization gap narrows in areas with better financial inclusion or weaker government intervention (supporting Proposition 2). Specifically, we measure financial inclusion development via the Peking University Digital Financial Inclusion Index of China (PKU_DFIIC) and government intervention intensity through entrepreneurs' 1–10 ratings of government officials' assistance impact in the ESIEC survey.

Robustness checks employing IV regression (with pre-entrepreneurship public-sector employment as an instrument) and PSM analysis affirm a positive correlation between political connections and formalization, underscoring the significance of disproportionately captured formalization benefits (e.g., bank loans, government contracts) for connected entrepreneurs. While these findings do not establish causality due to potential omitted variable biases, they demonstrate that the association between political connections and formalization benefits is central to understanding firm formalization process. Regionally, better business environments correlate with higher social insurance contribution rates, highlighting that promoting equitable access to formalization benefits—rather than expanding political connections—is pivotal for addressing Chinese SMEs' avoidance of formal employment.

This study contributes to existing research in three main ways. First, leveraging the unique and representative micro-level data, we provide a description of formal employment patterns among SMEs in China across industries, provinces and firm registration cohorts. Existing studies have documented widespread evasion of social insurance contributions by businesses in China, using data on enterprises of all sizes in

certain regions(e.g., Shanghai: Nyland et al., 2006, 2011; Zhejiang and Jiangsu: Song et al., 2021) or focusing on large-scale industrial firms (Feng, 2013; Zhao et al., 2016). However, micro-level evidence focused on SMEs remains limited. Although our sample size is smaller in scale, it captures detailed firm-level behaviors within a segment often underrepresented in mainstream datasets, thus providing a valuable complement to the existing literature. Moreover, even after China's sweeping business registration reform in 2013 that dramatically lowered entry barriers and spurred a surge in new firm registrations(Barwick et al., 2025; Cheng et al., 2024), we find that many formally registered SMEs continue to hire workers informally, failing to comply with social insurance obligations. While much of the literature focuses on the increase in registered firms as an indicator of successful formalization(e.g., Bruhn & McKenzie, 2014), analysis in this paper highlights the persistent challenges at the intensive margin of formalization, where progress in formalizing employment within firms remains sluggish.

Second, we contribute to the literature by incorporating political connections into the cost-benefit analysis framework of SMEs' formalization decisions. While models in the literature typically assume that formal firms automatically reap the benefits of formalization(e.g., Ulyssea, 2020), we emphasize that these benefits are often unequally distributed among SMEs: entrepreneurs without political connections capture very few of them. This refinement helps explain the limited impact of past formalization initiatives observed in the literature(De Mel et al., 2013; Bruhn & McKenzie, 2014; Floridi et al., 2020; Benhassine et al., 2018; Rocha et al., 2018; Ulyssea, 2020), and offers a nuanced explanation for why the formalization process remains sluggish in many developing countries: the majority of enterprises lack political connections and are thus unable to secure key formalization benefits, which leaves them with little incentive to formalize.

Finally, this study offers novel insights into the role of the business environment in shaping the formalization process and draws out policy implications for fostering SMEs'

formalization. Given the high prevalence of informality among SMEs, it is crucial to find effective ways to foster their formalization and development, so that they can create more formal jobs with decent benefits. Prior studies have highlighted the positive impact of such policy instruments as reductions in tax rates or social security contribution rates in promoting firms' willingness to provide social insurance for their employees (Feng, 2013; Nyland et al., 2011; Song et al., 2021; Zhao et al., 2016). However, governments in many less-developed regions often face fiscal constraints that make such measures difficult to sustain. We argue that lacking political connections and the resulting inability to access formalization benefits poses a significant barrier for SMEs' formalization. Importantly, we find this disadvantage is significantly reduced in regions with a fairer and more inclusive business environment, where the formalization premium of political connections is attenuated. Therefore, improving the business environment and dismantling the channels that link formalization benefits to political connections could serve as an policy alternative to encourage regulatory compliance among SMEs.

The remainder of the paper is organized as follows. Section II describes the ESIEC firm survey, study sample and data. Section III presents the status of formal employment among SMEs in China. Section IV introduces the conceptual framework followed by empirical results in Section V. Section VI concludes with policy implications.

II. Sample and Data

A. Survey Description and Sample Construction

The data are from the Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC), which is a field survey of Chinese private enterprises carried out by the Center for Enterprise Research, Institute of Social Science Survey of Peking University. It covers seven main areas: entrepreneurs' entrepreneurial history, the process of establishing enterprises, basic information, enterprise innovation, inter-enterprise relationships, and the business environment. The baseline survey was conducted in

2018, covering SMEs in the following six provinces: Henan, Guangdong, Zhejiang, Shanghai, Gansu, and Liaoning. Using a stratified sampling strategy in these six provinces, 117 counties were selected in total. In each county, private enterprises and self-owned businesses established in the 2010-2017 period from the China National Business Registration Database were randomly selected.[2] The total number of enterprises approached was 58,500, encompassing 18 industry categories.

In 2018, the number of completed interviews for the ESIEC sample was 6,198, comprising 4,148 corporate entities, 195 partnerships, and 1,855 *Getihu* (self-owned businesses). The ESIEC survey prioritized entrepreneurs and founders as respondents. If these individuals were unavailable, executives were approached as alternates.

First, we excluded *Getihu* (self-owned businesses) as their operations resemble individual or family-run enterprises, characterized by minimal employee hiring and typically low participation in social insurance programs. Furthermore, **93.61%** of *Getihu* in the sample reported having no political connections. Since this study examines the employment formalization of SMEs and the role of entrepreneurs' political connections, and considering that *Getihu* rarely pay social insurance or have political connections, including them in the sample could potentially overestimate the correlation and bias the results. Moreover, their inclusion might overshadow the issue we aim to highlight that formal firms face challenges in accessing benefits of formalization, which hinders the employment formalization of SMEs. This exclusion reduced the sample to 4,343 firms. Next, as our study focuses on the role of entrepreneurs' political connection at the startup phase, we retained only enterprises where the respondents were the entrepreneurs involved in the establishment of the business as shareholders or partners, and excluded those obtained through **inheritance or gifts**, resulting in a sample of 3,093 firms. Finally, we excluded enterprises with

[2] The China National Business Registration Database is from the State Administration of Industry and Commerce (SAIC), and contains the registration information for all the firms and self-employed businesses(*Getihu*), including registration dates, locations, contact information, business operation scope, and type of industry.

operational irregularities during the survey time, and those with missing or anomalous core variables such as information about firm history and employee data. The final sample size is 2,664. We provide a detailed explanation of the sample construction process and the specific reasons for each step in Fig. A.1.

B. Definition of the Political Connection

Previous studies have often identified firms' political connections based on the political backgrounds of company executives, such as board chairpersons and general managers (Faccio, 2006; Luo & Tang, 2009; Luo & Zhen, 2008; Shen & Zou, 2017; W. Yu et al., 2012). For instance, Yu et al. (2012) and Shen and Zou (2017) use positions such as membership in the Chinese People's Political Consultative Conference (CPPCC) as a measure of political connections. Since these studies predominantly focus on listed companies, political backgrounds are typically derived from public information about executives, potentially overlooking more subtle factors, such as entrepreneurs' social networks at the time of the firm's founding. Moreover, due to data limitations, there has been little focus on political connections in SMEs.

This paper identifies entrepreneurs' political connections based on five questions from the ESIEC (2018) questionnaire. These questions cover all relevant aspects that allow us to examine entrepreneurs' social networks, government-business interactions, and the special advantages firms gain in accessing external resources during the initial stages of firm creation, thereby helping us to identify the potential political connections of entrepreneurs.

These five specific indicators are categorized into three dimensions. The first dimension, *government startup support*, examines whether the firm received direct subsidies at its inception, and whether the entrepreneur was guided by the government in starting the business. The second dimension, *shareholder relationship*, captures a firm's political connections from the perspective of relationships between shareholders and the government. Related variables include whether there were any shareholders specifically designated to coordinate relations with the government, and whether the

entrepreneur formed partnerships with shareholders through government introductions. The third dimension, *social network*, assesses whether the entrepreneur had acquaintances working in government departments at the time of the firm's founding. If a firm meets any of the criteria within these dimensions, it is considered politically connected, and the dummy variable *political connection* is assigned a value of 1; otherwise, the value is 0, indicating no political connection.

Table 1 provides descriptive statistics for the variables used in this study. Although our method for identifying political connections was designed to capture all possible types of such linkages, the results show that the vast majority of SMEs remained unconnected, as only 23.61% of the sample firms had such ties. Specifically, 10.96% of firms had benefited from government support policies or subsidies, 1.69% had established equity relationships with government departments, and 14.49% had connections with individuals working in government departments.

C. A Brief Review of Institutional Background and Definition of Formal Employment

Social insurance is an essential component of employee benefits, making their administration central to labor relations. Over the past decades, numerous laws, regulations, and provisions have been gradually established or revised in China to ensure employees' lawful entitlement to social security.

In the mid-1980s, several cities in Guangdong, Jiangsu, Fujian, and Liaoning provinces took the lead in piloting reforms of the pension insurance system. The central government followed by introducing a series of policy documents that paved the way for the establishment of the pension insurance system for urban employees, which was implemented nationwide in 1997. Subsequently, additional policy measures were introduced to expand the coverage of this system while also initiating the development of other social insurance schemes, such as medical insurance. In 1994, the ***Labor Law of the People's Republic of China*** was enacted, and its Article 72 stipulated: "Employers and employees must participate in social insurance and pay social insurance premiums in accordance with the law", which legally established the social

insurance obligations of employers and employees for the first time. In 1999, the ***Interim Regulation on Social Insurance Premium Collection*** was issued and explicitly required the collection of pension, medical, and unemployment insurance premiums from both employers and employees for the first time (Zhao et al., 2016).

After more than two decades of active exploration, China promulgated the ***Social Insurance Law*** in 2010. As a dedicated legal framework for the field of social insurance, this law further clarified and reinforced the statutory obligations of enterprises to provide employees with social insurance, including pension, medical, unemployment, work injury, and maternity insurance. It also provided a more detailed legal basis for the management and resolution of social insurance-related disputes.

In this paper, we define formal employees as those who are covered by social insurance in the context of China. We consider compliance with social insurance contributions a key indicator of formal employment, primarily because it is a legal obligation for businesses to provide such benefits to employees with formal labor relationships. As long as an employment relationship is established and a labor contract is signed, the employer must pay social insurance premiums for employees on time and in full. Additionally, although less emphasized, existing literature commonly considers this as a characteristic of formality at the intensive margin (e.g., Perry et al., 2007).

The ESIEC survey inquired about the number of full-time employees and how many of them were covered by social insurance (including the five basic insurances and housing fund). This enables us to describe the formal employment status of SMEs at the firm level. We construct a continuous variable, *formal employee share,* which represents the proportion of employees covered by social insurance within the firm. Firms that reported not employing any full-time employees or that did not provide social insurance to any of their employees in 2017 are assigned a value of zero. Additionally, to present descriptive statistics more clearly later in the paper, we classify firms into two discrete categories based on their adherence to social insurance requirements: those without any full-time employees or without providing social

insurance to any of their employees (intensive informal firms, IIFs) and those that provided social insurance for at least one employee (intensive formal firms, IFFs).

III. Formal Employment of SMEs in China

A. Distribution of Formal Employment among SMEs Registered across Different Years

Our data show that, in 2017, the average proportion of employees covered by social insurance (*formal employee share*) in the sample SMEs was 36.4 percent.[3] Specifically, 48.4 percent of the sample SMEs provided social insurance to no one, while 74.6 percent failed to cover all their employees. Additionally, as shown in Fig. 1, the average *formal employee share* of SMEs has actually decreased despite the implementation of the business registration reform initiated in 2013, which significantly lowered the entry barriers and formalization costs.[4] Empirical studies have established a causal relationship between the reduction in formal entry costs and the significant rise in the number of registered businesses (Barwick et al., 2025; Cheng et al., 2024). While much of the literature focuses on the surge in registered firms as an outcome of policies aimed at formalization, our analysis reveals an important nuance: despite the rise in formal registrations, a significant portion of these registered businesses do not fully comply with social insurance contribution requirements, which reflects a persistent challenge in formalization at the intensive margin. Registered firms still face trade-offs between operating formally versus informally, such as whether to provide social insurance for their employees. The overall proportion of formal employees covered by social insurance has not kept pace with the increase in the number of registered firms, underscoring the ongoing difficulties in advancing employment formalization.

[3] As noted by Feng (2013), firms may also reduce their social insurance contribution base by methods such as not signing labor contracts or underreporting the number of employees, thereby avoiding the costs associated with social insurance contributions. Therefore, it must be acknowledged that the measurement of formal employment in this paper may be somewhat overestimated.

[4] Fig. A.2. in appendix represents the average proportion of intensive formal firms (i.e., SMEs with at least one formal Employee) across different registration year and shows similar pattern to Fig. 1.

Furthermore, we describe the key differences between the intensive informal firms and intensive formal firms in their basic characteristics in Table 2. We find that intensive informal firms were generally smaller in size and had lower levels of both entrepreneurial and employee capabilities compared to intensive formal firms. These intensive informal firms, which were entirely informal in terms of offering formal employment, exhibited characteristics strikingly similar to those of informal firms at the extensive margin, specifically, unregistered firms, as highlighted in the existing literature (La Porta & Shleifer, 2008, 2014; Perry et al., 2007; Ulyssea, 2020).

In terms of firm characteristics, we find that intensive formal firms tended to be larger, which is consistent with Ulyssea's findings, suggesting that the level of informality at the intensive margin decreases as firm size increases(Ulyssea, 2018). Specifically, row (4) shows intensive formal firms had a longer average lifespan (calculated as the time from the firm's registration to the time of the survey). Additionally, they had on average, almost twice the total assets of intensive informal firms and four times the operating income in 2017, as shown in row (5) and (6). In terms of employee numbers, row (8) reports that intensive formal firms had an average of 22.51 full-time employees, while intensive informal firms had an average of 7.68 full-time employees in 2017. These differences were all statistically significant at the 1% level.

There were also significant differences in capabilities of entrepreneurs and employees between the two types of firms. Specifically, entrepreneurs' abilities are measured by whether they have received education at the junior college level or above and by their self-assessed pre-entrepreneurial income relative to local standards. Row (1) and (3) show that intensive formal firms were more likely to be operated by entrepreneurs with higher educational attainment and pre-entrepreneurial income levels, with these differences being statistically significant at the 1% level. Besides, row (9) and (10) reveal that intensive formal firms hired a significant larger proportion of educated and skilled employees compared to their intensive informal counterparts.

B. Distribution of Formal Employment among SMEs across Provinces and Industries

This subsection examines the distribution of formal employment provided by SMEs across different provinces and industries. Fig. 2 reports the average formal employee share within SMEs in six representative provinces. The results show that formal employees were more concentrated in better developed regions. The *formal employee shares* were 57.6% in Shanghai, 50.6% in Guangdong, and 46.7% in Zhejiang. In Liaoning, Henan, and Gansu, however, these shares dropped to 30.8%, 22.6%, and 21.4%, respectively. A similar pattern is observed for the proportion of intensive formal firms in Fig. A. 2, with rates of 71.0%, 73.9%, and 62.2% in Shanghai, Guangdong and Zhejiang, compared to significantly lower rates of 43.2%, 34.1%, and 32.7% in Liaoning, Henan, and Gansu, respectively.

The distribution of formal employment among SMEs also varied significantly across industries, as illustrated in Fig. 3. We rank sectors from highest to lowest based on the average formal employee share. Overall, industries with higher regulatory requirements and capital intensity tended to have higher average formal employee share, while those with lower regulatory thresholds or labor intensity exhibited more informality in employment. Particularly, Scientific Research, Energy, and Mining sectors had highest average formal employee share, all of which exceed 50%, with shares of 76.2%, 75.8%, and 50.8%, respectively. In contrast, Agriculture has the lowest formal employment rate at 20.9%, followed by Leisure & Entertainment (31.3%), Hospitality (33.6%), Retail (33.8%), and Residential Services (34.1%). Similarly, when examining the proportion of intensive formal firms, we find that the Mining and Energy stood out, with all firms in these industries offering social security for at least one full-time employee, closely followed by Health (74.1%). In contrast, the proportions in Agriculture (29.6%), Leisure & Entertainment (41.5%), Retail (44.8%), and Residential

(49.0%) remained below 50%.[5]

These findings suggest that even as business registration serves as a crucial first step toward formalization, which spurs a large number of firms to formalize at the extensive margin, SMEs still face challenges in formalizing at the intensive margin, particularly in hiring formal employees and advancing employment formalization.

IV. Conceptual Framework

In this section, we develop a conceptual framework to analyze entrepreneurs' decisions regarding formal employment. We proceed by first establishing a simple benchmark model drawing on the cost-benefit analysis framework of Ulyssea (2020). Next, we extend this model by incorporating two important factors, namely firms' political connections and the quality of the business environment, and analyze how they alter the incentives and returns to formalization. Finally, we derive two testable propositions and validate the core argument of the conceptual framework using survey data.

A. Benchmark Model

To begin, we consider the simplified static optimization model without incorporating political connections. We assume that each firms' productivity is predetermined and heterogeneous, denoted as θ. The cumulative distribution function of θ is $F(\cdot)$. As rational agents, entrepreneurs weigh the costs and benefits of employment formalization.[6] While complying with formal regulations incurs a social

[5] In Table A1 and Table A2, we present the distribution of formal employment among SMEs across provinces and industries in detail based on three variables, including the average proportion of formal employees within SMEs (formal employee share), the proportion of SMEs with at least one formal employee (intensive formal firms) and the proportion of SMEs where all employees are formally employed.

[6] Following Ulyssea (2020), we model the firm's decision to formalize employment as a binary choice of whether to hire formal employees. This approach allows us to focus on the core mechanism: how political connections influence marginal firms' incentives to comply with formal regulations. We acknowledge that this simplification abstracts from intermediate scenarios, such as partially formal firms that contribute social insurance for only a subset of employees. In the ESIEC 2018 survey, 48.4% of SMEs provide no social insurance to employees, 25.4% of SMEs cover all employees, and 26.2% of firms fall into the "partial coverage" category. A more realistic extension would allow firms to choose the proportion of formal workers and receive an intermediate interest rate, resulting in two productivity cutoffs: firms below the lower threshold would remain fully

insurance burden, denoted as τ in the model, it also brings benefits, such as improved access to bank loans or capital market financing (Feng, 2013). In our model, we operationalize formalization benefits primarily through low capital costs. This conceptual choice is empirically grounded: in practice, banks can assess creditworthiness and risk based on social insurance contributions, such that only formal firms (those hiring formal employees) qualify for the loans from formal financial institutions. Formal loans in China carry significantly lower capital costs than informal financing. These capital costs are denoted as r_f and r_i, where the subscripts *f* and *i* represent formal and informal, respectively.[7] Besides, intensive formal firms are required to bear a one-time cost, denoted as C_f, for registering their employees with the government's social insurance agency.

Intensive formal firms and intensive informal firms respectively make capital investment decisions each period with the aim of maximizing their current profits. We assume a Cobb-Douglas production function $y = y(k;\theta) = \theta k^{\alpha}$, then the profit functions for these two types of firms are as follows:

$$\pi_f(\theta) = (1-\tau)\theta {k_f}^{\alpha} - r_f k_f \tag{1}$$

$$\pi_i(\theta) = \theta {k_i}^{\alpha} - r_i k_i \tag{2}$$

From the first-order conditions, we derive the optimal capital k_f^* and k_i^*:

$$k_f^* = [\frac{r_f}{\alpha\theta(1-\tau)}]^{\frac{1}{\alpha-1}} \tag{3}$$

$$k_i^* = [\frac{r_i}{\alpha\theta}]^{\frac{1}{\alpha-1}} \tag{4}$$

Substituting these optimal capital allocations back into the profit functions yields the following expressions for the profits of both types of firms:

$$\pi_f(\theta) = \frac{1-\alpha}{\alpha} r_f \left[\frac{\alpha\theta(1-\tau)}{r_f}\right]^{\frac{1}{1-\alpha}} \tag{5}$$

informal, those above the upper threshold would fully formalize, and those in between would partially formalize. We formalize this extension in Appendix 2.

[7] For simplifications, we center our analysis entirely on capital considerations. A large strand of literature indicates that the formal credit from banks is cheaper than informal lending. Thus, we can initially conceive that $r_f < r_i$. Additionally, we assume $r_i > r_f(1-\tau)^{\frac{1}{\alpha}}$.

$$\pi_i(\theta) = \frac{1-\alpha}{\alpha} r_i \left[\frac{\alpha\theta}{r_i}\right]^{\frac{1}{1-\alpha}} \quad (6)$$

Firms face a binary decision on whether to hire formal employees and to become intensive formal. They will choose to be intensive formal if the discounted future benefits of formalizing employment, minus the registration costs, outweigh the future benefits of remaining intensive informal. This decision criterion can be represented mathematically as follows:

$$V_f(\theta) - V_i(\theta) \geq C_f \quad (7)$$

Assume β is the discount rate and firms are infinitely lived, the value functions are:

$$V_f(\theta) = \sum_{t=0}^{\infty} \beta^t \pi_f(\theta) \equiv \frac{1-\alpha}{\alpha(1-\beta)} r_f \left[\frac{\alpha\theta(1-\tau)}{r_f}\right]^{\frac{1}{1-\alpha}} \quad (8)$$

$$V_i(\theta) = \sum_{t=0}^{\infty} \beta^t \pi_i(\theta) \equiv \frac{1-\alpha}{\alpha(1-\beta)} r_i \left[\frac{\alpha\theta}{r_i}\right]^{\frac{1}{1-\alpha}} \quad (9)$$

Combining and solving for equations (7), (8), and (9) yield the productivity threshold θ^*, as shown in equation (10). When θ exceeds θ^*, the firm chooses to become an intensive formal firm. Consequently, the proportion of intensive formal firms is given by $1 - F(\theta^*)$.

$$\theta^* = \left\{ \frac{\frac{\alpha(1-\beta)}{1-\alpha} C_f}{\left[\frac{(\alpha(1-\tau))^{\frac{1}{1-\alpha}}}{r_f^{\frac{\alpha}{1-\alpha}}} - \frac{\alpha^{\frac{1}{1-\alpha}}}{r_i^{\frac{\alpha}{1-\alpha}}}\right]\alpha^{\frac{1}{1-\alpha}}} \right\}^{1-\alpha} \quad (10)$$

A simple comparative static analysis reveals that several factors raise the threshold for employment formalization: increases in the registration cost (C_f) and the social security burden (τ), a higher capital cost for intensive formal firms (r_f), and a decrease in the capital cost for intensive informal firms (r_i). This, in turn, leads to a decline in the proportion of intensive formal firms.

B. Extended Model Incorporating Political Connection and Business Environment

We then extend this benchmark model by introducing a key distinction: while Ulyssea's model treats formalization status as the sole determinant of access to formalization benefits, represented in the model as lower capital rental costs, we argue

that such benefits are not universally accessible to all formalized SMEs. Instead, they are primarily available to intensive formal firms led by politically connected entrepreneurs. We also consider the role of the business environment in determining the extent to which firms without political connections can access lower capital rental costs.

In fact, due to the underdevelopment of the formal institution, most private enterprises in China, especially small and micro enterprises, face significant financing difficulties (Lin & Li, 2001). Evidence shows that many firms in developing countries fail to access the purported advantages of formal status, even after formal registration (Benhassine et al., 2018; De Andrade et al., 2016; De Mel et al., 2013). Moreover, the benefits of formalization, such as access to bank loans(Faccio, 2006; Khwaja & Mian, 2005; Li et al., 2008; M. Yu & Pan, 2008) and government contracts (Brogaard et al., 2020; Dicko, 2016; Goldman et al., 2013), are typically accessed by entrepreneurs with political connections, especially in settings with weak institutions.

To formally capture this reality, we incorporate two additional variables into the modified conceptual framework: *pc* (political connection) is a binary variable that equals 1 if the entrepreneur has political connections. *be* (business environment) is a continuous variable ranging from 0 to 1, with higher values indicating a more favorable institutional setting in the region where the firm operates.

We further refine the assumptions regarding capital costs faced by different types of firms. We set subscripts h and l to represent high cost and low cost, respectively. For intensive formal firms, those with political connections benefit from low capital costs r_l, while those without such connections encounter higher capital costs which are influenced by the business environment, as shown in equation (11). For example, in regions with the worst business environments, specifically when be=0, firms without political connections face high capital costs r_h identical to that faced by intensive informal firms. For intensive informal firms, non-compliance with social insurance requirements renders them ineligible for low-cost formal credit regardless of whether they have political connections. Thus, they consistently face high capital costs, denoted

as r_h, as shown in equation (12).

$$r_f = \begin{cases} r_l & when\ pc = 1 \\ r_h - (r_h - r_l) \cdot be & when\ pc = 0 \end{cases} \tag{11}$$

$$r_i = \begin{cases} r_h & when\ pc = 0 \\ r_h & when\ pc = 1 \end{cases} \tag{12}$$

After redefining the costs faced by firms with political connections and those without, we rewrite the productivity threshold for employment formalization as follows:

$$\theta_{pc=1}^{*} = \left\{ \frac{\frac{\alpha(1-\beta)}{1-\alpha} C_f}{\left[\frac{(\alpha(1-\tau))^{\frac{1}{1-\alpha}}}{r_l^{\frac{\alpha}{1-\alpha}}} - \frac{\alpha^{\frac{1}{1-\alpha}}}{r_h^{\frac{\alpha}{1-\alpha}}} \right] \alpha^{\frac{1}{1-\alpha}}} \right\}^{1-\alpha} \tag{13}$$

$$\theta_{pc=0}^{*} = \left\{ \frac{\frac{\alpha(1-\beta)}{1-\alpha} C_f}{\left[\frac{(\alpha(1-\tau))^{\frac{1}{1-\alpha}}}{(r_h - (r_h - r_l) \cdot be)^{\frac{\alpha}{1-\alpha}}} - \frac{\alpha^{\frac{1}{1-\alpha}}}{r_h^{\frac{\alpha}{1-\alpha}}} \right] \alpha^{\frac{1}{1-\alpha}}} \right\}^{1-\alpha} \tag{14}$$

Clearly, when $be < 1$, $\theta_{pc=0}^{*} > \theta_{pc=1}^{*}$, indicating that firms without political connections face a higher productivity threshold when deciding whether to provide formal employment. Therefore, we propose that:

Proposition 1: *Firms led by entrepreneurs with political connections are more likely to hire formal employees.*

Further comparative static analysis shows that as the business environment improves, $\theta_{pc=0}^{*}$ gradually decreases. Since the proportion of intensive formal firms is given by $1 - F(\theta_{pc=0}^{*})$, an improving business environment leads to a higher proportion of firms without political connections willing to hire formal employees. Meanwhile, as the business environment improves, the additional advantages conferred by political connections in motivating formalization decline, which will weaken the effect of political connections on the probability of hiring formal employees. Based on this, we propose that:

Proposition 2: The formalization premium associated with political connections diminishes as the business environment improves.

C. Model Validation

Before the empirical analysis, we use data from the ESIEC (2018) to provide preliminary graphical evidence to validate two key aspects of our conceptual framework. First, we assess the premise that productivity is not the sole determinant of formalization. Under the benchmark model—where formalization decisions are driven solely by productivity-based selection—theoretically, formal and informal firms should exhibit distinct size distributions, with larger firms more likely to formalize due to higher productivity. However, Fig. 4 shows a substantial overlap in the size distributions of intensive formal and informal firms, despite the former having higher average total fixed assets (in Panel A) and total revenue (in Panel B) in 2017. This suggests that another critical predetermined factor alongside productivity, such as political connections, influences firms' decisions to formalize, which provide initial support for the argument of our conceptual framework.

Second, we directly examine the heterogeneous formalization benefits faced by different types of firms—an extension captured by key assumptions in our augmented model (formalized as Equations 11 and 12), which explicitly differentiate the formalization benefits for politically connected versus unconnected firms. Panels A and B in Fig. 5 compare the prevalence of two common formalization benefits in reality (i.e., access to bank loans and government contracts) between connected and unconnected entrepreneurs, separately for intensive formal and informal firms. Both bank loan and government contract are coded as dummy variables, with the former defined only for firms seeking finance. The results clearly show that, overall, firms that provide social insurances to employees are more likely to obtain these formalization benefits. However, the formalization premium, which refers to the benefit of being formal, is significantly larger for connected entrepreneurs than for their unconnected counterparts. This pattern of heterogeneous returns validates the core argument of our conceptual framework: political connections unlock the true value of formal status, explaining why many firms in developing countries, the majority of which lack political connections, have little incentive to formalize.

V. Empirical Analysis

A. *Model specification*

In this section, we test the two propositions from the conceptual framework through regression analysis. The baseline regression model used in this paper is as follows:

$$Y_{ijk} = \alpha_0 + \alpha_1 connected_{ijk} + \alpha_2 \mathrm{X}_{ijk} + \lambda_j + \delta_k + \varepsilon_{ijk} \quad (15)$$

Subscripts *i, j,* and *k* denote the firm, the city and the industry in which the firm operates, respectively. Y_{ijk} represents the dependent variable, which can be either a dummy variable, *intensive formal firm,* representing whether the firm employed at least one formal employee in 2017, or a continuous variable, *formal employee share,* measuring the proportion of formal employees within the firm. $connected_{ijk}$ is a dummy variable indicating whether the firm had political connections, as defined in detail in previous section. X_{ijk} represents a set of control variables that may potentially affect firm formalization and are related with the political connections. We control for entrepreneur's demographic variables that are correlated with the probability of having political connection and hiring formal employees, such as the gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income. Additionally, the firm's age and age squared after registration are included as controls to account for the firm size. λ_j and δ_k represent city and industry fixed effects, respectively. ε_{ijk} is the random disturbance term. All regressions use heteroskedasticity-robust standard errors. In the baseline regressions, we use ordinary least squares (OLS) estimation, as the coefficients are directly interpretable in marginal terms, which facilitates economic interpretation. Additionally, we report results from alternative estimation methods that account for the distributional features of different dependent variables. Specifically, we use a Logit model for binary outcomes and a Tobit model for the censored continuous variable measuring the share of formal employees.

B. *Baseline results*

To test Proposition 1, we empirically examine the association between entrepreneurs' political connections and firms' formal employment status. Table 3 shows that firms whose entrepreneurs did not have political connections were less likely to hire formal employees. In Columns (1)-(4), the dependent variable is a dummy variable that equals 1 if the firm had at least one formal employee in 2017(intensive formal firm). Columns (5)-(8) use a continuous variable representing the proportion of formal employees within the firm (formal employee share), which equals zero when the firm had no formal employee. Columns (1) and (5) present results based on the most parsimonious model without control variables and fixed effects. The simplest regressions suggest that there is a significant positive correlation between firms' formalization decision and entrepreneurs' political connections. Firms whose entrepreneur had political connection were 7.8 percentage points more likely to be intensive formal firm and had an 8.2 percentage point larger share of formal employees at the survey time. In columns (2) and (6), we further control for city and industry fixed effects, and both OLS coefficients remain statistically significant at the 1% level. In columns (3) and (7), we add control variables that proxy for entrepreneurs' ability and firm size. Entrepreneurs' pre-entrepreneurial income and education both positively predict firms' employment formalization, indicating the role of entrepreneurial capacity. While the OLS coefficients on political connection decrease slightly, they remain significant with 7.1 percentage points for being an intensive formal firm and 7.5 percentage points for formal employee share. To better account for the distributional features of the dependent variables, we further adjust our estimation strategies by employing a Logit model for the binary outcome and a Tobit model for the censored continuous variable in Columns (4) and (8), respectively. The Logit coefficient is 0.361, yielding a relative marginal effect of 0.071, while the Tobit coefficient is 0.132 with an average marginal effect of 0.073. Both the Logit and Tobit estimates confirm the significant positive effect of political connections, and their magnitudes are close to those of the OLS estimates, reinforcing the robustness of our baseline results. These

findings suggest that entrepreneurs' political connections, in addition to their ability, play a critical role in firms' employment formalization.[8]

Our core argument hinges on the premise that political connections confer a premium in accessing formalization benefits. To validate this foundational element of our conceptual framework, we provide further regression evidence by examining the relation between political connections and the likelihood of firms securing formalization benefits, specifically in terms of access to bank loans and government contracts. We focus on these outcomes because they are both crucial for business development and often hard for SMEs to secure. As found by La Porta and Shleifer(2008, 2014) in their survey of private enterprises, both informal and formal entrepreneurs commonly cited access to financing as a major barrier to business development, meanwhile the benefits of formalization in expanding the customer base were also relatively limited(De Mel et al., 2013; McKenzie & Sakho, 2010).

The ESIEC (2018) questionnaire asked about the types of financing entrepreneurs used at the startup phase, distinguishing between loans from acquaintances (informal lending) and loans from non-acquaintances, such as financial institutions (formal lending). Additionally, the survey inquired whether the firm secured a government or state-owned enterprise (SOE) contract in 2017, which indicates a more stable and broader revenue stream. Based on these questions, we constructed two dummy variables, *bank loan* and *gov contract*. The variable *bank loan* equals 1 if the firm

[8] To test the robustness of our findings, we examine two alternative indicators of formalization commonly used in the literature: whether the firm exceeds an employment threshold of eight employees, and whether the firm reported making donations in the previous year. While neither measure is legally mandated, both reflect a firm's degree of formal engagement with institutional and public systems. Regression results using these alternative outcomes are reported in Table A4 in appendix and are consistent with our main findings. As raised by a reviewer, one of the five criteria used to construct the political connection indicator, whether the firm has a shareholder specifically designated to coordinate government relations, may not necessarily reflect privileged access to state resources. Instead, it could capture firms' attempts to cope with rent-seeking or administrative intervention, reflecting the "grabbing hand" of local governments. To assess the robustness of our results, we re-estimate the baseline regressions using an alternative measure of political connection that excludes this criterion. As shown in Table A5, the results remain highly consistent in sign and magnitude, suggesting that our main findings are not driven by this component. In Fig. A. 3, we also analyze the relationship between the employment formalization of SMEs and five disaggregated indicators that constitute the ***political connection*** variable through the same empirical method.

obtained a loan from the five major state-owned commercial banks, policy banks, or other formal financial institutions at the startup phase. The variable *gov contract* equals 1 if the firm received a government contract in 2017.[9]

Table 4 presents the results, controlling for city and industry fixed effects, as well as firm age effects and entrepreneur's demographic characteristics. Columns (1) and (3) show OLS estimates, indicating that firms whose entrepreneur had political connections were 8.4 and 7.9 percentage points more likely to obtain bank loans and government contracts, respectively. Both coefficients are statistically significant at the 1% level. Similarly, the logit estimates in Columns (2) and (4) confirm that political connections had a significant positive effect on the likelihood of obtaining bank loans and government contracts, with relative marginal effects of 0.081 and 0.078, respectively. These marginal effects are very close to the magnitude of OLS estimates, further reinforcing the robustness of the findings.

C. Heterogeneous analysis

To test the second theoretical proposition outlined in our conceptual framework that the premium associated with political connections diminishes in regions with more favorable business environments, we conduct two heterogeneity analyses.

In many developing and transition economies, access to the benefits of formalization remains highly uneven due to weak institutions and distorted resource allocation mechanisms. When formal market institutions are underdeveloped, private firms often turn to informal alternatives, with political connections serving as one of the most critical substitutes (e.g., Yu & Pan, 2008). In such environments, political connections can help firms bypass institutional frictions and obtain tangible economic benefits, such as access to credit and government contracts. Conversely, when

[9] Although the ESIEC (2018) survey includes questions on both current and startup-phase financing, only the startup module provides detailed distinctions between informal borrowing (from acquaintances) and formal lending sources. Therefore, we construct the "bank loan" variable based on startup-phase data, which offer greater granularity in loan source information. Moreover, firms are more likely to seek external financing during the startup phase, making this measure more informative for assessing access to formal credit.

institutional quality improves, the marginal value of political connections may decline.

Leveraging the substantial heterogeneity in business environments across the 62 cities in six provinces covered by the ESIEC data, we identify two key dimensions of the local business environment: First, we examine the development level of financial inclusion at the city level, which reflects how much credit is allocated through market-based financial institutions. In regions with higher financial inclusion, political connections may play a smaller role, as firms without such connections may still access bank loans, a key benefit of formalization, without relying on informal ties. We measure the development level of financial inclusion using the Peking University Digital Financial Inclusion Index of China (PKU_DFIIC). This index is compiled by the Institute of Digital Finance Peking University. Second, we assess the level of government intervention, captured by entrepreneurs' perception of the importance of government help for firm success. In regions where governments have a stronger hand in economic activities and market mechanisms are less dominant, political connections become more crucial role in accessing administratively allocated resources. ESIEC (2018) asked entrepreneurs about their views of government. Individuals were asked on scale of 1 to 10 how much they evaluate the influence of the help of government officials on their successfulness, where 10 denoted the most important and 1 denoted the least. We average these ratings at the city level to measure the degree of local government intervention in the economy. Cities are subsequently classified into high- and low-condition subsamples based on the median values of each business environment indices. Subgroup regressions are then conducted to re-estimate the baseline and mechanism models within each group.

Table 5 highlights that political connections play a more prominent role in regions with underdeveloped financial markets, where firms rely more heavily on informal ties to access low-cost loans from formal banks. Columns (1) and (2) show that in regions where financial inclusion was lower than the median level, firms with political connections were 12.5 percentage points more likely to obtain a bank loan and had a

formal employee share that was 8.5 percentage points higher than those without political connections. In comparison, in regions where financial inclusion was higher than the median level, while political connections still have a positive effect, their influence is weaker. As shown in columns (3) and (4), firms with political connections were 4.8 percentage points more likely to obtain a bank loan, and their formal employee share was 6.7 percentage points higher than those without political connections. Although both effects remain statistically significant, they are smaller in regions with higher financial inclusion, as formal firms in these areas can more easily access financial support without relying on informal ties, such as political connections.

Table 6 presents the results from dividing the sample based on whether the city where the firm was located fell above or below the median level of government intervention in the economy. As shown in columns (1) and (2), in regions with higher government intervention, firms with political connections were 9.0 percentage points more likely to obtain government contracts and had a formal employee share that was 8.9 percentage points higher than those without political connections. Both effects are statistically significant at the 1% level. In contrast, columns (3) and (4) reveal that in regions with lower government intervention, while political connections still have a positive effect, their impact is weaker. Firms with political connections had a 5.2 percentage point higher formalization rate (significant at the 10% level) and were 6.9 percentage points more likely to secure government contracts (significant at the 5% level). By examining how the accessibility of formalization benefits varies across regions, these two heterogeneity consistently suggest that political connections matter more when formal institutions fail to provide equal access to these benefits.[10]

[10] An alternative explanation posits that politically connected firms may provide better employee benefits, such as comprehensive social insurance and above-market wages, to discourage whistleblowing and conceal collusion. However, this rent-sharing mechanism does not predict systematic heterogeneity across institutional environments. In contrast, our heterogeneity results are consistent with the view that political ties are particularly valuable when formalization benefits are not equally accessible, thereby supporting the benefit-access channel proposed in this paper. Furthermore, we assess this alternative explanation by examining whether political connections are associated with higher levels of employee compensation through wage regressions. As shown in Table A6 in appendix, we find no consistent evidence that political connections are associated with higher wages across employee types, which does not support the rent-sharing hypothesis.

D. Further Analysis: Causality Discussion

While our research does not aim to test whether political connections have a causal effect on formalization, but rather emphasizes that the association between political connections and formalization benefits is central to understanding the firm formalization process, we nonetheless conducted a series of robustness checks to exclude the possibility that the positive relationship between connections and formalization is not robust due to omitted variable issues. Our baseline analysis reveals that politically connected entrepreneurs are more likely to provide formal employment benefits (e.g., social insurance) and exhibit higher formalization rates. However, potential endogeneity concerns persist: although we use politically connections measured at firm founding (prior to 2017 formalization outcomes) and control for demographic variables (proxying entrepreneurial ability), firm size, and city/industry fixed effects, two sources of bias may still affect the results—reverse causality (entrepreneurs of formalized firms have more chances to know government officials) and omitted variables correlated with unobserved entrepreneurial ability or endowments.

To further mitigate these concerns, we conduct two robustness checks. First, we employ an instrumental variable (IV) approach by using whether the entrepreneur had prior work experience in government agencies, state-owned enterprises, or public institutions before starting the business as the IV for political connections. Second, we use a propensity score matching (PSM) approach under counterfactual framework to estimate the average treatment effect on the treated (ATT), aiming to mitigate potential self-selection and confounding bias.

For the IV estimation, given our cross-sectional dataset, we employ a predetermined variable—prior work experience in government agencies, state-owned enterprises, or public institutions—as an instrumental variable (IV) for political connections. The validity of this IV is supported by two key criteria: (1) Relevance: Prior public-sector experience enhances entrepreneurs' opportunities and channels to establish political

connections, ensuring a strong correlation with the endogenous variable; (2) Exclusion Restriction: Since this experience occurred prior to business establishment, it is plausibly exogenous to formalization decisions, satisfying the exclusion condition.

Table 7 shows the regression results from the 2SLS. In column (1), the first-stage regression results indicate that entrepreneurs with prior work experience in the public sector are significantly more likely to have political connections at the time of firm founding. The F-statistic is 11.24, exceeding the conventional threshold of 10, suggesting that the instrument is relevant to the endogenous variable. In column (2)-(5), two-stage least squares (2SLS) estimates show that, although the statistical significance weakens somewhat, the overall direction of the effect remains consistent with the main results, and the findings lend suggestive support that our estimates are unlikely to be driven by reverse causality or omitted variable bias. Moreover, the estimated coefficient rises from 0.071 (Table 3, OLS) to 0.584 (Table 7, Column 2, IV regression), consistent with attenuation bias from measurement error in the baseline OLS specification. This pattern further indicates that the local average treatment effect (LATE) of political connection is substantially larger than the OLS estimates imply.

While this IV helps address reverse causality concerns, it does not fully resolve omitted variable biases. For instance, entrepreneurs with government backgrounds may systematically choose government-related or high-tech industries where formal employment is more prevalent—a potential confounding factor. To mitigate this, we control for city×industry fixed effects to account for unobserved heterogeneity at these levels. This specification yields estimates in Table A7 that are quantitatively similar to results in Table 7. Additionally, Fig. 6 presents supplementary evidence: connected and unconnected entrepreneurs exhibit no statistically significant differences in total factor productivity (TFP, measured by log Y/K), further supporting the exclusion restriction by suggesting no systematic pre-existing productivity differences between the two groups.

Besides, we augment the IV approach with propensity score matching (PSM) to

estimate a more precise average treatment effect (ATT). The consistent and significant estimates obtained from both the IV approach and the propensity score matching (PSM) method suggest a robust positive relationship between political connections and formalization. This convergent evidence enhances our confidence that the observed association is not merely an artifact of endogeneity.

Table 8 shows the results from the second robustness check. Since political connections are not randomly assigned across firms, the OLS estimates may suffer from self-selection and confounding bias if systematic differences exist between politically connected and unconnected firms. To address this, we apply the PSM method to construct a control group under the counterfactual framework by matching SMEs whose entrepreneurs did not have political connections with those in the treatment group that share similar observable characteristics. The average treatment effect on the treated (ATT) is then estimated as the average difference in the outcome variable between the matched treatment and control groups(Heinrich et al., 2010).

We start by estimating the propensity score, defined as the probability that an SME is led by a politically connected entrepreneur, using a logit model with a set of covariates. The major control variables are entrepreneur demographic characteristics that may influence the likelihood of forming political connections, including gender, education level, and self-assessed pre-business income. In addition, we include dummies for industries and cities to control the industrial and geographical fixed effect. Following common practice in the literature (Duan et al., 2022; Huang & Quan, 2025; Jiang & Guo, 2022), we first implement a one-to-one nearest neighbor matching algorithm to match firms in the treatment group (those whose entrepreneurs had political connections) with firms in the control group (those without such connections).

To assess the quality of matching, we conduct a set of standard diagnostics for covariate balance and common support, with results presented in the Appendix. As shown in Fig. A.4, the standardized mean differences across covariates are substantially reduced after matching, with nearly all post-matching values falling below the

commonly accepted 10% threshold. Table A3 further confirms that covariates are well balanced, as the post-matching differences in major covariates are small and statistically insignificant based on t-tests. In addition, we plot the kernel density distributions of propensity scores between politically connected and unconnected firms before and after matching in Fig. A. 5. The two groups exhibit significantly improved overlap after matching, indicating that the common support condition is satisfied and the matched samples are highly comparable. Overall, these diagnostics confirm that the matched sample is well balanced and suitable for estimating treatment effects.

Panel A in Table 8 reports the ATT estimated by one-to-one nearest neighbor matching algorithm. The results suggest that firms with politically connected entrepreneurs are 7.3 percentage points more likely to be intensive formal firms and have a 6.5 percentage point higher share of formal employees, relative to similar firms without such connections. In addition, they are 10.9 and 10.4 percentage points more likely to obtain bank loans and government contracts, respectively. Compared to the baseline regression results, the PSM results are consistent in direction, magnitude, and statistical significance. In panels B to E, we further test the robustness of the results using alternative matching algorithms, including k-nearest neighbor (with varying values of k), caliper matching, and kernel matching. Across all specifications, the estimated effects remain consistent with the baseline results.

Furthermore, the regional heterogeneity we observe aligns with logical intuition: better business environments (e.g., higher financial inclusion) correlate with stronger relationships between connections and formalization benefits, as well as higher social insurance contribution rates. Importantly, our analysis aims to explore the “causal” impact of environmental factors in reducing the connection between political connections and exclusive formalization benefits, thereby promoting formalization among unconnected firms. We provide descriptive evidence at the city level in Fig. 7, illustrating a positive correlation between the quality of the business environment and the average share of formal employees within firms. As shown in Panels A and B, firms

in cities with higher levels of financial inclusion or lower degrees of government intervention tend to cover a larger proportion of their employees with social insurance. This pattern aligns with the inference of our conceptual framework: by weakening the link between political connections and access to formalization benefits, a sound business environment can raise the incentive for employment formalization among firms, the vast majority of which lack political connections. This insight underscores that promoting equitable access to formalization benefits—rather than expanding political connections—is pivotal for addressing Chinese SMEs' avoidance of formal employment.

VI. Conclusion

Small and medium-sized enterprises (SMEs) play a vital role in fostering employment in developing countries. However, their transition toward employment formalization remains challenging, particularly in providing social insurance for their employees. Understanding the decision-making processes of SMEs regarding social insurance payments is essential for addressing broader challenges related to job quality and labor protection. This paper aims to describe the current state of formal employment among SMEs and analyze the constraints they commonly face in hiring formal employees.

Using unique survey data (ESIEC) collected through random sampling and field interviews in six representative provinces, we find that the average formal employee share in sample SMEs is 36.4 percent. Specifically, 51.6 percent of sample SMEs provide social insurance for at least one of their employees while 25.4 percent do so for all their employees. Compared to intensive formal firms, intensive informal firms exhibit characteristics similar to unregistered informal enterprises documented in previous literature, such as fewer employees, lower assets, and less educated entrepreneurs. These findings reveal that, despite government efforts to dramatically increase the number of formally registered businesses, many firms' behavior remains largely informal at the intensive margin. Additionally, we observe significant variation

across provinces and industries, with formal employees were more concentrated in better developed regions and in sectors with higher regulatory thresholds or lower labor intensity.

To explain why employment formalization among SMEs remains sluggish, we build on Ulyssea's (2020) formalization decision model and introduce entrepreneurs' political connections and business environment as key factors influencing firms' formalization process. We argue that access to formalization benefits is restricted, as formal credit resources and government contracts are not allocated purely through market mechanisms. Instead, entrepreneurs with political connections are more likely to secure these benefits, gaining disproportionate advantages from formalization. Furthermore, as the business environment improves, the reliance on political connections for accessing formalization benefits diminishes, reducing its premium for intensive formal firms.

Our regression analysis supports these propositions, highlighting the significant role political connections play in employment formalization. Firms without political connections are 7.1 percentage points less likely to provide formal employment, and the proportion of employees covered by social insurance decreases by 7.5 percentage points in these firms. The underlying mechanism reveals that firms without political connections have a lower probability of accessing formalization benefits. Specifically, these firms are 8.4 percentage points less likely to receive loans from state-owned or policy banks during the startup phase by, and 7.9 percentage points less likely to secure contracts from government departments or state-owned enterprises (SOEs) at the time of the survey. We further explore how the business environment influences the relationship between political connections and firm formalization through heterogeneity analyses. Our findings indicate that the impact of political connections on formalization is less pronounced in regions with better-developed financial inclusion and lower government intervention. Additionally, we assess the validity of our core argument by testing whether the cost of formalization affects the reliance on political

ties. We find that political connections are more influential in incentivizing formalization in industries where compliance imposes a greater financial burden on firms. To address potential endogeneity, we use instrumental variable (IV) and propensity score matching (PSM) methods. The robust test results are consistent with our baseline regressions. However, while these methods mitigate endogeneity, we cannot fully rule it out. Therefore, we refrain from interpreting our empirical findings as establishing a definitive causal relationship. Instead, we present suggestive evidence to support our theoretical argument.

In many developing countries, the benefits of formalization, such as access to bank loans or government contracts, are often limited and primarily available to those with political connections. As a result, operating formally and lawfully does not necessarily guarantee access to these resources, which has become a barrier preventing firms from providing more formal employment opportunities. Consequently, policies that prioritize the number of formalized enterprises as a key metric while neglecting the business environment may inadvertently encourage the registration of firms led by politically connected but less efficient entrepreneurs. This selection effect may distort resource allocation and exacerbate economic inefficiencies, ultimately undermining the broader objectives of formalization policies.

To foster meaningful formalization, policymakers should adopt a more comprehensive approach that accounts for both the extensive and intensive margins of formalization. Beyond improving formalization benefit and reducing its cost, efforts should also focus on enhancing the business environment and decoupling formalization benefits from political connections to ensure that all firms, particularly SMEs transitioning to formal status, can access these benefits. Establishing transparent and equitable government procurement processes can help create a level playing field, enabling SMEs without political connections to compete fairly for contracts. Expanding access to financial resources, such as by promoting inclusive financial systems and reducing reliance on political ties for obtaining bank loans, can further incentivize firms

to formalize. By ensuring that formalization benefits are equitably distributed and not dependent on political connections, policymakers can promote greater compliance with labor regulations and facilitate the transition of SMEs toward employment formalization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4o in order to check grammar and improve the language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Reference

Aga, G., Francis, D., & Rodríguez-Meza, J. (2015). SMEs, age, and jobs: A review of the literature, metrics, and evidence. *World Bank Policy Research Working Paper*, *7493*.

Barwick, P. J., Chen, L., Li, S., & Zhang, X. (2025). Entry deregulation, market turnover, and efficiency: China's business registration reform. *Review of Economics and Statistics*, 1–46.

Benhassine, N., McKenzie, D., Pouliquen, V., & Santini, M. (2018). Does Inducing Informal Firms to Formalize Make Sense? Experimental Evidence from Benin. *Journal of Public Economics*, *157*, 1–14.

Brogaard, J., Denes, M., & Duchin, R. (2020). Political Influence and the Renegotiation of Government Contracts. *The Review of Financial Studies*, *34*(6), 3095–3137.

Bruhn, M., & McKenzie, D. (2014). Entry regulation and the formalization of microenterprises in developing countries. *The World Bank Research Observer*, *29*(2), 186–201.

Cheng, H., Ding, S., & Liu, Y. (2024). The effectiveness of entry deregulation: Novel evidence from removing minimum capital requirements. *Journal of Development Economics*, *170*, 103304.

De Andrade, G. H., Bruhn, M., & McKenzie, D. (2016). A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms. *The World Bank Economic Review*, *30*(1), 24–54.

De Mel, S., McKenzie, D., & Woodruff, C. (2013). The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka. *American Economic Journal: Applied Economics*, *5*(2), 122–150.

De Soto, H. (1989). *The other path: The economic answer to terrorism*. Basic Books.

Dicko, S. (2016). Firms political connections and winning government contracts. *International Journal of Economics and Finance*, *8*(2), 19–32.

Djankov, S., Ganser, T., McLiesh, C., Ramalho, R., & Shleifer, A. (2010). The effect of corporate taxes on investment and entrepreneurship. *American Economic Journal: Macroeconomics*, *2*(3), 31–64.

Duan, Y., Zhang, H., Wang, W., & Ao, X. (2022). The effects of China's higher education expansion on urban and rural intergenerational mobility. *China Economic Review*, *73*, 101793.

Faccio, M. (2006). Politically Connected Firms. *The American Economic Review*, *96*(1), 369–386.

Faye, I., & Goldblum, D. (2022, June 14). Quest to better understand the relationship between SME finance and job creation: Insights from new report. *Quest to Better Understand the Relationship between SME Finance and Job Creation: Insights from New Report*. https://blogs.worldbank.org/en/psd/quest-better-understand-relationship-between-sme-finance-and-job-creation-insights-new-report#:~:text=SMEs%20account%20for%2090%20percent,by%20insufficient%20access%20to%20finance

Feng, J. (2013). The Incentive of Participation in Urban Social Insurance System in China: A Study Based on Manufacturing Firms. *Economic Research Journal*, *48*(7), 104-117 (in Chinese).

Floridi, A., Demena, B. A., & Wagner, N. (2020). Shedding light on the shadows of informality: A meta-analysis of formalization interventions targeted at informal firms. *Labour Economics*, *67*, 101925.

Goldman, E., Rocholl, J., & So, J. (2013). Politically Connected Boards of Directors and The Allocation of Procurement Contracts. *Review of Finance*, *17*(5), 1617–1648.

Heinrich, C., Maffioli, A., & Vázquez, G. (2010). *A primer for applying propensity-score matching* (Technical Note No. IDB-TN-161). Inter-American Development Bank (IDB).

Huang, T., & Quan, Y. (2025). Narrowing the digital divide: The growth and distributional effect of internet use on income in rural China. *China Economic Review*, *91*, 102387.

Jiang, S., & Guo, Y. (2022). Reasons for college major-job mismatch and labor market outcomes: Evidence from China. *China Economic Review*, *74*, 101822.

Khwaja, A. I., & Mian, A. (2005). Do lenders favor politically connected firms? Rent provision in an emerging financial market. *The Quarterly Journal of Economics*, *120*(4), 1371–1411.

Kok, J. de, & Berrios, M. (2019). *Small matters: Global evidence on the contribution to employment by the self-employed, micro-enterprises and SMEs*. ILO.

La Porta, R., & Shleifer, A. (2008). *The unofficial economy and economic development*. National Bureau of Economic Research.

La Porta, R., & Shleifer, A. (2014). Informality and Development. *Journal of Economic Perspectives*, *28*(3), 109–126.

Li, H., Meng, L., Wang, Q., & Zhou, L.-A. (2008). Political connections, financing and firm performance: Evidence from Chinese private firms. *Journal of Development Economics*, *87*(2), 283–299.

Lin, J. Y., & Li, Y. (2001). Promoting the growth of medium and small-sized enterprises through the development of medium and small-sized financial institutions. *Economic Research Journal*, *1*, 10-18 (in Chinese).

Luo, D., & Tang, Q. (2009). Political Connections, Social Capital, and Access to Policy Resources: Empirical Evidence

from China's Privately Listed Firms. *The Journal of World Economy*, *7*, 84-96 (in Chinese).

Luo, D., & Zhen, L. (2008). Private control, political relationship and financing constrain of private listed enterprises. *Journal of Financial Research*, *12*(2), 164-78 (in Chinese).

McKenzie, D., & Sakho, Y. S. (2010). Does it pay firms to register for taxes? The impact of formality on firm profitability. *Journal of Development Economics*, *91*(1), 15–24.

MOHRSS. (2024). *2023 Annual Statistical Bulletin on the Development of Human Resources and Social Security Undertakings*. Ministry of Human Resources and Social Security of the People's Republic of China. https://www.mohrss.gov.cn/xxgk2020/fdzdgknr/ghtj/tj/ndtj/202406/t20240617_520366.html

Nyland, C., Smyth, R., & Zhu, C. J. (2006). What determines the extent to which employers will comply with their social security obligations? Evidence from Chinese firm-level data. *Social Policy & Administration*, *40*(2), 196–214.

Nyland, C., Thomson, S. B., & Zhu, C. J. (2011). Employer attitudes towards social insurance compliance in Shanghai, China. *International Social Security Review*, *64*(4), 73–98.

Perry, G. E., Arias, O., Fajnzylber, P., Maloney, W. F., Mason, A., & Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion*. The World Bank.

Rocha, R., Ulyssea, G., & Rachter, L. (2018). Do lower taxes reduce informality? Evidence from Brazil. *Journal of Development Economics*, *134*, 28–49.

Shen, G., & Zou, J. (2017). Firm Size, Government-Business Relations, and Effective Tax Rates: Evidence from the World Bank's "Investment Climate Survey." *Journal of Management World*, *6*, 23-36 (in Chinese).

Song, H., Feng, J., & Yang, W. (2021). The Effect of a Reduction of the Social Security Contribution Rate on Enterprise Social Security Participation and labor employment. *Economic Research Journal*, *56*(1), 90-104 (in Chinese).

Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, *108*(8), 2015–2047.

Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, *12*(1), 525–546. https://doi.org/10.1146/annurev-economics-082119-121914

World Bank. (2013). *Doing Business 2014: Understanding Regulations for Small and Medium-Size Enterprises*. The World Bank Group.

World Bank. (2020). *Doing Business 2020*. The World Bank Group.

Xu, L., Wang, R., Wang, R., Li, B., Tian, C., & Zhang, X. (2023). The Development of Small and Medium-sized Enterprises in China: Data and New Evidence. *Economic Science*, *1*, 5-26 (in Chinese).

Yu, M., & Pan, H. (2008). Political Connections, Institutional Environment, and Bank Loans for Private Enterprises. *Journal of Management World*, *8*, 9-21 (in Chinese).

Yu, W., Wang, M., & Jin, X. (2012). Political connection and financing constraints: Information effect and resource effect. *Economic Research Journal*, *9*(2012), 125-139 (in Chinese).

Zhao, J., Mao, J., & Zhang, L. (2016). Social Security Contribution Rate, Participation Probability and Contribution

Level: An Empirical Study on Contribution Evasion of Employees and Firms. *China Economic Quarterly*, *15*(1), 341-372 (in Chinese).

Tables and Figures

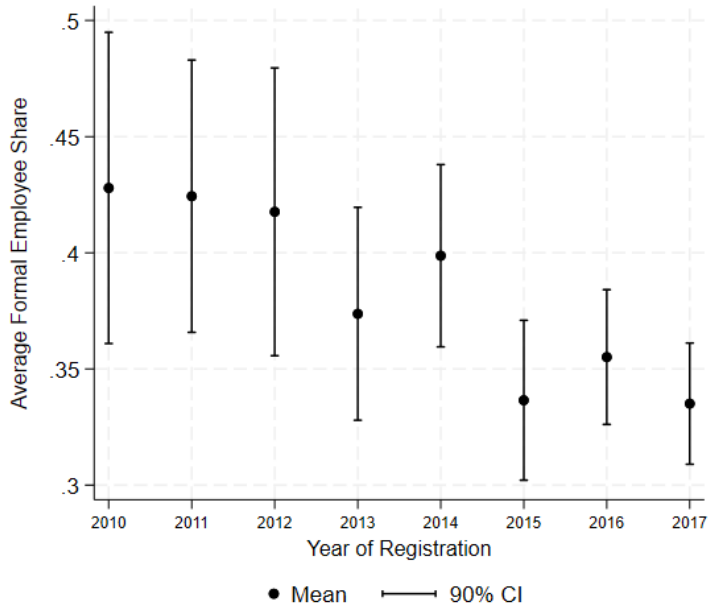


Fig. 1 The Average Formal Employee Share within SMEs, by Year of Registration

Note: Based on data from the ESIEC (2018), this figure illustrates the trend in the average proportion of formal employees within SMEs over the registration time in China. Formal employee are defined as those covered by social security. Note firms registered after a nationwide implementation of the commercial system reform since 2013 exhibit a significant lower formal employee share. A 90% confidence interval is plotted as vertical lines.

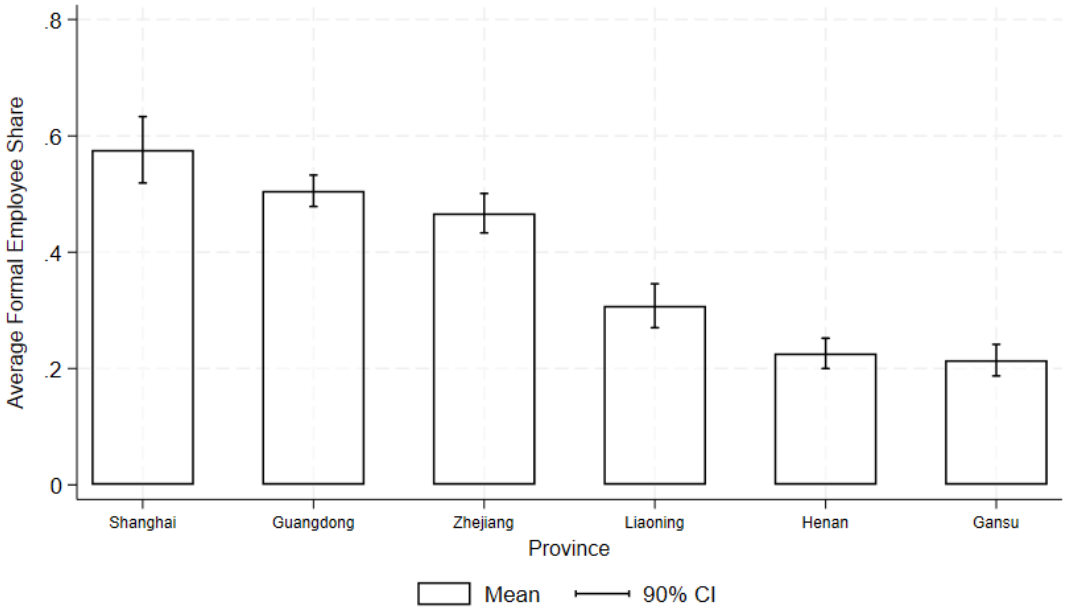


Fig. 2 The Average Formal Employee Share within SMEs, by Province

Note: Based on data from the ESIEC (2018), this figure illustrates the average proportion of formal employees of six provinces in our sample. Formal employees are defined as those covered by social security. A 90% confidence interval is plotted as vertical lines.

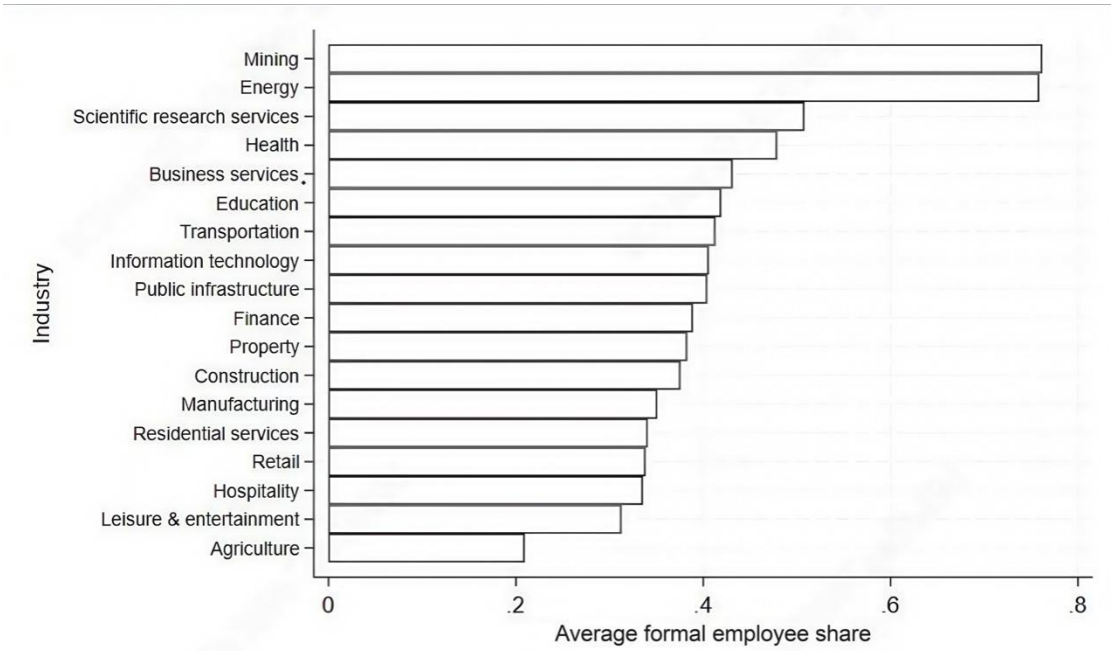
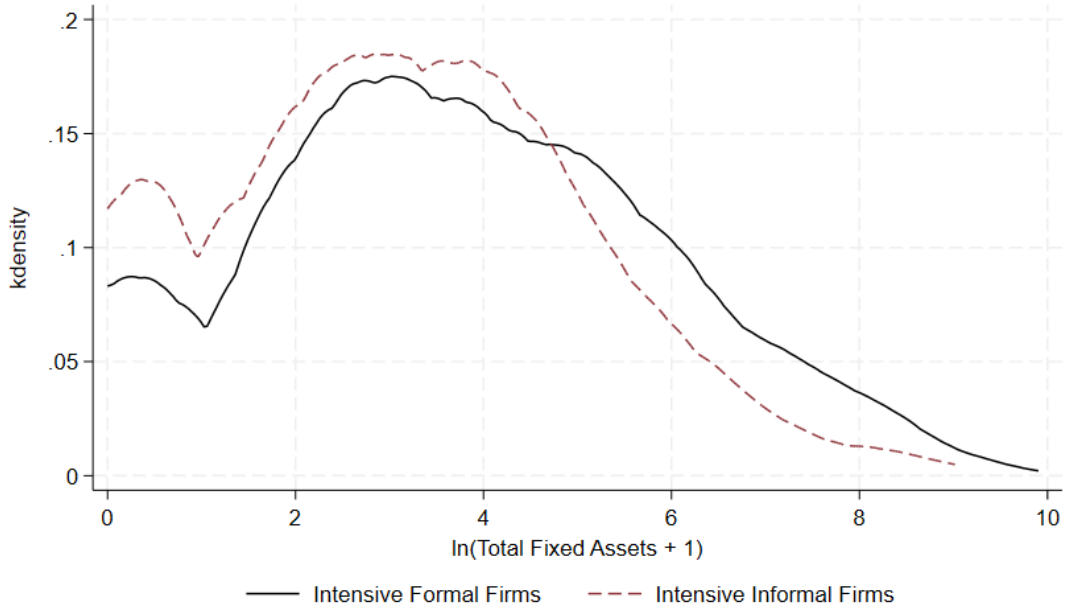
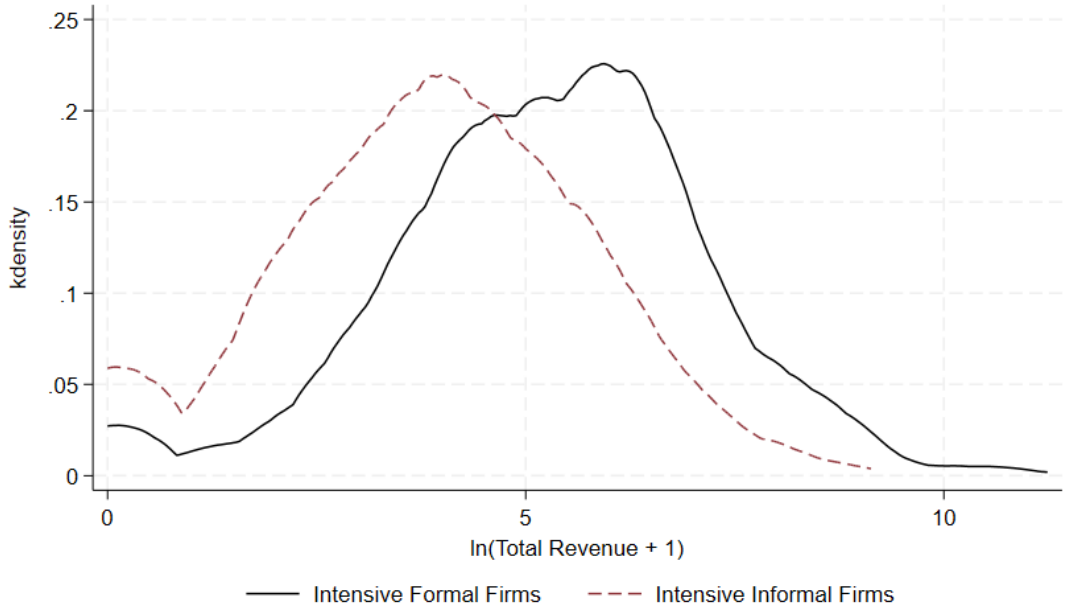


Fig. 3 The Average Formal Employee Share within SMEs, by Industry

Note: Based on data from the ESIEC (2018), this figure illustrates the average proportion of formal employees within SMEs across 18 industries in China. The vertical axis lists 18 industries at the SIC one-digit level, ranked from highest to lowest proportion, while the horizontal axis shows the average formal employee share of each industry in our sample. Formal employees are defined as those covered by social security.



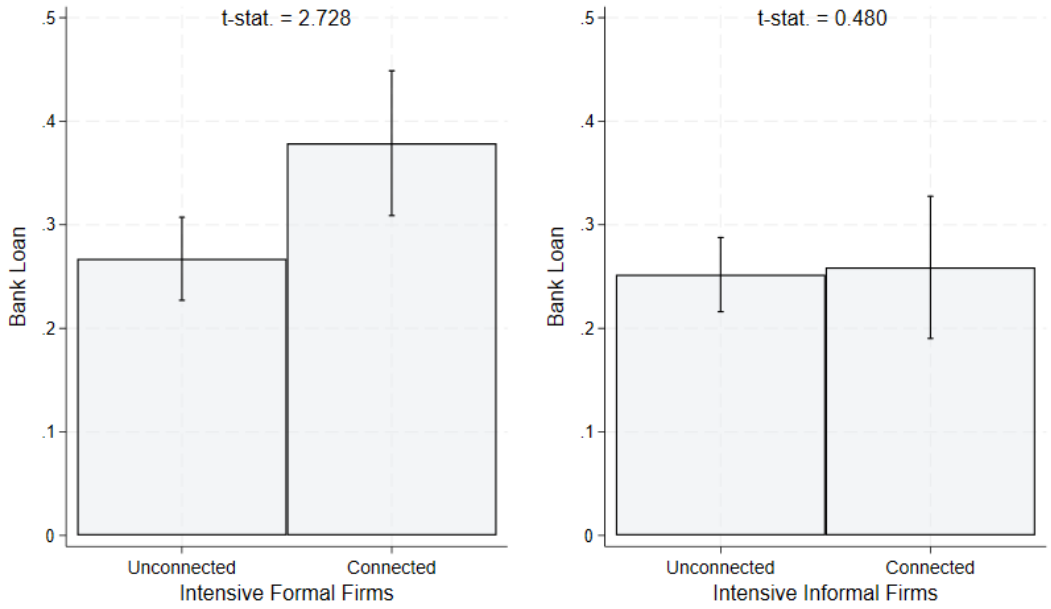
Panel A. Total Fixed Assets



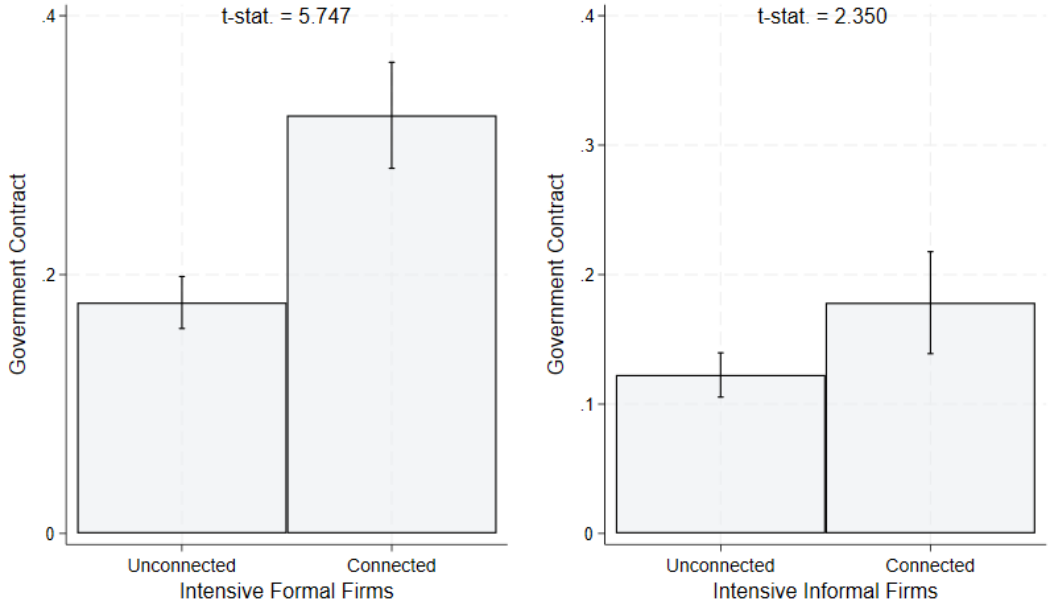
Panel B. Total Revenue

Fig. 4 Distribution of Firm Size for Intensive Formal and Informal Firms

Note: The Authors' calculations are based on ESIEC(2018). Panel A and B show the distribution of intensive formal and informal firms in terms of total fixed asset and total revenue size in 2017. A firm was an intensive informal firm if it did not employ any formal employee. A firm was an intensive formal firm if it employed at least one formal employee. Formal employees are defined as those covered by social security.



Panel A. Bank Loan



Panel B. Government Contract

Fig. 5 Differential Access to Formalization Benefits by Political Connection and Formal Status

Note: Based on data from the ESIEC (2018), this figure presents the difference in the probability of obtaining formalization benefits between connected and unconnected firms, separately for intensive formal and informal firms. F-statistics are derived from mean-comparison t-tests within each firm type, with 90% confidence intervals. A firm is classified as an intensive informal firm if it employed no formal employees, and as an intensive formal firm if it employed at least one formal employee. Formal employees

are defined as those covered by social security. A firm is classified as connected if its entrepreneur had political connections at the startup phase, and as unconnected otherwise. Bank loan is a binary variable indicating whether a firm received a loan from one of the five major state-owned commercial banks, policy banks, or other formal financial institutions in 2017, defined only for firms that sought finance. Government Contract is a binary variable indicating whether a firm received a government contract in 2017.

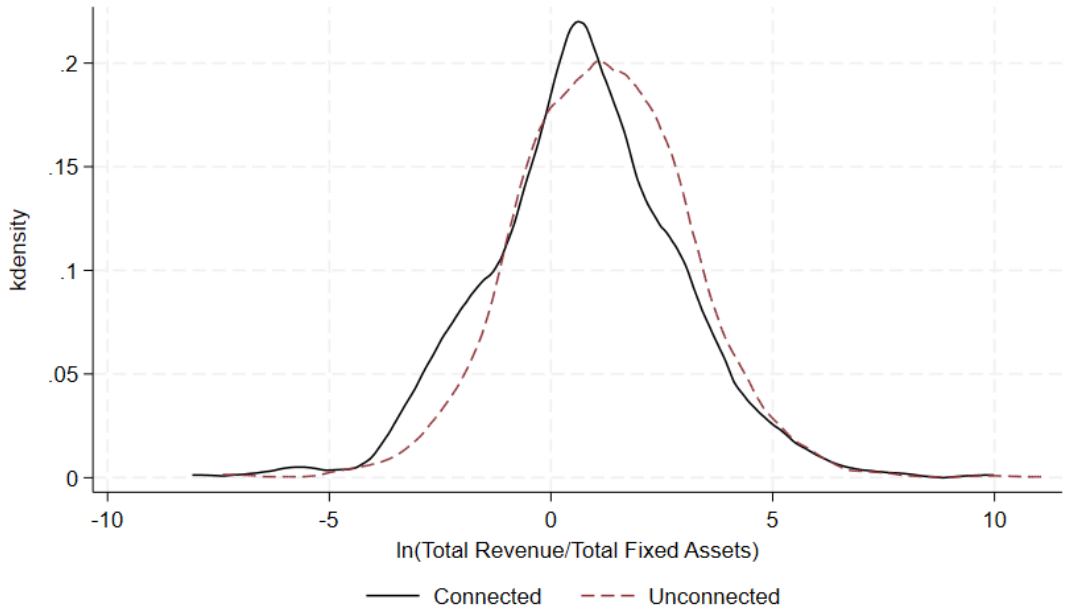
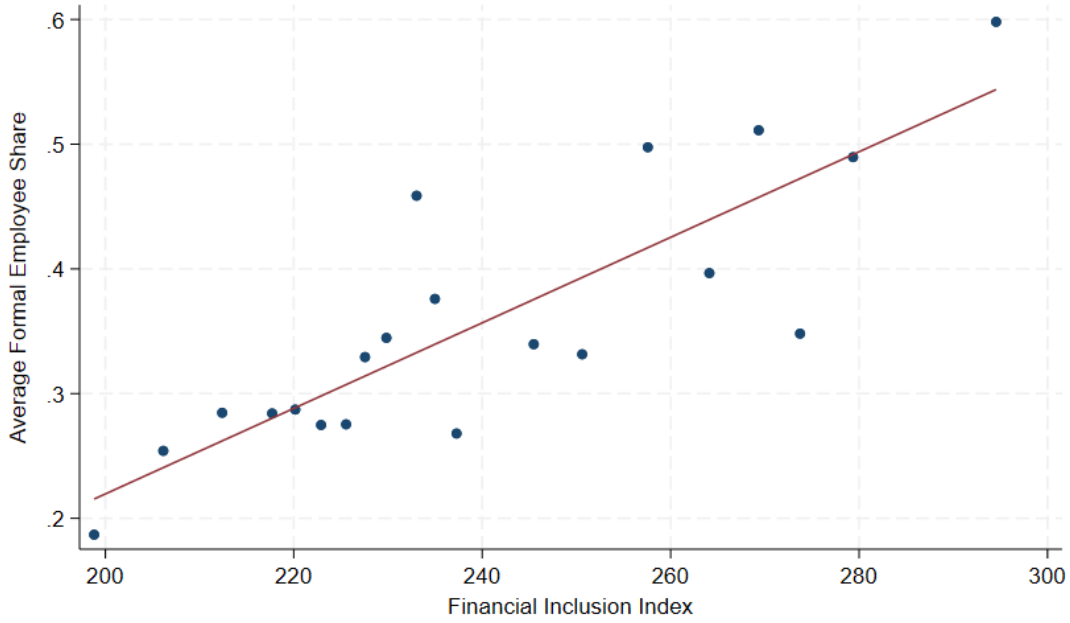
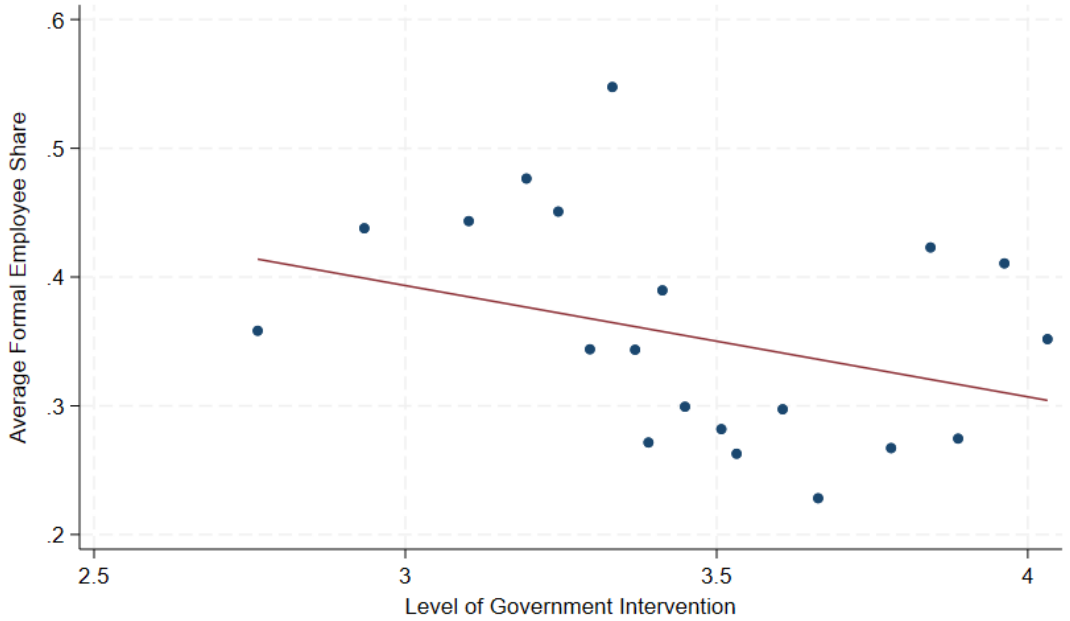


Fig. 6 Distribution of Total Factor Productivity for Connected and Unconnected Firms

Note: The Authors' calculations are based on ESIEC (2018). This figure presents the kernel density distributions of total factor productivity (TFP) for connected and unconnected firms. TFP is measured by log(Y/K), where Y is proxied by total revenue and K is proxied by total fixed assets. A firm is classified as connected if its entrepreneur had political connections at the startup phase, and as unconnected otherwise.



Panel A. Financial inclusion index



Panel B. Level of government intervention

Fig. 7 Business Environment and Average Formal Employee Share

Note: This figure presents binned scatterplots based on city-level data from the ESIEC (2018), illustrating the correlation between the quality of the local business environment and the average share of formal employees within firms. Formal employees are defined as those covered by social insurance. Panel A examines the dimension of financial development, measured by the Peking University Digital Financial Inclusion Index of China (PKU_DFIIC). Panel B examines the level of

government intervention, captured by an index constructed from entrepreneurs' perceptions regarding the importance of government help for firm success.

Table 1: Summary Statistics

Variable	Definition	Obs.	Mean	SD
intensive formal firm	=1 if a firm employed at least one formal employee, =0 if a firm did not hire formal employee in 2017	2664	0.516	0.500
formal employee share	the proportion of formal employees within the firm in 2017; =0, if a firm did not have any formal employee	2664	0.364	0.426
bank loan	=1 if the firm received a loan from one of the five major state-owned commercial banks, policy banks, or other formal financial institutions at the startup phase, and =0 otherwise	2634	0.200	0.400
gov contract	=1 if received a government contract in 2017, and 0 otherwise	2610	0.177	0.381
political connection	=1 if a firm met any of the 5 criteria bellow, =0 otherwise	2664	0.236	0.425
gov_acquaintance	=1 if had acquaintances working in government at inception, =0 otherwise	2664	0.145	0.352
gov_guided	=1 if firm was established under the guiding of the government, =0 otherwise	2664	0.046	0.209
gov_subsidy	=1 if received government subsidies or services at inception, =0 otherwise	2664	0.079	0.269
shareholder_coord	=1 if existed shareholders specifically designated to coordinate with the government, =0 otherwise	2664	0.017	0.128
shareholder_link	=1 if had established equity relationships with government, =0 otherwise	2664	0.0004	0.019
college	=1, if entrepreneur had received education at the junior college level or above, =0 otherwise	2664	0.450	0.498
male	=1 if the entrepreneur interviewed was a male	2664	0.778	0.416
age of firm	lifespan of the firm calculated as the time from the registration to the time of the survey	2664	3.110	2.030
pre-business income	entrepreneur's self-assessed pre-entrepreneurial income	2664	2.599	1.016

	relative to local standards, ranging from 1 to 5			
priorgovjob	=1 if the entrepreneur previously held a job in the public sector, including positions in government agencies, state-owned enterprises, or public institutions	2664	0.233	0.423
employees	number of the full-time employees in 2017	2664	15.326	39.212
asset	total fixed assets in 2017	2664	259.067	937.044
revenue	total revenue in 2017	2664	609.739	2875.040
gross profit rate	gross profit rate in 2017	2664	23.333	18.276
technician share	the proportion of technicians in the full-time employees	2131	0.426	0.363
college employee share	the proportion of employees with junior college education or above	2131	0.367	0.386
Cities with higher financial inclusion index	=1 if the firm was located in a city with financial inclusion level higher than the median, =0 otherwise	2664	0.594	0.491
Cities with higher level of government intervention	=1 if firm was located in a city with government intervention higher than the median, =0 otherwise	2664	0.533	0.499
Cities with higher fairness index	=1 if firm was located in a city with fairness index higher than the median, =0 otherwise	2664	0.485	0.500
Industries with higher labor cost share	=1 if firm operated in an industry with an average labor cost share higher than the median, =0 otherwise	2664	0.501	0.500

Note: The authors calculate the data from the ESIEC (2018).

Table 2: Characteristics of Intensive Informal and Intensive Formal Firms

Variables	(1) Obs.	(2) Intensive informal firms	(3) Obs.	(4) intensive formal firms	(2)-(4) Mean Diff
college	1,290	0.340	1,374	0.553	-0.213***
male	1,290	0.777	1,374	0.779	-0.002
pre-business income	1,290	2.507	1,374	2.685	-0.178***
age of firm	1,290	2.938	1,374	3.271	-0.333***
total fixed asset	1,290	158.305	1,374	353.670	-195.365***
total revenue	1,290	224.2	1,374	971.7	-747.571***
gross profit rate	1,290	24.11	1,374	22.61	1.498**
employees	1,290	7.677	1,374	22.51	-14.830***
technician share	859	0.403	1,272	0.441	-0.038**
college employee share	859	0.282	1,272	0.424	-0.142***

Note: The authors calculate the data from the ESIEC (2018). Column (5) presents the differences in entrepreneurs' personal characteristics and firms' operation performance in 2017 between intensive formal and intensive informal firms. A firm was an intensive informal firm if it did not employ any formal employee. A firm was an intensive formal firm if it employed at least one formal employee. Formal employees are defined as those covered by social insurance. The symbols *, ** and *** denote significance levels at the 10%, 5%, and 1% levels, respectively.

Table 3: OLS Regressions of Political Connection on Employment Formalization of SMEs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	OLS	OLS	OLS	Logit	OLS	OLS	OLS	Tobit
Dependent variable:	*intensive formal firm* (dummy)				*formal employees share* (continuous)			
political connection	0.078***	0.092***	0.071***	0.361***	0.082***	0.094***	0.075***	0.132***
	(0.023)	(0.021)	(0.021)	(0.109)	(0.020)	(0.019)	(0.019)	(0.033)
male			0.009	0.054			-0.001	0.003
			(0.022)	(0.108)			(0.019)	(0.035)
college			0.148***	0.723***			0.126***	0.238***
			(0.020)	(0.096)			(0.017)	(0.030)
pre-business income			0.021**	0.112**			0.026***	0.046***
			(0.009)	(0.045)			(0.008)	(0.014)
age of firm			0.036**	0.181**			0.024	0.047*
			(0.017)	(0.088)			(0.015)	(0.027)
age square of firm			-0.002	-0.011			-0.001	-0.003
			(0.002)	(0.011)			(0.002)	(0.003)
Mean of dependent variable	0.516	0.516	0.516	0.516	0.364	0.364	0.364	0.364
City Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Industry Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Relative/Average Marginal Effect				0.071				0.073
Observations	2,664	2,664	2,664	2,661	2,664	2,664	2,664	2,664

Note: In Columns (1)-(4), the dependent variable is a dummy that equals 1 if the firm is an intensive formal firm that hire at least one formal employee, and equals 0 otherwise. Columns (5)-(8) use a continuous dependent variable representing the proportion of formal employees within the firm. Columns (1) and (5) only perform a simple linear OLS regressions, while all other columns include city and industry fixed effects. Columns (3) and (7) further control for entrepreneurs' demographic variables,

such as gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. Columns (4) and (8) employ a Logit model and a Tobit model, respectively, to better account for the distributional features of the dependent variables. Due to perfect prediction in the logit model, some observations are omitted, resulting in fewer observations in Logit compared to the OLS estimation. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table 4: Regressions of Political Connection on Bank Loan and Government Contract

	(1)	(2)	(3)	(4)
Method:	OLS	Logit	OLS	Logit
Dependent variable:	*bank loan*		*gov contract*	
political connection	0.083***	0.527***	0.079***	0.567***
	(0.020)	(0.123)	(0.020)	(0.125)
male	-0.008	-0.053	0.056***	0.491***
	(0.018)	(0.129)	(0.016)	(0.148)
college	-0.051***	-0.356***	0.048***	0.371***
	(0.016)	(0.119)	(0.017)	(0.125)
pre-business income	0.011	0.071	0.010	0.098*
	(0.008)	(0.054)	(0.007)	(0.059)
age of firm	0.001	0.002	0.027*	0.226**
	(0.015)	(0.104)	(0.014)	(0.113)
age square of firm	-0.001	-0.005	-0.002	-0.020
	(0.002)	(0.012)	(0.002)	(0.014)
Mean of dependent variable	0.200	0.200	0.177	0.177
City Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Relative Marginal Effect		0.080		0.078
Observations	2,634	2,561	2,610	2,546

Note: In columns (1) and (2), the dependent variable is a dummy variable that equal to one if the firm obtained a loan from one of the five major state-owned commercial banks, policy banks, or other formal financial institutions at the startup phase, and zero otherwise. In columns (3) and (4), the dependent variable, is a dummy variable that equal to one if the firm secured a government contract in 2017, and zero otherwise. For each dependent variable, we conduct both OLS and Logit estimations. Due to perfect prediction in the logit model, some observations are omitted, resulting in fewer observations in Logit compared to the OLS estimation. All regressions include city and industry fixed effects. Additionally, we control for entrepreneurs' demographic variables, such as gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table 5: Heterogeneity Analysis Based on the Development of Financial Inclusion

	(1)	(2)	(3)	(4)
Locations of firms	Cities with lower financial inclusion index		Cities with higher financial inclusion index	
Dependent variable:	*formal employee share*	*bank loan*	*formal employee share*	*bank loan*
political connection	0.085***	0.117***	0.067***	0.054**
	(0.029)	(0.032)	(0.025)	(0.024)
male	-0.061**	-0.091***	0.039	0.045**
	(0.030)	(0.033)	(0.025)	(0.020)
college	0.071***	-0.105***	0.163***	-0.013
	(0.027)	(0.028)	(0.022)	(0.020)
pre-business income	0.016	0.026	0.032***	-0.0001
	(0.012)	(0.013)	(0.011)	(0.009)
age of firm	0.002	-0.013	0.041**	0.013
	(0.023)	(0.026)	(0.019)	(0.017)
age square of firm	0.001	0.004	-0.003	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)
Mean of dependent variable	0.268	0.267	0.430	0.154
City Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,082	1,072	1,582	1,562

Note: Columns (1) and (3) use a continuous dependent variable representing the proportion of formal employees within the firm. Columns (2) and (4) use a dummy dependent variable that equal to one if the firm obtained a loan from one of the five major state-owned commercial banks, policy banks, or other formal financial institutions at the startup phase, and zero otherwise. All cities are classified into two groups based on the median value of the financial inclusion index (PKU_DFIIC). Columns (1) and (2) include firms located in cities where the financial inclusion index is higher than the median, while Columns (3) and (4) include firms located in cities where the financial inclusion index is lower than the median. All regressions are estimated using the OLS method, including city and industry fixed effects. Additionally, we control for entrepreneurs' demographic variables, such as gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table 6: Heterogeneity Analysis Based on the Level of Government Intervention

	(1)	(2)	(3)	(4)
Locations of firms	Cities with higher level of government intervention		Cities with lower level of government intervention	
Dependent variable:	*formal employee share*	*gov contract*	*formal employee share*	*gov contract*
political connection	0.089***	0.090***	0.052*	0.069**
	(0.025)	(0.026)	(0.029)	(0.029)
male	0.013	0.048**	-0.026	0.066***
	(0.026)	(0.024)	(0.028)	(0.022)
college	0.139***	0.043*	0.108***	0.051**
	(0.023)	(0.024)	(0.025)	(0.023)
pre-business income	0.031***	0.005	0.021*	0.019*
	(0.011)	(0.011)	(0.012)	(0.010)
age of firm	-0.003	0.030	0.056**	0.031
	(0.021)	(0.020)	(0.022)	(0.021)
age square of firm	0.001	-0.003	-0.004*	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Mean of dependent variable	0.328	0.177	0.405	0.176
City Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,421	1,392	1,243	1,218

Note: Columns (1) and (3) use a continuous dependent variable representing the proportion of formal employees within the firm. Columns (2) and (4) use a dummy dependent variable that equal to one if the firm received a government contract in 2017, and zero otherwise. All cities are classified into two groups based on the degree of local government intervention. Columns (1) and (2) include firms located in cities where the level of government intervention is higher than the median, while Columns (3) and (4) include firms located in cities where the level of government intervention is lower than the median. All regressions are estimated using the OLS method, including city and industry fixed effects. Additionally, we control for entrepreneurs' demographic variables, such as gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table 7: IV Regressions of Political Connection on Employment Formalization of SMEs

	（1）	(2)	(3)	(4)	（5）
Model	First stage	2SLS	2SLS	2SLS	2SLS
Dependent variable:	*political connection*	*intensive formal firm* (dummy)	*formal employees share* (continuous)	*bank loan*	*gov contract*
priorgovjob	0.069*** (0.020)				
political connection		0.584* (0.347)	0.681** (0.329)	0.520* (0.293)	0.635* (0.359)
male	0.030 (0.019)	-0.007 (0.026)	-0.020 (0.024)	-0.021 （0.022）	0.037 （0.023）
college	0.049*** (0.018)	0.118*** (0.029)	0.091*** (0.028)	-0.076*** （0.024）	0.016 （0.029）
pre-business income	0.039*** (0.008)	0.001 0.017	0.002 0.016	-0.006 （0.015）	-0.012 （0.017）
age of firm	0.021 (0.016)	0.025 (0.021)	0.011 (0.019)	-0.008 (0.017)	0.015 （0.018）
age square of firm	-0.002 (0.002)	-0.001 (0.002)	-0.0002 (0.002)	-0.001 （0.002）	-0.002 （0.002）
Mean of dependent variable	0.236	0.516	0.364	0.200	0.177
City Fixed Effects	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,664	2,664	2,664	2,634	2,610
F-Statistics	11.24				

Note: Column (1) presents the first-stage regression results, where the instrument is a dummy variable indicating whether the entrepreneur previously held a job in the public

sector (*priorgovjob*), including positions in government agencies, state-owned enterprises, or public institutions. Columns (2)-(5) report the two-stage least squares (2SLS) estimates for the following dependent variables: *intensive formal firm* (dummy), *formal employees share* (continuous), *bank loan*, and *government contract*. All specifications include city and industry fixed effects and control for entrepreneurs' demographic variables, including gender, education (measured by whether they had a junior college degree or higher), their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table 8 ATT Estimates from Propensity Score Matching (PSM)

	(1)	(2)	(3)	(4)
	intensive formal firm	*formal employees share*	*bank loan*	*gov contract*
A. one-to-one nearest neighbor matching				
ATT	0.073**	0.065**	0.105***	0.104***
S.E.	（0.029）	（0.033）	（0.027）	（0.026）
T-stat	2.24	2.25	3.85	4.04
observation		*2,554*		
B. k-Nearest Neighbor Matching (k=3)				
ATT	0.073***	0.063**	0.105***	0.090***
S.E.	(0.028)	(0.025)	（0.024）	（0.023）
T-stat	2.62	2.58	4.45	3.90
observation		*2,554*		
C. k-Nearest Neighbor Matching (k=5)				
ATT	0.079***	0.068***	0.105***	0.088***
S.E.	(0.027)	(0.024)	（0.023）	（0.022）
T-stat	2.95	2.88	4.58	3.98
observation		*2,554*		
D. caliper matching				
ATT	0.749**	0.052*	0.085***	0.082***
S.E.	(0.033)	(0.029)	(0.027)	(0.026)
T-stat	2.27	1.80	3.11	3.14
observation		2,488		
E. kernel matching				
ATT	0.079***	0.075***	0.093***	0.082***

S.E.	(0.025)	(0.022)	(0.021)	(0.021)
T-stat	3.21	3.48	4.38	4.00
observation		*2,554*		

Note: This table reports the estimated average treatment effect on the treated (ATT) for four outcome variables: whether the firm has at least one formal employee (*intensive formal firm*), the share of formal employees (*formal employee share*), whether the firm obtained a bank loan (*bank loan*), and whether it secured a government contract (*gov contract*). Estimates are presented for five matching algorithms: one-to-one nearest neighbor matching (Panel A), k-nearest neighbor matching with k=3 (Panel B) and k=5 (Panel C), caliper matching (Panel D), and kernel matching (Panel E). All matching procedures use political connection as the treatment variable. Each specification controls for pre-treatment covariates and includes industry and city dummies in the matching process. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Appendix A

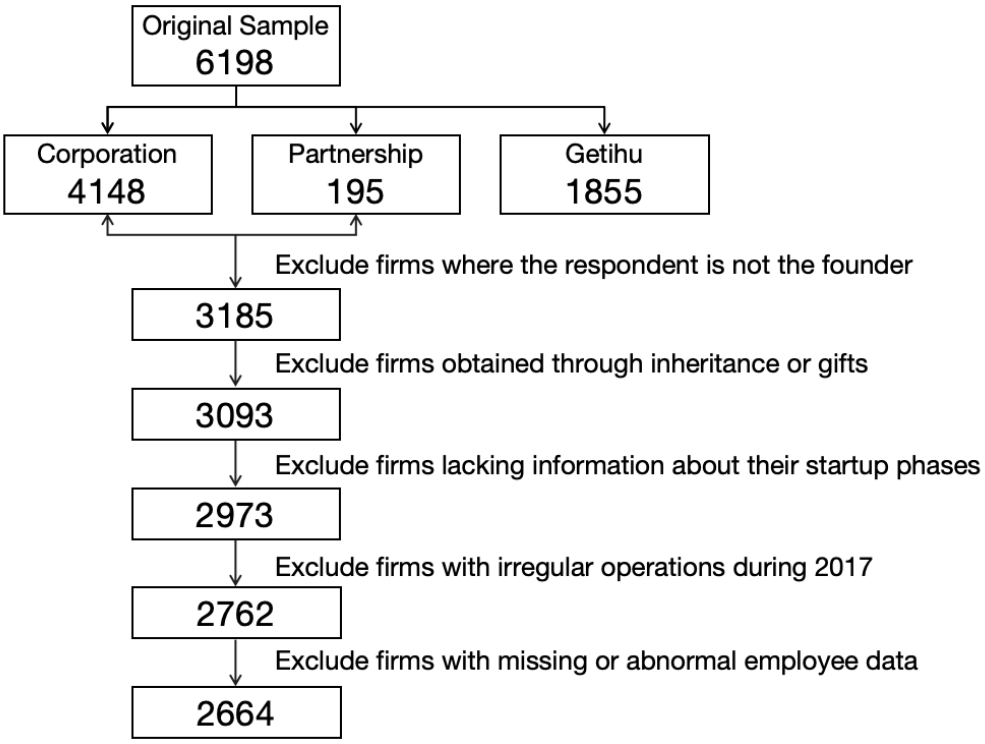


Fig. A. 1 Sample Construction Process.

Note: This figure illustrates the sample construction process used in this study. The numbers inside the boxes represent the number of firms retained at each stage. The ESIEC (2018) dataset initially comprised 6,198 completed interviews. First, Getihu (self-employed businesses) were excluded, as their operations resemble individual or family-run enterprises with minimal employee hiring. Moreover, 93.61% Getihu reported no political connections, and including them could potentially overestimate the correlation between employment formalization and political connections. This exclusion reduced the sample to 4,343 firms. Next, we retained only firms where the respondent was directly involved in the firm's creation as a shareholder or partner (3,185 firms left), followed by excluding those obtained through inheritance or gifts (3,093 firms left). Firms lacking startup-phase information and those reporting irregular operations in 2017 were also removed, as such cases would lead to missing key analytical variables, leading to 2,762 firms retained. Lastly, firms with missing or abnormal employee data were excluded, resulting in a final sample of 2,664 firms.

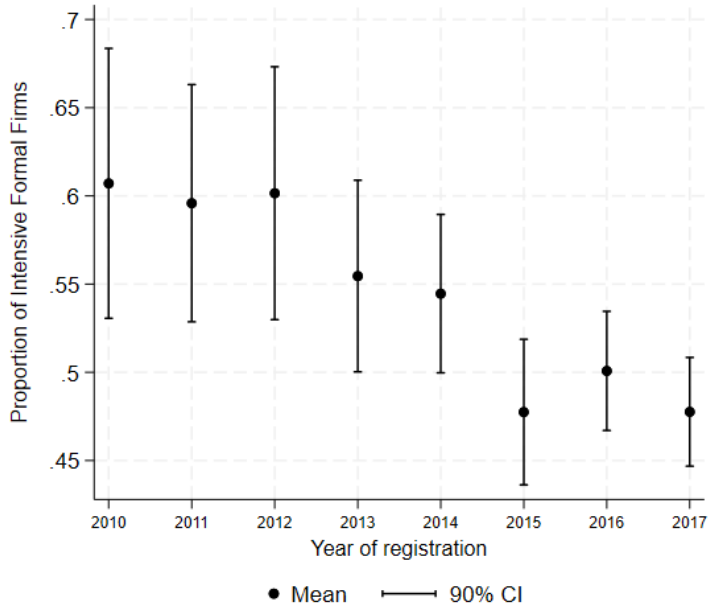


Fig. A. 2 The Proportion of Intensive Formal Firms, by Year of Registration

Note: Based on data from the ESIEC (2018), this figure illustrates the trend in the proportion of intensive formal firms over the firms' registration time in China. A firm is classified to an intensive formal firm if it employed at least one formal employee in 2017. Formal employees are defined as those covered by social insurance. Note firms registered after a nationwide implementation of the commercial system reform since 2013 exhibit a significant lower likelihood of being intensive formal. A 90% confidence interval is plotted as vertical lines.

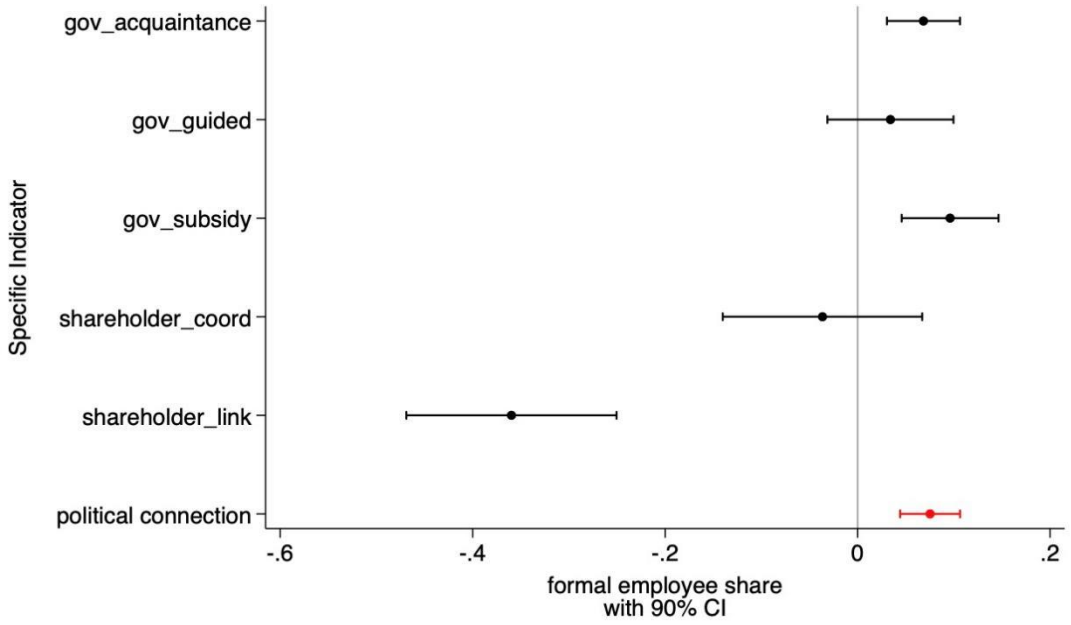


Fig. A. 3 OLS Estimates of Formal Employment on Disaggregated Indicators of Political Connections

Note: This figure analyzes the relationship between the employment formalization of SMEs and five disaggregated indicators that constitute the political connection, which serves as core explanatory in the main regression. Additionally, we present the baseline regression estimate for comparison. The Y-axis lists the variable political connection along with its five disaggregated components, all of which are dummy variables derived from the ESIEC (2018) and capture firm's political ties through different dimensions. gov_acquaintance equals 1 if the entrepreneur had acquaintances working in the government at the time of the firm's establishment; gov_guided equals 1 if firm was established under the guiding of the government; gov_subsidy equals 1 if the firm received government subsidies or services at inception; shareholder_coord equals 1 if there existed shareholders in the firm specifically designated to coordinate with the government; shareholder_link equals 1 if the firm had established equity relationships with government; political connection at the bottom of the Y-axis, equals 1 if any of the five indicators listed above equals 1, categorizing the firm as having political connection. The X-axis presents the OLS estimates for these variables in relation to the proportion of formal employees within the firm. Formal employees are defined as those covered by social insurance. All regressions are estimated using the OLS method, including city and industry fixed effects. Additionally, we control for entrepreneurs' demographic variables, such as gender, education (measured by whether they had a junior college degree or higher), and their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. Horizontal lines indicate the 90% confidence intervals for each estimate.

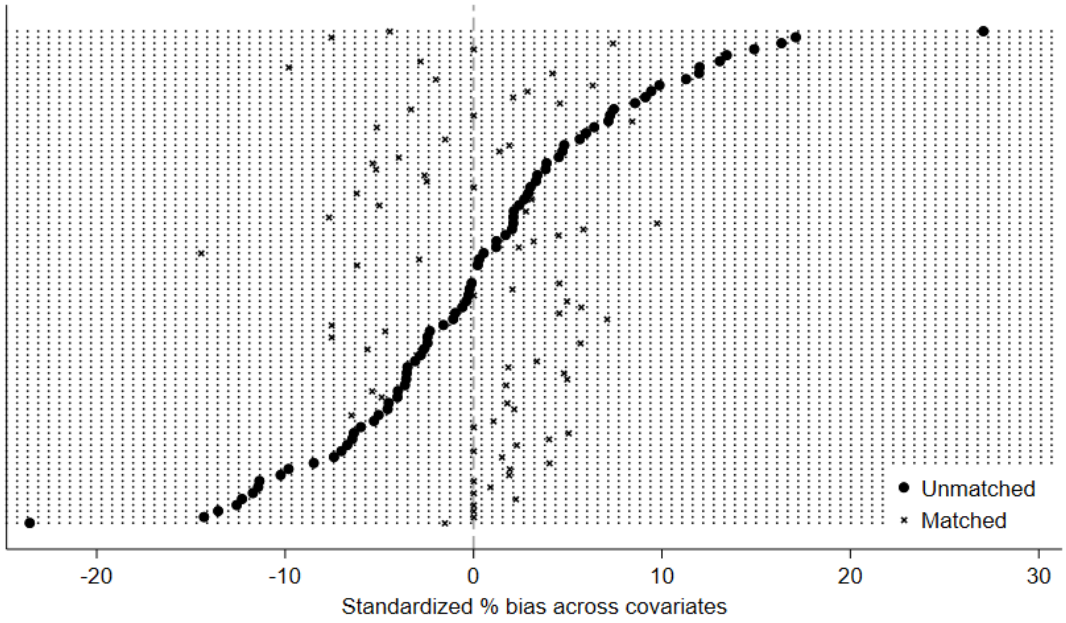


Fig. A. 4 Standardized Mean Differences Plot for Cobariate Balance (Before and After Matching)

Note: This figure presents the standardized mean differences in covariates before and after propensity score matching (PSM). The solid dots represent the bias before matching, while the crosses show the bias after one-to-one nearest neighbor matching. The plot shows that matching substantially reduces the standardized bias across most covariates, with nearly all post-matching values falling below the commonly accepted 10% threshold.

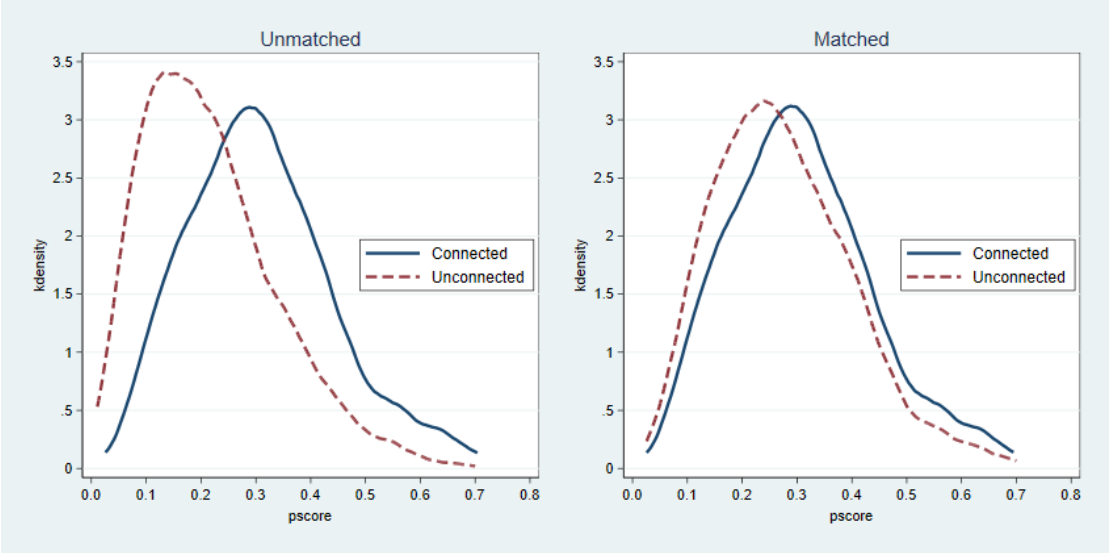


Fig. A. 5 Kernel Density of Propensity Scores for Connected and Unconnected Firms (Before and After Matching)

Note: This figure compares the kernel density distributions of propensity scores between politically connected and unconnected firms before and after propensity score matching (PSM). The left panel displays the distributions in the unmatched sample, showing substantial differences between the two groups. The right panel shows the distributions after one-to-one nearest neighbor matching. The solid line represents connected firms, while the dashed line represents unconnected firms. After matching, the overlap between the two distributions increases markedly, indicating improved covariate balance and support for the common support assumption.

Table A1 Examination of Formal Employment across Provinces by Three Different Variables

Province	(1) Average formal employees share	(2) Proportion of intensive formal firms	(3) Proportion of SMEs employ all employees formally
Shanghai	0.576	0.710	0.469
Guangdong	0.506	0.739	0.330
Zhejiang	0.467	0.622	0.335
Liaoning	0.308	0.432	0.213
Henan	0.226	0.341	0.163
Gansu	0.214	0.327	0.143

Note: Based on data from the ESIEC (2018), this table presents the status of employment formalization across provinces based on three variables. Column (1) reports the average proportion of formal employees within SMEs in different provinces. Columns (2) and (3) respectively show the proportion of intensive formal firms and the proportion of SMEs where all employees are formally employed across provinces. Intensive formal firms are defined as firms that hired at least one formal employee in 2017. Formal employees are defined as those covered by social insurance.

Table A2 Examination of Formal Employment across Industries by Three Different Variables

Industry	(1) Average formal employees share	(2) Proportion of intensive formal firms	(3) Proportion of SMEs employ all employees formally
Agriculture	0.209	0.296	0.150
Leisure & entertainment	0.313	0.415	0.252
Hospitality	0.336	0.518	0.227
Retail	0.338	0.448	0.242
Residential services	0.341	0.49	0.223
Manufacturing	0.351	0.573	0.213
Construction	0.376	0.559	0.226
Property	0.383	0.579	0.250
Finance	0.389	0.667	0.333
Public infrastructure	0.404	0.625	0.250
Information technology	0.406	0.527	0.296
Transportation	0.413	0.581	0.284
Education	0.419	0.654	0.231
Business services	0.431	0.569	0.329
Health	0.479	0.741	0.296
Scientific research	0.508	0.639	0.399
Energy	0.758	1	0.500
Mining	0.762	1	0.667

Note: Based on data from the ESIEC (2018), this table presents the status of employment formalization across industries based on three variables. Column (1) reports the average proportion of formal employees within SMEs in different industries. Columns (2) and (3) respectively show the proportion of intensive formal firms and the

proportion of SMEs where all employees are formally employed across industries. Intensive formal firms are defined as firms that hired at least one formal employee in 2017. Formal employees are defined as those covered by social insurance.

Table A3 Covariate Balance of Connected and Unconnected Firms Before and After Matching

	Unmatched	Mean			% reduct	t-test	
	Matched	Connected	Unconnected	% bias	|bias|	t	p>|t|
A. One by-one neighbour matching							
	U	0.815	0.769	11.3		2.38	0.017
male	M	0.814	0.823	-2.0	82.0	-0.37	0.710
	U	0.505	0.431	14.9		3.22	0.001
college	M	0.506	0.506	0.0	100.0	-0.00	1.000
	U	2.806	2.529	27.1		5.92	0.000
pre-business income	M	2.803	2.848	-4.5	83.5	-0.78	0.438
B. k-Nearest Neighbor Matching (k=3)							
	U	0.815	0.769	11.3		2.38	0.017
male	M	0.814	0.826	-2.9	74.3	-0.53	0.593
	U	0.505	0.431	14.9		3.22	0.001
college	M	0.506	0.513	-1.5	89.7	-0.27	0.789
	U	2.806	2.529	27.1		5.92	0.000
pre-business income	M	2.803	2.854	-5.0	81.4	-0.89	0.373
C. k-Nearest Neighbor Matching (k=5)							
	U	0.815	0.769	11.3		2.38	0.017
male	M	0.814	0.835	-5.2	54.1	-0.96	0.335
college	U	0.505	0.431	14.9		3.22	0.001

	M	0.506	0.521	-3.1	79.2	-0.54	0.591
	U	2.806	2.529	27.1		5.92	0.000
pre-business income	M	2.803	2.854	-5.0	81.4	-0.09	0.368
D. caliper matching							
	U	0.815	0.769	11.3		2.38	0.017
male	M	0.814	0.834	-4.9	56.8	-0.91	0.364
	U	0.505	0.431	14.9		3.22	0.001
college	M	0.503	0.544	-8.4	43.6	-1.44	0.149
	U	2.806	2.529	27.1		5.92	0.000
pre-business income	M	2.793	2.746	4.6	83.2	0.78	0.434
E. kernel matching							
	U	0.815	0.769	11.3		2.38	0.017
male	M	0.815	0.811	0.9	92.2	0.16	0.874
	U	0.505	0.431	14.9		3.22	0.001
college	M	0.505	0.516	-2.1	85.7	-0.37	0.711
	U	2.806	2.529	27.1		5.92	0.000
pre-business income	M	2.806	2.790	1.6	94.0	0.28	0.776

Note: This table reports the covariate balance between politically connected and unconnected firms before and after propensity score matching (PSM), based on five matching algorithms: one-to-one nearest neighbor (Panel A), k-nearest neighbor (Panels B and C), caliper matching (Panel D), and kernel matching (Panel E). "U" and "M" denote unmatched and matched samples, respectively. Each row presents the group-specific covariate means, the standardized percentage bias between groups, the percentage reduction in bias after matching, and the results of a two-sample *t*-test for mean differences. Due to space constraints, we report results for three major covariates, entrepreneur gender, education, and pre-business income, while industry and city dummies are included in the matching procedure but omitted from the table. Across all

matching strategies, post-matching differences are small and statistically insignificant, indicating good covariate balance between the connected and unconnected firms.

Table A4: Testing Alternative Measures of Firm Formalization

	(1)	(2)
Model	OLS	
Dependent variable:	*Employees>8*	*Donation*
political connection	0.072***	0.044*
	(0.022)	(0.024)
male	0.057***	0.005
	(0.022)	(0.024)
college	0.053***	0.060***
	(0.020)	(0.021)
pre-business income	0.036***	0.0001
	(0.009)	(0.010)
age of firm	0.025	0.005
	(0.018)	(0.019)
age square of firm	-0.001	0.0001
	(0.002)	(0.002)
Mean of dependent variable	0.423	0.561
City Fixed Effects	Yes	Yes
Industry Fixed Effects	Yes	Yes
Observations	2,664	2,573

Note: This table reports OLS regression results using two alternative measures of firm formalization. Column (1) uses a binary variable equal to one if the firm reported having more than eight employees, a commonly used regulatory threshold in the literature. Column (2) uses a binary variable equal to one if the firm reported making donations in the year prior to the survey, which reflects its engagement with formal institutions and serves as an indicator of organizational formality. All specifications include city and industry fixed effects and control for entrepreneurs' demographic variables, including gender, education (measured by whether they had a junior college degree or higher), their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table A5: Robustness Check of Political Connection Measure

	(3)	(5)	(3)	(5)
Model	OLS			
Dependent variable:	*intensive formal firm (dummy)*	*formal employees share (continuous)*	*bank loan*	*gov contract*
Political connection (alternative measure)	0.069***	0.079***	0.085***	0.062***
	(0.022)	(0.019)	(0.020)	(0.020)
male	0.009	-0.001	-0.008	0.056***
	(0.022)	(0.019)	(0.018)	(0.016)
college	0.148***	0.126***	-0.050***	0.049***
	(0.020)	(0.017)	(0.016)	(0.017)
pre-business income	0.021**	0.026***	0.011	0.011
	(0.009)	(0.008)	(0.008)	(0.007)
age of firm	0.036**	0.024	0.001	0.027*
	(0.017)	(0.015)	(0.015)	(0.014)
age square of firm	-0.002	-0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Mean of dependent variable	0.516	0.364	0.200	0.220
City Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,664	2,664	2,634	2,610

Note: This table aims to test the robustness of our main findings using an alternative measure of political connection. Specifically, the political connection variable is reconstructed by excluding the criterion indicating whether the firm had a shareholder specifically designated to coordinate government relations, which may reflect responses to administrative intervention rather than genuine political advantage. Columns (1) to (4) report the OLS estimates using this adjusted measure. Column (1) uses a dummy dependent variable equal to 1 if the firm is an intensive formal firm that hire at least one formal employee, and equals 0 otherwise. Column (2) use a continuous dependent variable representing the proportion of formal employees within the firm. In column (3), the dependent variable is a dummy variable that equal to one if the firm obtained a loan from one of the five major state-owned commercial banks, policy banks, or other formal financial institutions at the startup phase, and zero otherwise. In column (4), the dependent variable is a dummy variable that equal to 1 if the firm secured a

government contract in 2017, and zero otherwise. All specifications include city and industry fixed effects and control for entrepreneurs' demographic variables, including gender, education (measured by whether they had a junior college degree or higher), their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table A6: Testing a Rent-Sharing Explanation: Political Connections and Employee Compensation

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS					
Dependent variable:	*ln(technician wage)*	*ln(female manager wage)*	*ln(male manager wage)*	*ln(female worker wage)*	*ln(male worker wage)*	*Labor cost share*
political connection	0.028	-0.007	-0.060**	-0.026	-0.026	0.200
	(0.026)	(0.026)	(0.025)	(0.023)	(0.023)	0.471
male	0.044*	0.035	0.003	0.032	0.010	-0.320
	(0.026)	(0.028)	(0.029)	(0.022)	(0.023)	(0.354)
college	0.038*	0.105***	0.072***	0.063***	0.045**	0.303
	(0.022)	(0.024)	(0.022)	(0.020)	(0.020)	(0.499)
pre-business income	0.032***	0.042***	0.052***	0.020**	0.021**	0.040
	(0.011)	(0.012)	(0.012)	(0.010)	(0.010)	(0.059)
age of firm	0.009	-0.004	-0.008	0.012	-0.008	0.051
	(0.022)	(0.022)	(0.021)	(0.018)	(0.018)	(0.351)
age square of firm	-0.001	-0.001	0.0001	-0.002	-0.001	-0.018
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.043)
Mean of dependent variable	8.371	8.301	8.474	8.066	8.214	0.600
City Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,685	1,167	1,428	1,390	1,562	1,856

Note: This table examines whether politically connected firms offer higher employee compensation and incur greater labor costs, as a way to test the alternative rent-sharing hypothesis that such firms may share the benefits of political connections with their employees in order to discourage whistleblowing and conceal collusion. Columns (1)-(5) report OLS estimates of the association between political connections and the log of average monthly wages (in RMB) for different types of employees within the firm:

technicians, female managers, male managers, female frontline workers, and male frontline workers, respectively. Column (6) uses labor cost share as the dependent variable, defined as the ratio of total employee compensation, including wages, bonuses, social insurance contributions, and other benefit-related expenses to total operating revenue in 2017. All specifications include city and industry fixed effects and control for entrepreneurs' demographic variables, including gender, education (measured by whether they had a junior college degree or higher), their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. Due to data limitations, many firms did not distinguish between technicians, managers, and frontline workers, or did not report wage details by employee type, leading to a relatively high proportion of missing observations in Columns (1)-(5). The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Table A7: IV Regressions with City×Industry Fixed Effects

	（1）	(2)	(3)	(4)	（5）
Model	First stage	2SLS	2SLS	2SLS	2SLS
Dependent variable:	*political connection*	*intensive formal firm* (dummy)	*formal employees share* (continuous)	*bank loan*	*gov contract*
priorgovjob	0.070*** (0.023)				
political connection		0.784* (0.427)	0.664* (0.363)	0.530* (0.317)	0.618* (0.407)
male	0.021 (0.022)	0.002 (0.030)	-0.025 (0.026)	-0.012 （0.024）	0.053* （0.024）
college	0.058*** (0.020)	0.104*** (0.039)	0.093*** (0.033)	-0.063** （0.028）	0.011 （0.033）
pre-business income	0.047*** (0.009)	-0.012 0.023	0.001 0.020	-0.017 （0.018）	-0.015 （0.021）
age of firm	0.034* (0.018)	0.023 (0.028)	0.015 (0.024)	-0.023 (0.020)	0.012 （0.023）
age square of firm	-0.003 (0.002)	-0.002 (0.003)	-0.001 (0.003)	0.003 （0.002）	-0.003 （0.002）
Mean of dependent variable	0.236	0.516	0.364	0.200	0.177
City×Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,475	2,475	2,475	2,445	2,417
F-Statistics	9.226				

Note: Column (1) presents the first-stage regression results, where the instrument is a dummy variable indicating whether the entrepreneur previously held a job in the public sector (*priorgovjob*), including positions in government agencies, state-owned enterprises, or public institutions. Columns (2)-(5) report the two-stage least squares

(2SLS) estimates for the following dependent variables: *intensive formal firm* (dummy), *formal employees share* (continuous), *bank loan*, and *government contract*. All specifications include city×industry fixed effects and control for entrepreneurs' demographic variables, including gender, education (measured by whether they had a junior college degree or higher), their self-assessed pre-business income, as well as the firm' age and age square to account for firm size. The heteroscedasticity robust standard errors are reported in parentheses. The symbols *, **, and *** stand for the significance levels at 10%, 5%, and 1%, respectively.

Appendix 2: Extended Model Setup

In this section, we extend the conceptual framework to accommodate the empirical reality that some firms adopt a hybrid employment structure, employing both formal and informal workers.

The most straightforward way to incorporate this feature is to introduce three types of firms: fully informal, partially formal, and fully formal. We assume that partially formal firms, defined as those firms providing social insurance to a subset of employees, face an interest rate that lies between fully informal and fully formal firms, i.e. $r_f < r_p < r_i$, and pay a tax less than τ, i.e. $p\tau$, where p in (0,1). The production technology remains Cobb-Douglas, consistent with the main model. Therefore, a firm with productivity θ that chooses to operate as a partially formal firm has the following one-period profit function:

$$\pi_p(\theta) = (1 - p\tau)\theta {k_p}^{\alpha} - r_p k_p \quad (1)$$

From the first-order conditions, we derive the optimal capital k_p^*

$$k_p^* = [\frac{r_p}{\alpha\theta(1-p\tau)}]^{\frac{1}{\alpha-1}} \quad (2)$$

Substituting the optimal capital allocations back into the profit function yields the following expression for the profit of the partially formal firm:

$$\pi_p(\theta) = (1 - p\tau)\theta {k_p}^{\alpha} - r_p k_p \quad (3)$$

Assume β is the discount rate and firms live forever, the value function for the partially formal firm is defined as:

$$V_p(\theta) = \sum_{t=0}^{\infty} \beta^t \pi_p(\theta) \equiv \frac{1-\alpha}{\alpha(1-\beta)} r_p \left[\frac{\alpha\theta(1-p\tau)}{r_p}\right]^{\frac{1}{1-\alpha}} \quad (4)$$

We now extend the model by allowing firms to choose among three operational types: fully informal, where no social insurance is provided to employees; partially formal, where social insurance is provided to a subset of employees; and fully formal, where firms fully comply with labor regulations. This decision criterion can be represented mathematically as follows:

$$V_p(\theta) - V_i(\theta) \geq C_p \quad (5)$$

$$V_f(\theta) - V_p(\theta) \geq C_f - C_p \tag{6}$$

Firms will choose to become partially formal if the discounted future benefits of formalizing employment, net of registration costs, exceed the expected future benefits of remaining fully informal. This occurs when equation (5) holds. Similarly, becoming fully formal is more profitable than remaining partially formal when equation (6) is satisfied. For simplicity, we assume that the registration cost for partial formalization is half that of full formalization, i.e. $C_p = \frac{1}{2} C_f$.

Referring to the value functions for fully formal and fully informal firms in the main text (equations (8) and (9)), and combining them with equations. (4), (5), and (6) in this appendix, we can derive two productivity thresholds, θ_1^* and θ_2^*:

$$\theta_1^* = \left\{ \frac{\frac{\alpha(1-\beta)}{2(1-\alpha)} C_f}{\left[\frac{(\alpha(1-p\tau))^{\frac{1}{1-\alpha}}}{r_p^{\frac{\alpha}{1-\alpha}}} - \frac{\alpha^{\frac{1}{1-\alpha}}}{r_i^{\frac{\alpha}{1-\alpha}}}\right] \alpha^{\frac{1}{1-\alpha}}} \right\}^{1-\alpha} \tag{7}$$

$$\theta_2^* = \left\{ \frac{\frac{\alpha(1-\beta)}{2(1-\alpha)} C_f}{\left[\frac{(\alpha(1-\tau))^{\frac{1}{1-\alpha}}}{r_f^{\frac{\alpha}{1-\alpha}}} - \frac{(\alpha(1-p\tau))^{\frac{1}{1-\alpha}}}{r_p^{\frac{\alpha}{1-\alpha}}}\right] \alpha^{\frac{1}{1-\alpha}}} \right\}^{1-\alpha} \tag{8}$$

When parameters satisfy $\frac{(\alpha(1-\tau))^{\frac{1}{1-\alpha}}}{r_f^{\frac{\alpha}{1-\alpha}}} + \frac{\alpha^{\frac{1}{1-\alpha}}}{r_i^{\frac{\alpha}{1-\alpha}}} < \frac{2\alpha(1-p\tau))^{\frac{1}{1-\alpha}}}{r_p^{\frac{\alpha}{1-\alpha}}}$, it follows that $\theta_1^* < \theta_2^*$.

Therefore, we can infer the following regime classification: firms with productivity below the lower threshold θ_1^* will remain fully informal; firms above the upper threshold θ_2^* will fully formalize; and firms in between will choose to be partial formal.

Why Do Chinese SMEs Avoid Formal Employment? Political Connections and Unequal Access to Formalization Benefits

Highlight

- We draw on a novel firm-level survey and provide the first systematic documentation of the distribution of formal employment among Chinese SMEs across provinces, industries, and registration cohorts.
- We develop a conceptual framework grounded in the cost–benefit formalization model of Ulyssea (2020), incorporating how entrepreneurs' political connections and the local business environment affect their access to formalization benefits.
- We present robust empirical evidence that SMEs led by politically connected entrepreneurs are more likely to provide formal employment. Moreover, this positive association is weaker in regions with a more favorable business environment.
- We offer key policy insights: fostering inclusive formalization requires improving the business environment to decouple formalization benefits from political connections and to ensure that genuine formalization benefits extend broadly to all entrepreneurs.