Familiar Strangers: Overseas Chinese, Lineage Connection and Foreign Firms in China

Fanghao Chen   Ruichi Xiong   Xiaobo Zhang

PKU   UofT   PKU & IFPRI

August 11, 2021

Abstract

This paper studies the importance of international lineage network in facilitating the entry and survival of foreign firms in developing countries where formal institutions are largely immature. We build a unique dataset on the universe of foreign firms in China that are controlled by Huaren (overseas Chinese) during China's Reform Era (1981-1996). To identify the causal effects of lineage connection, we exploit variations from both China's stepwise openness and surname distributions across prefectures, to implement a Difference-in-Difference design with differential treatment to Huaren foreign firms whose legal representatives differ in their surnames, which allows us to control for unobserved factors at multidimensional level. We find that, following the local openness reform, lineage connection facilitates the entry and survival of Huaren foreign firms, through reducing information frictions, securing property right, and smoothing contract enforcement. And the lineage-driven FDI brings about an array of long-term economic benefits, including greater GDP growth, wage level, innovation, non-Huaren FDI inflows, export and import.

Key Words: International Migration, Lineage Network, Foreign Direct Investment, China

JEL Codes: F22, F23, L14

1   Introduction

Governments around the world strive to attract foreign direct investment (FDI). They hope, by introducing the investment from new and high-performing foreign firms, the domestic sectors can harvest an assortment of positive spillovers (Harrison and Rodríguez-Clare, 2010). Developing countries, however, often face difficulties in attracting FDI. Severe information frictions and immature institution building are often the major obstacles to blame. The case of China, however, is exceptional: Ever since China's dramatic reform and opening-up in 1978, it has quickly ascended from a closed state to one of the hubs for attracting multinationals and foreign firms, which challenges the conventional wisdom. Many anecdotes attribute this achievement to "Huaren" —overseas Chinese[1], who though as emigrants, are believed to have strong lineage connection with their ancestral hometown: They pioneered in an immature market where policies and institutions were uncertain and insecure, used informal networks to overcome entry barriers, as well as planted the seed investment for China's later industrial growth (Vogel, 1990). Yet no study so far, to the best of our knowledge, has ever quantitatively documented such claims, or empirically examined the causal relationship between lineage connection and the entry and survival of Huaren foreign firms in China.

In this paper, we use a comprehensive and disaggregated dataset on the universe of foreign firms in China that are controlled by Huaren (overseas Chinese) during China's Reform Era (1981-1996), and ask whether the entry and survival of a Huaren foreign firm is affected by its lineage connection with Chinese prefectures, when hard (such as, communication tools) and soft (institution) infrastructure for market were largely immature. In China, lineage refers to a group of descendants of a common patrilineal ancestor within multiple generations, who share the same surname. For thousands of years in China, lineages have provided local public goods and promoted reciprocal cooperation within the group, as an informal institution that is parallel to the formal institution like state (Greif and Tabellini, 2017). Activities such as compiling genealogy books and worshiping ancestors at lineage temple during family reunions last even to date (Clark, 2015). Many accounts suggest the same practices are still honored by overseas Chinese even if they emigrated and settled down in a foreign country (Kuhn, 2008). By identifying the legal representative's surname and the type of ID document, we first document with full-sample administrative data that a substantial portion of foreign firms in China founded since

---

[1]"Huaren" is the Mandarin Chinese pronunciation for overseas Chinese. Throughout this paper, Huaren refer to ethnic Chinese that hold citizenship of foreign countries or Hongkong, Macau, and Taiwan, and Huaren and overseas Chinese are interchageable in meanings.

the 1980s are controlled by overseas Chinese. We then show that Huaren-controlled foreign firms are disproportionately attracted to and survive in prefectures where their kin surnames are concentrated, i.e. where they tend to have stronger lineage ties, and this effect appears to be driven mainly by reducing information frictions and facilitating contract enforcement upon entry. These "zero-to-one" foreign entrants, as we find, amount to differential FDI stocks in the long run, which further leads to divergent economic development in various dimensions across China.

The analysis proceeds in several steps. In the first step, we introduce the new database that we assemble for this research. We start with firm registration data on the entry and exit information for the universe of firms ever registered in China, provided by the State Administration of Industry and Commerce of China (SAIC). Then we screen out all of the foreign firms established between 1981 to 1996, according to the official ownership code assigned by the SAIC. The research period we choose, earilier than the WTO accession, is often referred to as the "Reform Era", when China was de facto open, but along with massive policy uncertainties. Huaren businessmen, as we will show later, are the dominant type of foreign investors during this period. Next, we determine whether a foreign firm is controlled by a Huaren by telling from the spelling of the legal representative' surname. The above steps lead to our final sample of Huaren foreign firms where we can use the observed surnames of the legal representatives to proxy the firm's lineage connection with each prefecture.

In the second step, we describe and implement our main empirical design to identify the effect of lineage connection. As is obvious, unobserved factors at surname level, prefecture level or even surname-prefecture level, may simultaneously affect lineage connection and the presence of "connected" Huaren foreign firms, creating a spurious correlation. For example, southeastern China, where most Huaren originated from, has natural advantages for attracting foreign firms because of their proximity to the sea and desirable placed-based policies such as special economic zones (Wang, 2013). Moreover, entrepreneurial culture and tacit knowledge for conducting international business can be passed on from generation to generation within certain lineage groups. To overcome these concerns, we adopt a Difference-in-Difference (DID) strategy that exploits variations from both China's staggered openness to foreign capital and predetermined surname distributions across space. With this framework, we are able to control for fixed effects at surname-prefecture, surname-year, and prefecture-year level, so as to guard us against a wide range of confounding factors and arrive at causal estimates of the effects of lineage connection.

We find large positive effects of lineage connection following openness shocks on the presence of Huaren foreign firms at surname-prefecture-cohort level, for the following three measures: the counts of entrants, the urvival-adjusted counts of s entrants, and the counts of survived firms in 2014. Our findings remain robust to examining extensive margin of and alternative function forms for firm entry, using alternative lineage connection measures, excluding FDI-intensive prefectures, regressing on subsamples within or without emigration-intensive provinces and surname, and several placebo tests. The long run effects of lineage connection on the survival of Huaren foreign firms can be further broken down to short term advantages in lowering entry barriers or long term advantages in increasing chances of survival. We separate the two margins by looking at the effect of lineage connection on survival ratio conditional on entry. The results are ambiguous. Although we find effects of lineage connection on the counts of survived firms, we did not discover benefits of lineage connection on survival probability of Huaren foreign firms after their establishment. This motivates us to interpret the role of the lineage connection as mainly lowering entry barrier in the short run regardless of Huaren firm's latent quality. Besides, the results also help to rule out an alternative hypothesis which emphasizes the potential selection of Huaren firms with higher latent quality into location with higher lineage connections.

The third step sheds light on the mechanisms through which lineage connection helps overcome the entry barriers for Huaren foreign firms. We either assess heterogeneous effects of lineage connection across prefectures and cohorts, or at even finer granularity. We find that the effects of lineage connection are stronger in prefectures where there is higher social fragmentation or better telecommunication infrastructure. We also find that Huaren investors with higher lineage connection are more likely to start and maintain a business as an imitator of existing industries, a headquarter firm (relative to a branch firm), and in labor-intensive and input-intensive industries, which reveals the role of lineage connection as a kind of social network in transmitting information and enforcing contracts (Rauch, 2001).

The final question is whether the lineage-driven FDI that came along with Huaren entrants during the early stage of China's opening-up has a persistent effect on economic development today. We implement a strategy similar to Sequeira, Nunn and Qian (2020) to isolate lineage-driven FDI across prefectures that is independent of various confounding factors. We find large spillover effects of these lineage-driven FDI on economic development along a variety of dimensions, including greater GDP growth, wage level, innovation, non-Huaren FDI inflow, export and import. This suggests that the benefits of introducing FDI as industrial seeds seem to outweigh its costs during the take-off stage

of an developing economy.

Existing Literature. Our work speaks to a large literature on the role of social affinities in facilitating trade and investment flows across countries and regions (Rauch, 2001). Specifically, empirical evidence has found that social affinities, in the form of ethnic ties (Rauch and Trindade, 2002), trust (Guiso, Sapienza and Zingales, 2009), linguistic proximity (Melitz and Toubal, 2014), migrant network (Javorcik et al., 2011; Parsons and Vézina, 2018), and even Facebook relationships (Kuchler et al., 2020) have led to increased bilateral trade and investment.

We contribute to this literature in several ways. First, we quantify a new type of social affinity that functions between a locality and a surname in China—the lineage connection, sustained and carved into the second-nature of Chinese people by the repeated interactions within the kinship community throughout history (Greif, 1993). We then provide a causal identification for the effect of lineage connection, by aggregating the entry and survival of Huaren foreign firms to the cohort-surname-prefecture level and controlling for high-dimensional fixed effects, which helps guard against a wide range of possible confounding factors such as geographic advantages, place-based policies, surname-specific expertise. The paper closest to ours is Burchardi, Chaney and Hassan (2019) and Ma (2018). Burchardi, Chaney and Hassan (2019) identifies the causal effect of ancestry compositions on the present-day FDI inflows into U.S. counties exploiting the information from historical immigration. Ma (2018) presents evidence on the effect of cultural similarity on attracting culture-specific FDI, employing the geographical borders of Chinese linguistic dialect zones. Our paper differs from theirs in examining the role of lineage connection in a dynamic perspective during China's gradual opening up. We make best use of our data structure to separate the respective effects on entry and survival, which helps shed new light on the mechanisms of how lineage connection works.

This paper also complements the literature on the long term implications of international migration. Overseas Chinese networks are created as a product of historical migration. What are the long term effects on the economic outcomes of the origin and host regions remains an open question, concerning both developed and developing countries. Existing research primarily focus on examining the effect of immigration on the developed countries (see for example, Sequeira, Nunn and Qian (2020)). But research studying the effect of emigration on the origin (usually developing) countries are pretty scarce. Nanda and Khanna (2010) is one of the few exceptions, exploring the ties between the Indian expatriate community and local entrepreneurs in India's software industry. But they only use a surveyed sample of firms in a particular industry while we draw upon

the universe of Huaren foreign firms in China. We add to this literature by showing that an unexpected benefit of emigration is to facilitate foreign investments through migrant networks that provide seed investments for the home country. In addition, our usage of surnames to recover the ancestral origins of Huaren investors advances a literature on the informational content of names (Abramitzky, Boustan and Eriksson, 2020; Bazzi, Fiszbein and Gerresilasse, 2020). Unlike precedents that mainly draw implications from given names, we extract the information of surname in the context of Chinese lineage groups.

A narrower literature related to ours has looked at the determinants and spillovers of inward FDI in China. China's swift absorption of FDI following its opening up challenges the conventional wisdom that the poor market environment in developing countries would deter foreign investments. Therefore, understanding the case of China has direct policy implications for the vast developing world, which helps answer the inquiry of Lucas (1990): "Why doesn't capital flow to developing countries?" Prior studies have primarily focus on common factors such as market access (Amiti and Javorcik, 2008), infrastructure and industrial policies (Cheng and Kwan, 2000), institutional quality (Du, Lu and Tao, 2008). While a notable feature of China's FDI policies is to make use of the overseas Chinese business network (Vogel, 1990), few studies have offered a quantitative description, perhaps due to data limitation. We make progress in this strand of literature in three aspects: First, we draw upon a unique dataset and offer a systematic look at Huaren foreign firms across Chinese prefectures from 1981 to 1996, and document that a substantial portion of foreign firms in China founded since the 1980s are run by overseas Chinese. Prior literature typically focused on a specific place at a particular time, making it difficult to establish empirical regularities (Tan, 2006). Moreover, we discuss the effect of lineage connection on Huaren foreign firms at a fairly disaggregated level, providing a causal identification framework. Third, isolating lineage-driven FDI proves to help account for the large regional differences in FDI inflows and divergent growth path within China, which past framework from literature has not perfectly addressed (Lin, Liu and Zhang, 2009; Lu, Tao and Zhu, 2017; Ma, 2018),

Finally, our research advances an emerging literature using firm registration data of China. Dai et al. (2019) and Bai et al. (2020) explores the role of community networks and state power underlying the rapid growth of China's private sector, respectively. While most existing studies pay attention to the private domestic firms, we focus on foreign firms, a relatively small but indispensable sector that this line of literature has yet touched on.

The rest of the paper is organized as following. Section 2 introduces the historical

background on overseas Chinese and China's gradual opening-up. Section 3 describes our data. Section 4 presents the empirical findings. Section 5 explores the mechanism of our findings. Section 6 concludes.

2 Historical Background

Ever since its dramatic turn toward market economy in 1978, within thirty years China has swiftly ascended from a completely closed state to a hub for multinational and foreign firms. Figure 1 plots the entry and survival of foreign companies in mainland China by entry cohort using firm registration data from SAIC database, suggesting that the establishments of foreign firms were already active even before China's accession to WTO. From 1979 to 1999, China became one of the most popular destinations for foreign direct investment (FDI), only second to the United States (Huang, 2003). This achievement is unusual not only for its volume, but also for its speed, which challenges the conventional observation that the information frictions and poor institutional environment in developing countries would deter foreign investments. Peering into the historical details, one may faithfully posit that China's widespread networks with overseas Chinese must have played a crucial role.
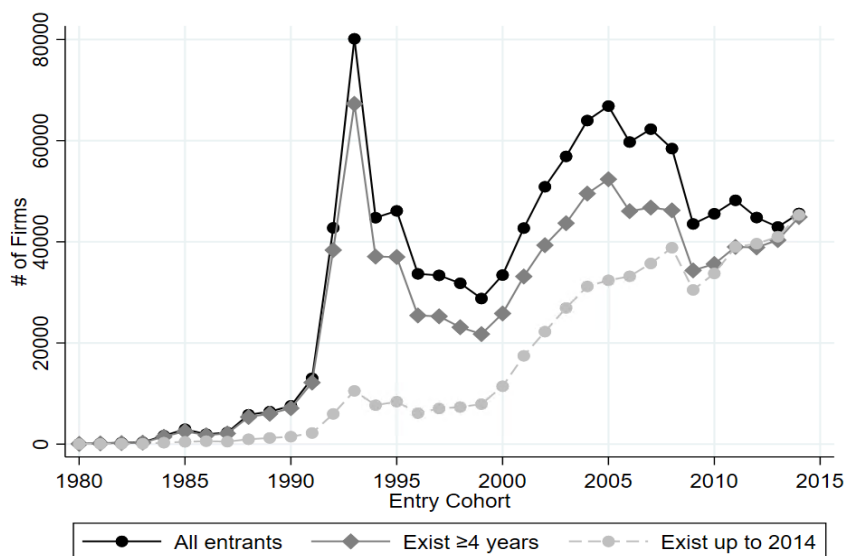

Figure 1: Entry and Survival of Foreign Firms by Cohort

China has a long history of emigration that can date back to at least 1300s A.D. (Kuhn, 2008). According to the Global Migrant Origin Database of the Development Research

Centre, the stock of overseas Chinese has reached 5.8 million in 2007, making China the world's 6th largest country in exporting migrants[2]. Besides, many overseas Chinese have prospered and even become affluent class in the host countries. A typical example is in South East Asia. According to the Economist's estimate, in 2019 more than three-quarters of 369$ billion in South East Asian billionaire wealth was controlled by Huaren[3].

Despite a 30-year seal-off since 1949 when China turned closed, the bond between overseas Chinese and their ancestral hometowns has never been truly cut off. The overseas Chinese, though physically abroad, are emotionally attached to their ancestral hometowns. Many anecdotal facts show that overseas Chinese remained socially in touch with their lineage members, through sending letters and remittance (Tan, 2006). These behaviors, as Greif and Tabellini (2017) illustrates, reflect some persistent and fundamental values that is rooted in the Chinese culture. For centuries, people from the same lineage groups exchange information, smooth contract enforcement, and weather crises by helping out on each other.

After 1978, when the political leader of China, Deng Xiaoping initiated the great Reform and Opening Up, China took steps to open up its territories to international trade and foreign investments (Branstetter and Lardy, 2006). Table 1 describes the staggered opening process across Chinese regions in the Reform Era. In 1979, the Law on Joint Ventures was passed, providing, for the first time in thirty years, legal framework under which foreign firms were allowed to operate in mainland China. In the following year, four "Special Economic Zones" (Shenzhen, Zhuhai, Shantou, Xiamen) were established where foreign corporations enjoyed broad autonomy and preferential tax treatment. Later on in 1984, additional 14 cities were granted the status of "Coastal Open City" in a bid to absorb FDI. Yingkou, Weihai, and Hainan provinces were added to the list of territories open to foreign capital in 1985, 1987, and 1988 respectively. Although the economic and political crisis in 1989 were a huge blow to the country, Deng's visit to the southern China in 1992 restored people's confidence in China's resolution to embrace market economy and globalization, marking the start of comprehensive reform and national openness. Figure 1 shows a spike of foreign firm entry around 1992, which is consistent with this historic event.

Meanwhile, many local governments in the reformed regions realized the complementarity between the liberalized FDI regime and the local lineage connection with overseas Chinese. They enacted policies and coordinated through specialized organizations to encourage overseas Chinese businessmen to invest back in their ancestral hometowns, in

---

[2]See http://www.sussex.ac.uk/Units/SCMR/drc/about/index.html
[3]Chinese Diaspora Inc: High-Wire Act", The Economist, May 30th 2020

Table 1: Stepwise Openness Across Chinese Regions in the Reform Era

Year	Event	Region
1980	Special Economic Zone	Shenzhen, Zhuhai, Shantou, Xiamen
1984	Coastal Open City	Dalian, Qinhuangdao, Tianjin, Yantai, Qingdao, Lianyungang, Nantong, Shanghai, Ningbo, Wenzhou, Fuzhou, Guangzhou, Zhanjiang, Beihai
1984	Coastal Open City	Yingkou
1987	Coastal Open City	Weihai
1988	Special Economic Zone	Hainan Province
1990	Special Economic Zone	Shanghai Pudong District
1992	Deng Xiaoping's South Tour	All other regions

pursuit of capital, technological capacity, and export channels (Huang, 2003). To be sure, having been shut down from the world market for three decades, China was at a huge disadvantage to engage in international trade. However, these transitional policies set the lineage network in motion, and remain persistently effective even if they had phased out. With the help of the international ethnic ties, China' share in international trade volumes and FDI inflows rocket up even before accession to WTO. And the pionnering foreign entrants are believed to plant the seed investments for China's later industrial growth, and bring about a myriad of benefits in the long run.

3   Data Description

3.1   Data Sources

To empirically test the effect of lineage connection on foreign firms, we merge aggregated data at surname-prefecture-cohort level from multiple large-scale micro datasets.

Firm Registration Data. The main dataset we use is based on firm registration data, maintained by the State Administration of Industry and Commerce of China (SAIC), that track the universe of firms ever registered in China up till 2014. For each firm, we observe its entry date, exit date (left blank if the firm still survives), its latest-reported registered capital by the end of 2014, 4-digit industry code, county-level location code, ownership code, and a complete list of immediate shareholders and personnel.

2005 Population Census. The 2005 population Census reports individual surname for a 0.2% representative sample of the population, from which we recover the geographic

distribution of surnames across prefectures. We then leverage the geographic distribution of surnames to construct our measure of lineage connection for each surname-prefecture pair, EG index for each surname, and social fragmentation for each prefecture.

Other Sources. Additional variables on economic development at prefecture level are retrieved from CSMAR database, supplemented by various China City Statistical Yearbooks. The variables include GDP per capita, number of registered patents, wage, export and import volumn. The capacity of telephone exchanger at province-year level is from China Compendium of Statistics (2009). Sector-level capital intensity (capital to labor ratio) is constructed based on manufacturing samples in the 2004 Industrial Census, and contract intensity is based on 2002 Input-Output Table of China following the methodology of Levchenko (2007), respectively.

3.2 Foreign Firm Database and Surname-Based Lineage Connection

Identifying Foreign Firms. To construct a comprehensive dataset on the foreign firms in China, we first need to have a consistent definition of foreign firms. Two approaches coexist in literature. The first approach uses the ownership codes of firms, and the second approach uses shareholder information. In this paper we opt for the first approach for two reasons. First, it is often the case that immediate shareholding structure does not imply ultimate control structure. For example, investors can exert control over a firm through holding shells (Bai et al., 2020). Second, joint ventures (JVs) and branch firms, as specific forms of foreign firms[4], can only be distinguished from ownership code.

One might be concerned that ownership of a firm may change as a result of privatization that took place during the late 1990s. However, privatization in China mostly took the form of Management But-Outs (MBO) rather than sales to foreigners. Moreover, any change of ownership will trigger a change in firm identifier as well, creating a new legal entity (Chen et al., 2020). Thus, it is unlikely that the changeable nature of ownership type will threaten our identification of foreign firms.

---

[4]There are several subcategories of registration type for foreign firms, depending on the investor's origin and the entry form. First, investors coming from non-mainland China, namely Hongkong, Macau, and Taiwan (HMT), would also be classified as foreign investors and distinct from investors from other parts of the world. Second, wholly-owned foreign firms are also separately coded; Third, joint ventures (JVs) which are set up by partnership between a foreign and a domestic firm, are also identified as on specific form of foreign firms. Unlike other foreign firms, a JV is not determined by a certain threshold of foreign equity share (normally 25% according to China's law), but emphasizes the joint efforts of both parties into a contracted project, whose formation is most likely to be related to FDI regulation and technology transfer (Jiang et al., 2018); Fourth, we can also know whether a foreign firm is set up as a headquarter, or a branch which does not have an independent legal entity.

Therefore, we conclude it is better to screen out foreign firms according to the ownership code that the SAIC assigns to each firm. This process compiles a comprehensive database on the universe of foreign firms in China.

Identifying Surnames of Huaren Firms. An additional advantage of SAIC database is that we are able to observe each firm's personnel members' full names and their ID information as they reported to the SAIC upon registration. We design an algorithm as described in Appendix A to extract the surname of each foreign firm's personnel member, and determine whether he or she is an ethnic Chinese. If an ethnic Chinese also holds a non-mainland Chinese ID—including passport of a foreign country, travel permit for Hongkong, Macau, Taiwan (HMT) residents—he or she is defined as a "Huaren", i.e. an overseas Chinese.

We then use the surname of the Huaren legal representative to represent the lineage endowment of a foreign firm. Having worked through the list of the personnel and shareholders of foreign firms, we arrive at several findings to support our choice of legal representative among all the personnel to represent the firm. First, we show in Appendix B, that 97% of foreign firms have unique legal representative within the firm, while they might have none or multiple directors, CEOs, and top-rank executives. Second, we do not observe the surnames of most foreign shareholders of foreign firms. 92% of immediate foreign shareholders hold equity stakes of foreign firms through legal persons rather than natural persons. Even a backing-out through ownership chain to identify each foreign firm's ultimate natural person owners will not work perfectly, because the SAIC does not have information on the owners of foreign legal persons that are registered abroad (Bai et al., 2020). Third, Appendix B reports that the chance that a legal representative overlaps a top-rank executive position within a foreign firm is 93%, which means a legal representative is highly likely to be the person in charge.

Our sample for empirical analysis consists of all foreign firms that are represented by an overseas Chinese. There are several caveats to make. First, our sample would exclude all the foreign firms that are represented by a Chinese mainlander. Such firms may be related to what Huang (2003) coins "roundtrip FDI", which refers to capital that a domestic firm first exports to overseas (majorly Hongkong) and then imports back into mainland China as a new legal entity. Since establishments of such firms are most likely to be motivated by the tax breaks and preferential treatment granted toward foreign firms during early periods, taking this subgroup of firms into account would exaggerate the presence of foreign firms in China. On the other hand, however, we would also leave out foreign firms that are delegated to the mainlander (perhaps a family member of

the overseas Chinese investor). Hence, our estimates should be interpreted as a lower bound for the presence of lineage-related foreign firms. Moreover, our main dataset also excludes foreign firms represented by non-Huaren since we want to keep our sample as comparable as possible[5]. Last, since the legal representative is not necessarily the major shareholder of the firm, it is possible that a non-Chinese-owned company, but represented by a Huaren, is classified as Huaren foreign firm, according to our definition. Yet, the choice of appointing Huaren legal representative still reveals the motive of the foreign firm to mobilizing lineage connection in some sense.
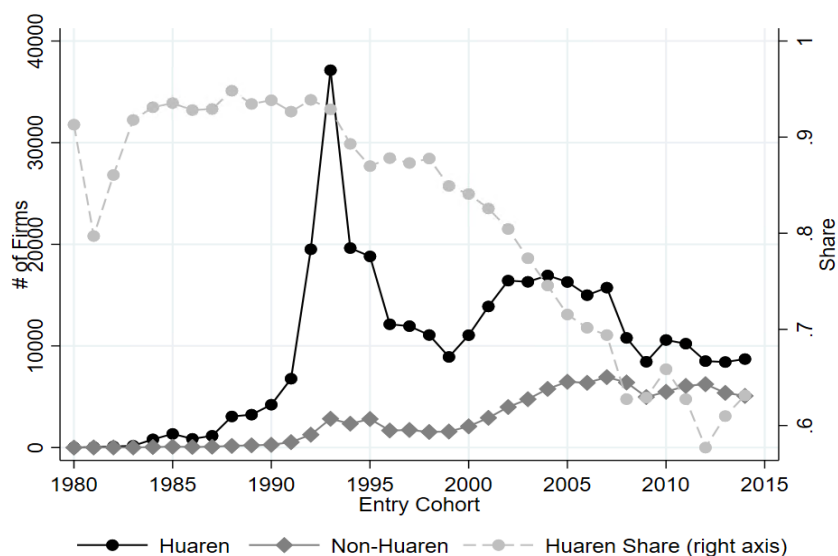

Figure 2: Breakdown of Foreign Firms by Cohort

Figure 2 provides a breakdown of foreign firms by Huaren and entry cohort. The share of Huaren foreign firms is astonishingly high, which echoes previous qualitative research. Before 1995, the portion of Huaren firms among all foreign firms can be above 90%. Over time, however, the share dwindled to 60% in 2014, suggesting a secular decline in the importance of lineage network as market institutions formalize.

Surname-Based Lineage Connection. The 2005 population Census reports individual surname for a 0.2% representative sample of the population. To exclude rare surnames, we first rank the sample size of each surname from common to rare, and then select surnames sequentially until they account for a cumulative 99% of population. We calculate the size of the selected surname groups to the prefecture-surname level. This gives us

---

[5]Including non-Huaren foreign firms would mechanically increase the size of the control group. Yet, most of our findings are robust to including non-Chinese foreign firms. See section 4.3.

the geographic distribution of 233 surnames across 314 prefectures[6]. We then leverage the geographic distribution of surnames to construct our measure of lineage connection for each surname-prefecture pair, EG index in terms of surnames for each surname, and social fragmentation in terms of surname for each prefecture.

We define the lineage connection m_{sp} between a surname s and prefecture p as follows:

$$m_{sp} = \frac{E_{sp}}{\sum_p E_{sp}} \tag{1}$$

where E_{sp} denotes the population of surname s in prefecture p. The construction of social fragmentation, defined as one minus the Herfindahl-Hirschman Index (HHI) across all surnames in a prefecture, follows Padró i Miquel, Qian and Yao (2012). Notably, our measure of lineage connection is "size-free" in the sense that populous surnames will not be associated with universally higher connection with all prefectures[7]. So we explicitly rule out the heterogeneity of surname size in determining the lineage connection, across all our specifications[8].

Moreover, the novelty of our approach lies in the usage of surname instead of migration patterns to proxy lineage connections. The reasons are two-fold. One is for practical purpose. While there are relatively rich data for immigration patterns into developed countries such as the United States (Burchardi, Chaney and Hassan, 2019; Sequeira, Nunn and Qian, 2020), so far the data on emigration, especially at sub-national level for developing countries like China, are extremely scarce. Our surname-based measure makes best use of the fact that lineage institution in China are operating within a surname (Clark, 2015). The other reason is for methodological purpose. We construct a pre-determined and time-invariant variable, instead of an endogenous and rolling variable (such as em-

---

[6]There are three reasons that we choose prefecture (instead of county) as the geographic unit for our research. First, measurement noises for the presence of rare surnames are large at small counties. Second, boundaries of counties may change over time while that of prefectures are fairly stable. Third, we speculate that lineage should function in a broader geographic area than county, though within a certain range (such as province).

[7]For purpose of comparison, if we use population size of surname s in in prefecture p as a measure of lineage connection, we can expect significant correlation even with the function of lineages absent, since populous surnames like "陈" ("Chen", "Chan", "Tan") can be dominant both domestically and abroad. Moreover, Appendix C indicates that our lineage connection is actually negatively correlated with prefecture population at surname-prefecture level.

[8]Another dimension with respect to the lineage connection is how strong the lineage group it. The Chinese lineage culture values the notion of "branching out", which highlights the preference for family size, and for the unification of the family (all offsprings should belong to the same "tree"). In Appendix D, we consider the possibility that the strengths of lineage connection are heterogeneous across surnames measured by the EG index of that surname (Ellison and Glaeser, 1997). The rationale is that the EG index of a surname measures the extent to which a person of this surname benefits from co-locating with his or her kin surname. However, we find little evidence in support of this heterogeneity.

igration flows) to proxy lineage connection. This significantly lessens the endogeneity concerns.
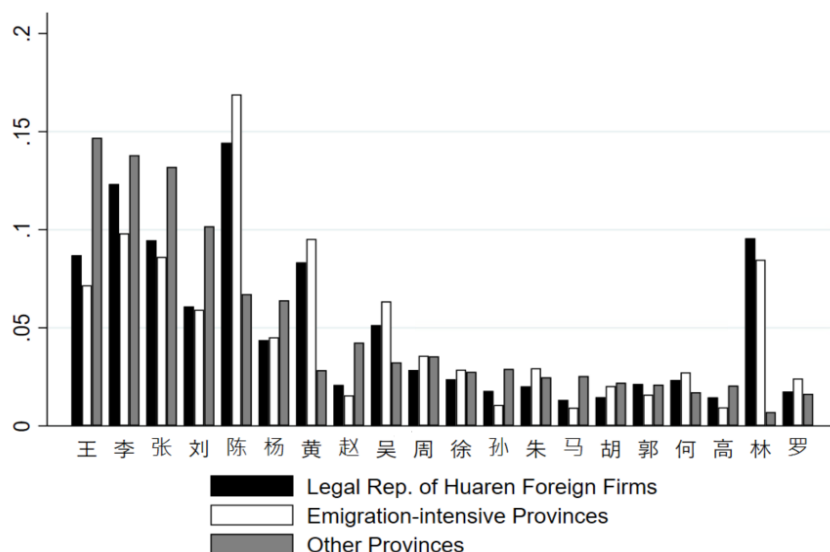

Figure 3: Surname Distribution: Domestic Population and Huaren Foreign Firms

Figure 3 visualizes the geographic distributions of 20 most populous surnames in China. To compare, we draw the distributions separately for emigration-intensive provinces (including Guangdong, Fujian, Zhejiang), other provinces, and all legal representatives of Huaren foreign firms from 1980 to 2014. We discover informative difference and similarity between the three distributions. For example, "陈" ("Chen", "Chan", "Tan")[9] is the most common surname among Huaren legal representatives and in emigration-intensive provinces, but ranks only fifth in the total population. "林" ("Lin", "Lim", "Lam") and "黄" ("Huang", "Wong") are another two typical Huaren-intensive surnames . Overall, the surname distribution using all Huaren foreign firm legal representatives are closer to the that of the population in emigration-intensive provinces. This exercise suggests that the variations of surname distribution across regions can help with our identification.

In our empirical analysis, we treat our lineage connection measure as time-invariant and predetermined, assuming that the geographic distributions of surnames are relatively stable over time[10]. While this assumption may be arguable, we find strong support

---

[9]Due to Chinese-English translation, there could be more than one spellings for the same Chinese surname. Figure 7 illustrates the complexity of spelling-character mapping between English and Chinese with the example of "陈", the most common Huaren surname in our database. In our study, we aggregate Huaren surnames to the Chinese character level. See Appendix A for details.

[10]Internal migration might threat the stability of surname distributions. From 1949 to 1978, however, there is almost none internal migration due to stringent Hukou restriction. Even after China gradually loosened

from literature. Clark (2015) documents that geographic distribution of certain elite surnames is rather stable throughout history even with warfare and revolutions. Bai (2020) compares the share of each surname at the prefecture level using 2005 population Census with that using Harvard's China Biographical Database (CBDB), confirming that the two measures are highly correlated.

3.3 Foreign Firm Entry and Survival

To empirically test the effect of lineage connection on Huaren foreign firms, all of our analyses are conducted at cohort-surname-prefecture level, a highly disaggregated level. We restrict our attention to the period from 1981 to 1996, known as the Reform Era when China was constantly adopting gradual reforms to transition from planned economy to market economy, and institutions were yet to be formalized. This Reform Era also saw the Huaren firms represent a dominant share of all foreign firms (see Figure 2). Utilizing the entry and exit information in the SAIC, we select three subgroups of foreign firms: new entrants at each cohort, survival-adjusted entrants at each cohort (defined as those lasting for more than 4 years), the long term survived firms at the end of 2014. For the three subgroups, we use the surname of the Huaren legal representative to proxy a foreign firm's lineage connection, and aggregate the number of foreign firms to the cohort-surname-prefecture triplet.

We exclude four ethnic minority provinces, including Xinjiang, Tibet, Ningxia, Inner Mongolia, since surname-based lineage mostly operate among ethnic Han Chinese. We further exclude four tier-1 cities in China, including Beijing, Shanghai, Guangzhou, and Shenzhen, because the outlying economic performance of these super cities which massively attract internal migrants could be one potential confounding factor driving both our lineage connection measure and foreign firm entry and survival.

Table 2 presents the summary statistics for all the variables used for empirical analyses. Panel A shows that a lot of the variation of the presence of Huaren foreign firms is at the extensive margin given our disaggregation level. Only 5.2% of surname-prefecture-year triples has foreign firm entry while the magnitude is even smaller (0.9%) for foreign firm survival. Panel B suggests the lineage connection measure varies widely with a mean of 0.005 and a standard deviation of 0.011. Appendix C provides a visual check that our lineage connection measure is not positively correlated with population size, which is meaningful for identification because not all surnames are concentrated in large cities

---

its grip on internal migration since 1984, it only formalized free mobility of internal migrants until early 2000s (Tombe and Zhu, 2019).

proportionately. In Panel F, we can also find that Huaren foreign entrants tend to be headquarter firm rather than branch firms, labor-intensive rather than capital-intensive, followers rather than leaders in new industries. These characteristics are largely in line with descriptions of Huaren firms by past narratives (Huang, 2003). Besides, only 57.8% of Huaren firms during this period originated from HMT, suggesting that HMT entrants may not be a good proxy for Huaren entrants[11].

Appendix E illustrates the strengths of our dataset as relative to those used by prior literature, and cross-validate our dataset with other data sources.

4 Lineage Connection and Foreign Firms in China

4.1 Identification Strategy

To identify the causal effect of lineage connection on the presence of Huaren foreign firms, we estimate the following baseline specification:

$$Y_{spc} = \eta_{sp} + \theta_{sc} + \phi_{pc} + \beta \times Openness_{pc} \times m_{sp} + \lambda \times S_{s,p,c-1} + \epsilon_{spc} \qquad (2)$$

where Y_{spc} represents the outcome variables. We are primarily interested in how lineage connection affects the number of Huaren foreign entrants. Nevertheless, if the duration of a firm is transient, the welfare implication of firm entry shall be undermined. Therefore, we supplement another two measures to address the concerns: the number of survival-adjusted entrants (defined as entrants that lasted for more than 4 years), and number of survived firms by 2014.

Under all specifications, we control for surname-prefecture fixed effects η_{sp}, surname-cohort fixed effects θ_{sc} and prefecture-cohort fixed effects ϕ_{pc}. The broad set of fixed effects help us guard against a wide range of confounding factors including geographic advantages, place-based policies, and surname-specific expertise. We also include $S_{s,p,c-1}$, the number of incumbent firms one year prior to the entry cohort, in order to capture potential agglomeration benefits from incumbent firms.

The main independent variable is the interaction term $Openness_{pc} \times m_{sp}$. $Openness_{pc}$ is a time-variant measure of China's openness to foreign capital that varies across regions, which equals 1 if a prefecture p was open in year c and afterwards (see Table 1), and equals 0 otherwise. m_{sp} is the measure of lineage connection between surnames and

---

[11]See Appendix E for a detailed discussion.

Table 2: Summary Statistics

	N	Mean	Std.Dev.	Min	Max
	(1)	(2)	(3)	(4)	(5)
Panel A: Surname-Prefecture-Cohort					
At Least One Entrant	1,345,024	0.021	0.142	0	1
At Least One Survived Firm	1,345,024	0.005	0.03	0	1
Number of Entrants	1,345,024	0.060	1.437	0	1,006
Number of Survival-adjusted Entrants	1,345,024	0.052	1.234	0	795
Number of Survived Firms in 2014	1,345,024	0.009	0.181	0	32
Survived Registered Capital ($10^4 CNY$)	1,345,024	27.017	940.778	0	161,000
Survival Ratio Conditional on Entry	27,846	0.168	0.338	0	1
Number of Incumbent Firms	1,345,024	0.177	4.671	0	2,511
Panel B: Surname-Prefecture					
Lineage Connection (2005 Census)	48,179	0.005	0.011	0	0.450
Lineage Connection (SAIC Subsample)	57,802	0.005	0.012	0	0.452
Panel C: Province-Year					
Telephone Exchanger Capacity (10^4)	270	61.527	107.107	1.030	1007.100
Panel D: Prefecture					
Social Fragmentation (2005 Census)	314	0.980	0.032	0.174	1
Distance to Seaboard (KM)	314	575.776	501.694	2.360	3451.803
Special Economic Zone	314	0.022	0.041	0	1
Coastal Open City	314	0.051	0.220	0	1
GDP per capita (1996;$10^4 CNY$)	261	0.717	0.775	0.074	9.190
GDP per capita (2014;$10^4 CNY$)	282	4.935	2.877	1.017	20.015
Avg. Wage (2014;$10^4 CNY$)	273	4.823	0.989	2.721	10.340
Non-Huaren FDI (2014;$10^4 CNY$)	277	25.307	82.916	0	898.088
Number of Patents (2014;10^4)	282	0.403	0.880	0.002	7.247
Import (2017;$10^4 CNY$)	280	80.967	333.801	0	3160.789
Export (2017;$10^4 CNY$)	280	85.991	279.868	0.001	2443.579
Panel E: Surname					
EG Index (2005 Census)	233	0.062	0.319	-0.012	3.286
Panel F: Firm					
Pioneer Firm	157,875	0.155	0.362	0	1
Branch Firm	157,875	0.131	0.337	0	1
Foreign-controlled	157,875	0.469	0.499	0	1
Joint-venture	157,875	0.487	0.500	0	1
HMT Firm	157,875	0.578	0.494	0	1
Labor Intensive	92,951	0.685	0.465	0	1
Contract Intensive	138,925	0.684	0.465	0	1

Note: Panel A reports the entry and survival characteristics at surname-prefecture-cohort level based on Huaren Foreign Firms database we assemble from SAIC. Panel B summarizes the measures of lineage connection constructed from 2005 population Census and SAIC, respectively. Specifically, the latter is calculated based on the surnames of the personnel working in domestic firms established prior to 1992. The data source of Panel C is China Compendium of Statistics (2009). Variables in Panel D are retrieved from official documents or CSMAR database, except for social fragmentation which we construct with 2005 Census following Padró i Miquel, Qian and Yao (2012), and non-Huaren FDI which we sum up registered capital of non-Huaren foreign firms by 2014 using SAIC database. In Panel E, we follow Ellison and Glaeser (1997) to construct an EG index in terms of surname to capture the overall geographic concentration of a surname adjusted for its size. Panel F extracts information for the population of Huaren foreign firms whose entry cohort ranges from 1981 to 1996, from their industrial or ownership classification. A pioneer firm is defined as a first-ever entrant of a 2-digit industry in a prefecture. A firm is labor intensive if the capital/labor ratio is below the median level across all 2-digit manufacturing industries in 2004 Industrial Census. A firm is contract intensive if the contract intensity measure, proposed by Levchenko (2007) and calculated using China Input-Output Table in 2002, is above the median level across all 2-digit industries.

prefectures, defined in Equation 1. We mainly use the population Census of China in 2005 to construct this measure. But we also show that our main results are robust even if we use alternative data source to construct the measure (see section 4.3). The error term ϵ_{spc} captures all the idiosyncratic disturbances. Standard errors are clustered at surname-prefecture level.

The coefficient of interest is β, which measures the effect of lineage connection following the local openness reform. Our empirical design is essentially a DID design with differential treatment. Hence the identification assumption for the causal interpretation of our estimates to hold should be that there exist no differential effect of lineage connection before the openness shock, which can be empirically tested. To perform a formal test, we employ an event-study framework as follow:

$$Y_{spc} = \eta_{sp} + \theta_{sc} + \phi_{pc} + \sum_{\tau \in \{\leq -4, -3, -2, 0, ...3, \geq 4\}} \beta^{\tau} \times Openness_{pc}^{\tau} \times m_{sp} + \lambda \times S_{spc} + \epsilon_{spc} \quad (3)$$

where $\tau = c - c_p$, referencing the time window as relative to the incidence of openness for the prefecture p as we use c_p the year at which openness shock occurs in prefecture p. $Openness_{pc}^{\tau} \times m_{sp}$ are indicator variables which equal one if year c is τ years away from openness shock and 0 otherwise for prefectures. The omitted group is $\tau = -1$, hence all estimates of β^{τ} should be interpreted as relative to the period one year prior to the openness shock. For our DID design to hold, we should expect β^{τ} for all $\tau < 0$ statistically insignificant.

To further shed light on the underlying mechanisms, we organize entrants counts to be by cohort-prefecture-surname-type level, where type indicates the "fourth" dimension that teases out differential responses to the lineage connection. In analogy to Equation 2, we estimate

$$Y_{spc} = \eta_{sp} + \theta_{sc} + \phi_{pc} + \alpha_x + \beta_1 \times Openness_{pc} \times m_{sp} + \beta_2 \times Openness_{pc} \times m_{sp} \times type_x + \lambda \times S_{s,p,c-1} + \epsilon_{spc}$$
$$(4)$$

where $type_x$ is a dummy variable for firm type x. The sign of the estimate of β_2 determines whether the type x of Huaren entrant is complementary or substitute to the the presence of lineage connection. If $type_x$ is informative of some functions of lineage network, then we can uncover the specific roles of lineage connections have played in producing the our baseline estimates.

4.2 Baseline Results

Table 3: The Effect of Lineage Connection on Entry and Survival of Huaren Foreign Firms

	Number of Entrants	Number of Survival-adjusted Entrants	Number of Survived Firms in 2014	Survival Ratio Conditional on Entry in 2014
	(1)	(2)	(3)	(4)
Mean of Dep. Var.	0.060	0.052	0.009	0.168
Openness × Lineage Connection	1.767***	1.517***	0.218*	-0.333
	(0.574)	(0.527)	(0.399)	(0.825)
Adj.R^2	0.570	0.555	0.399	0.364
N	1,344,421	1,344,421	1,344,421	20,633
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included).

We first estimate Equation 2 to evaluate how lineage connection affects the entry and survival of Huaren foreign firms following the local openness shock, and present results in Table 3. In all columns, we control for surname-prefecture, surname-cohort and prefecture-cohort fixed effects, as well as number of incumbent firms at surname-prefecture-cohort level. The coefficients of our interested variables remain both economically and statistically significant. The results in column (1) suggest that a 10% increase in the concentration of a particular surname in a prefecture can leads to approximately twice as many more Huaren entrants of that surname in a year following the openness shock during the Reform Era. Column (2) confirms that not all entrants are short-lived, by estimating a slightly smaller yet still significant effect on counts of survival-adjusted entrants. Even if we measure firm counts by survival in 2014, as in column (3), we still arrive at significant results. Column (1) to (3) altogether suggest that lineage connection facilitates the presence of Huaren foreign firms in both short run and long run. As a crucial check on the identification assumption, we also use the event-study specification and show in Figure 4 that the entry and survival of Huaren foreign firms across surnames and prefectures had similar trends prior to the openness shocks.

Column (4) additionally looks at the effect of lineage connection on survival ratio conditioning on entry, in order to separate the benefits of lineage with respect to different

stages of a firm's life cycle. Although we find effects of lineage connection on the counts of survived firms, we did not discover benefits of lineage connection on survival probability of Huaren foreign firms after their establishment, which suggests the mechanisms through which lineage takes effects are mainly lowering entry barrier. Also, the results overrule an alternative hypothesis which points to the potential selection of Huaren firms with higher latent quality into location with higher lineage connections—if only firms of high quality can take advantage of the lineage connection, the sorting of more capable Huaren entrepreneurs into surname-prefecture pair with high lineage connection could possibly be conducive to both entry and survival, leading to an overestimation of the effects of lineage connection. The results of column (4) motivate us to interpret the role of the lineage connection as lowering entry barrier in the short run regardless of Huaren firm's latent quality.

4.3 Robustness Checks

In this subsection, we test our results to a variety of robustness checks.

Alternative Dependent Variables. We note that the entry and survival of Huaren foreign firms are pretty scare at surname-prefecture-cohort level (see Table 2). Naturally we want to at first make sure our results still hold if we consider the extensive margin. We replace our firm counts measures with dummies indicating whether there is at least one firm present in a surname-prefecture-cohort triplet. The column (1) to (3) of Table 4 show our findings that the lineage connection significantly increases the entry of survival of Huaren foreign firms at the extensive margin as well. We are also concerned with the economic implications of firm entries since they are not directly related to welfare measures. While the SAIC database does not provide many variables for firm-level activities except for registered capital, Bai et al. (2020) and we in Appendix F both find that firm-level registered capital is strongly correlated with assets, employment and output, even with inclusion of firm fixed effects, using a samples of manufacturing firms from Annual Survey of Industrial Firms (ASIF). We transforms the total registered capital of all survived Huaren foreign firms in 2014 with an inverse hyperbolic function[12] and rerun the regressions. Consistently, the results of column (4) of Table 4 imply that in the long run, stronger lineage connection help introduce more foreign capital, thus probably producing more output, and providing more jobs.

---

[12]In practice, hyperbolic function has been increasingly used to replace natural logarithmization (Bellemare and Wichman, 2020). It has two properties: (1) it approximates the natural logarithm of the transformed variable and (2) allows retaining zero-valued observations.
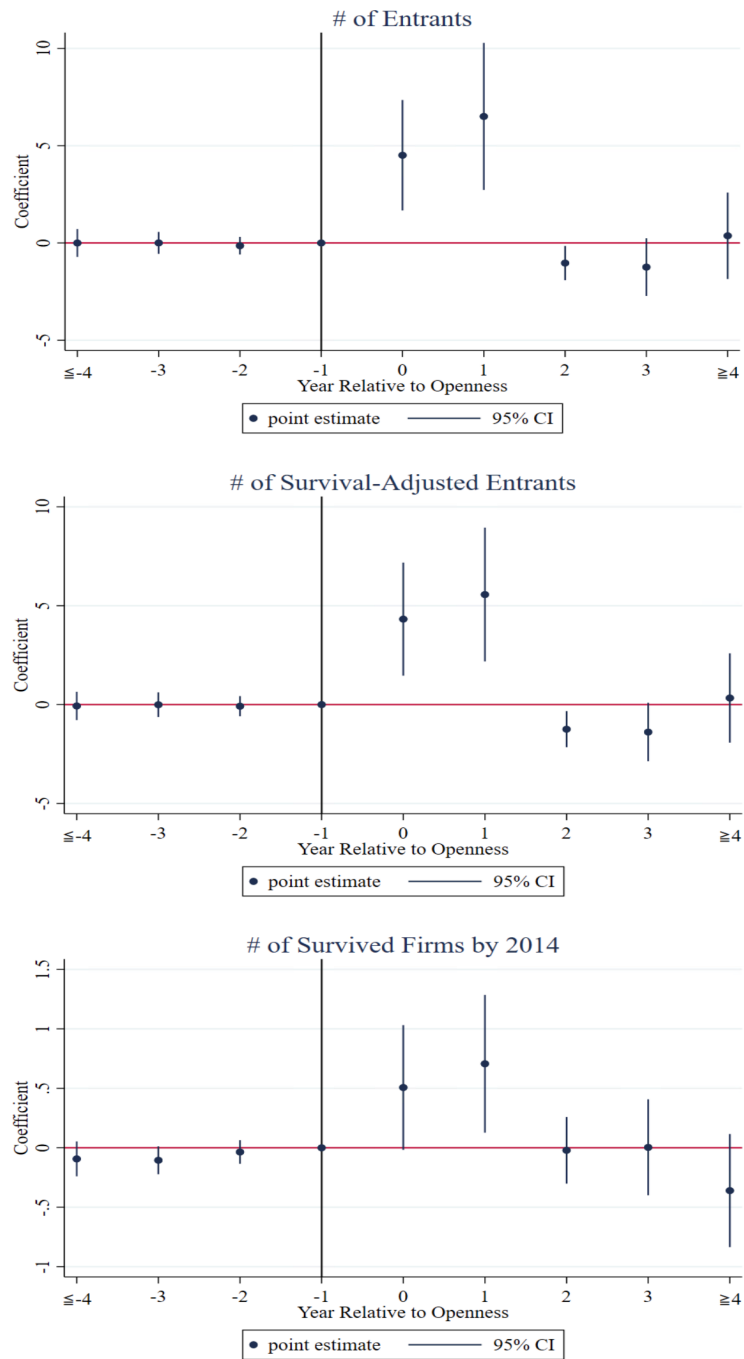
Figure 4: Event Study: Estimates of β^{τ} in Equation 3

Table 4: Alternative Dependent Variables

	At Least One Entrant	At Least One Survival-adjusted Entrants	At Least One Survived Firm in 2014	Arcsinh (Total Registered Capital in 2014)
	(1)	(2)	(3)	(4)
Openness × Lineage Connection	0.766***	0.672***	0.165***	1.229***
	(0.112)	(0.102)	(0.053)	(0.422)
Adj.R^2	0.383	0.373	0.255	0.271
N	1,344,421	1,344,421	1,344,421	1,344,421
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included). "Arcsinh" in column (4) is the inverse hyperbolic function, which converges to logarithmizing transformation yet remain meaningful for zero-valued observations.

Alternative Lineage Connection Measure. The second potential challenge to our results is the measurement error in our lineage connection measure. 2005 Census is the only data source where we could observe nationally representative distributions of surnames. Since our sample period starts from the 1980s, it is possible that the internal migration between the 1980s and 2005 could lead to changes in the spatial distribution of surnames. We gain confidence in our measure from past literature showing that the spatial distribution of surnames does not vary by much throughout the history, even with warfare and revolutions (Clark, 2015). Recent work by Bai (2020) also checks the stability of spatial distribution of surnames by comparing the surname distributions constructed from 2005 Census data and the Harvard China Biographical Database (CBDB), and finding a strikingly high similarity. We provide additional checks on our results taking advantage of the surname information available in our SAIC database. We construct an alternative lineage connection measure using surnames of individuals who are personnel members of domestic firms in SAIC database. We further restrict the samples to those who have worked in domestic firms established before 1992, to mitigate concerns on internal migration.

The results are presented in Panel A of Table 5. We still get significant estimates for most of our outcome variables as before and they remain economically large relative to the mean. The only difference is that we now do not have a significant estimate for the number of long run survived firms though the coefficient remains positive. Note that our

alternative lineage connection measure is constructed based on a non-representative sample of the population (entrepreneurs and managers in domestic firms), so we put more faith in our original results relative to the alternative results. We can, however, still reach the same conclusion as before that lineage connection exerts not only short run but also long run positive impacts on the entry of Huaren foreign firms.

Alternative Standard Error. Our baseline results cluster standard errors at surname-prefecture level to account for serial heteoskedasticity. However, the error terms could be correlated at a higher level. The results in the Panel B of Table 5 show that the significance of our estimates does not change by much even if we cluster our results at a higher prefecture level.

Subsample Regressions. The fourth challenge is the possibility that variations from certain regions or surnames versus the others could totally drive our findings. For example, emigration-intensive provinces like Guangdong, Fujian, Zhejiang, disproportionately exported more emigrants and attracted more foreign firms. And some surnames, whose population historically resided in emigration-intensive provinces, have persistently higher presence in both FDI-intensive regions and among Huaren group, due to surname-specific expertise. Although it does not qualify as an identification issue, it certainly challenges the generality of our findings. The coefficients for number of entrants and survival-adjusted number of entrants remian positive and statistically significant While the significance of the coefficients for the number of survived firms in 2014 varies from case to case, from Panel C to Panel F in Table 5, where we use different subsample regressions to test the robustness of our baseline findings. Panel C and Panel E together suggests that the lineage connection take effects not only within but also outside of emigration-intensive provinces (though the magnitude of the coefficients within emigration-intensive provinces are indeed larger). Likewise, Panel D verifies the robustness of results to the exclusion of emigration-intensive surnames. And finally Panel F overrules the possibility that outliers—FDI-intensive prefectures drive our observed patterns by excluding them from our regression samples.

Permutation Tests. To further check the extent to which our results are influenced by any measurement errors in the lineage connection measures, we perform two permutation tests by randomly shuffling our lineage connection measures within and across prefectures. Figure 5 and Figure 6 plot the distributions of the estimates for our outcome variables, from 100 simulations with reshuffled lineage connection of different Surnames within a prefecture and of same surnames across prefectures, respectively. All distributions are centered around zero and the actual estimates tend to lie at the far right end of

Table 5: Robustness Checks

	Number of Entrants	Number of Survival-adjusted Entrants	Number of Survived Firms in 2014	Survival Ratio Conditional on Entry in 2014
	(1)	(2)	(3)	(4)
Panel A: Alternative Lineage Connection Measure				
Openness × Lineage Connection (SAIC)	2.874***	2.511***	0.463	-0.111
	(0.946)	(0.903)	(0.403)	(0.537)
Adj.R^2	0.612	0.596	0.445	0.373
N	539,376	539,376	539,376	18,324
Panel B: Alternative Standard Error				
Openness × Lineage Connection	1.767***	1.517***	0.218*	-0.333
	(0.493)	(0.456)	(0.119)	(0.735)
Adj.R^2	0.570	0.538	0.247	0.389
N	1,162,898	1,162,898	1,162,898	10,236
Panel C: Excluding Emigration-intensive Provinces				
Openness × Lineage Connection	0.642***	0.482***	0.039	-0.015
	(0.233)	(0.167)	(0.050)	(0.496)
Adj.R^2	0.570	0.538	0.247	0.389
N	1,162,898	1,162,898	1,162,898	10,236
Panel D: Excluding Emigration-intensive Surnames				
Openness × Lineage Connection	1.733***	1.480***	0.258***	0.023
	(0.332)	(0.292)	(0.092)	(0.884)
Adj.R^2	0.409	0.394	0.208	0.294
N	1,250,024	1,250,024	1,250,024	1,250,024
Panel E: Within Emigration-intensive Provinces				
Openness × Lineage Connection	6.001***	5.152**	0.694	-0.953
	(2.476)	(2.186)	(0.532)	(1.261)
Adj.R^2	0.690	0.682	0.491	0.337
N	181,447	181,447	181,447	9,678
Panel F: Excluding FDI-intensive Prefectures				
Openness × Lineage Connection	1.534***	1.308***	0.172**	-1.870
	(0.440)	(0.380)	(0.072)	(1.373)
Adj.R^2	0.451	0.435	0.230	0.400
N	1,272,091	1,272,091	1,272,091	11,858
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included). Emigration-intensive provinces include Guangdong, Fujian, Zhejiang. Emigration-intensive surnames refer to the 20 most populous surnames among all Huaren legal representatives from 1981 to 2014. FDI-intensive prefectures are those whose cumulative number of hosted foreign firms rank top 20 among all prefectures during 1981 to 2014.

the distributions. Taken together, these observations suggest that the positive and significant effects of the lineage connection on entry and survival of Huaren foreign firms are not driven by unobserved factors.
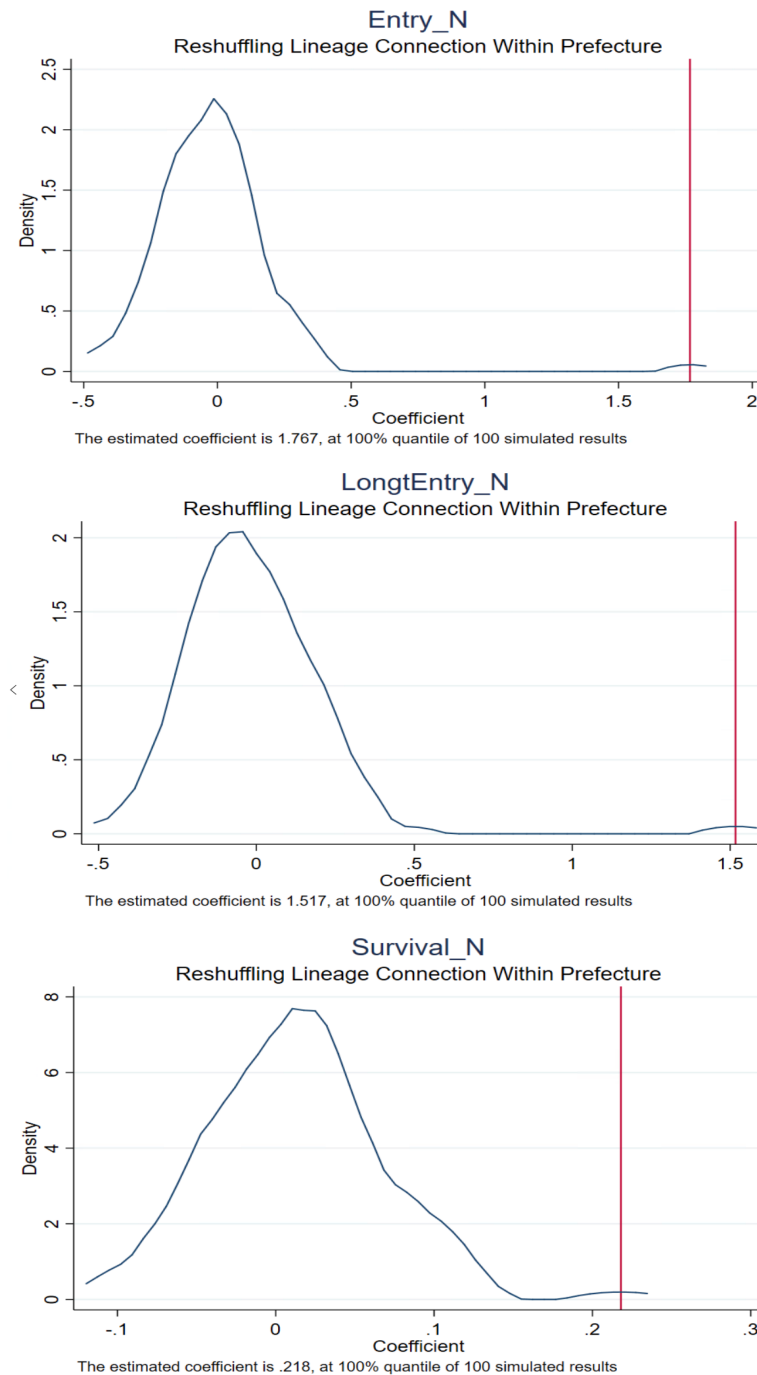

Figure 5: Permutation Test: Estimates of β in Equation 2 after Reshuffling Lineage Connection of Different Surnames within a Prefecture.
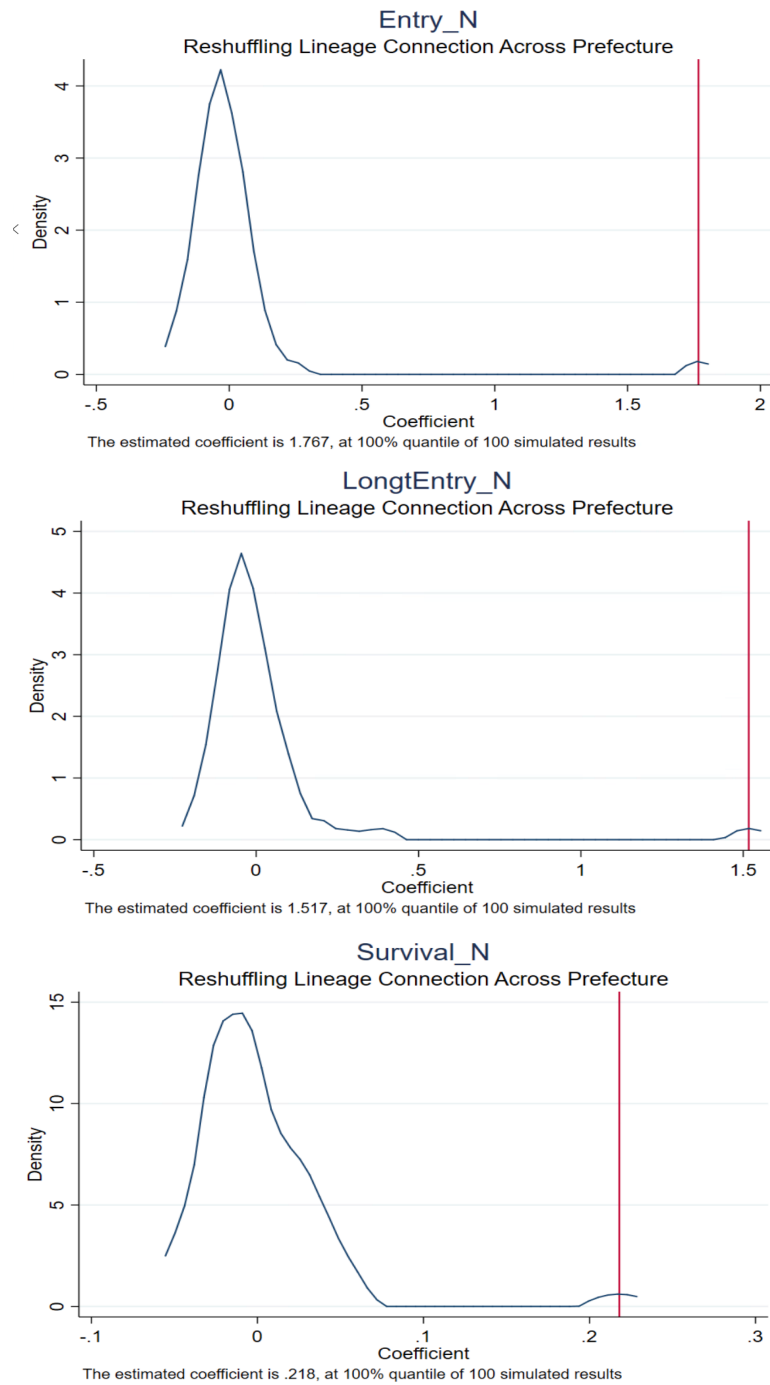
Entry_N
Reshuffling Lineage Connection Across Prefecture

The estimated coefficient is 1.767, at 100% quantile of 100 simulated results

LongtEntry_N
Reshuffling Lineage Connection Across Prefecture

The estimated coefficient is 1.517, at 100% quantile of 100 simulated results

Survival_N
Reshuffling Lineage Connection Across Prefecture

The estimated coefficient is .218, at 100% quantile of 100 simulated results

Figure 6: Permutation Test: Estimates of β in Equation 2 after Reshuffling Lineage Connection of Same Surnames across Prefectures.

4.4 Mechanisms

In this section, we explore the underlying mechanisms through which the lineage connection exerts the observed effects on the entry and survival of Huaren foreign firms. As summarized in Rauch (2001), any kinds of social networks in business play two major roles: sharing information and enforcing contracts. We discover evidence in support of both roles.

Overcoming Information Frictions. We hypothesize that the information frictions must be a critical barrier, since barely any foreign investors are allowed to pay business visits, let alone invest in mainland China before 1979, making China a vacuum of business information for 30 years. Huaren investors, however, remain socially in touch with their lineage members—anecdotal narratives reveal that they communicated via specially-authorized telephones and letters even during the time when communication infrastructure in China was in infancy (Tan, 2006). Lineage connection is inherently a kind of social tie and local information can be transmitted from the local lineage members through the social tie to the Huaren entrepreneurs abroad, which helps overcome the information frictions.

To empirically test our hypothesis, first, we assess the heterogeneous effects of lineage connection on Huaren foreign firm across regions and time with differential development of communication infrastructure. We include an additional term by interacting our main independent variable with province-year level of telephone exchanger capacity. Panel A of Table 6 show our results. The coefficients of the new interaction terms remain positive and significant for the number of entrants and survival-adjusted number of entrants. These findings are in line with expectation that communication infrastructure are most valuable and complementary with lineage connection upon entry since Huaren can use them to access information through local contacts following the openness reform. As information infrastructure, such as satellite telephone and Internet, has been constantly improved over time in China, the relative importance of telephone exchanger capacity wanes, so we do not find significant coefficients for the number of survived firms in the long run.

Likewise, we discover stronger effects of lineage connection in prefectures of higher social fragmentation as measured by one minus the HHI in terms of surnames. According to Padró i Miquel, Qian and Yao (2012), information access becomes harder at the presence of higher social fragmentation, and thus making lineage connection relatively more valuable. Panel B confirms that claim.

Third, we organize our data to the surname-prefecture-cohort-pioneer firm level as

Equation 4. A pioneer firm refers to a first entrant in a 2-digit industry in a prefecture. Panel C demonstrates that Huaren firms of stronger lineage connection tend to enter into existing industries rather than new industries, as relative to others. This suggests that information on proven business opportunities passes through lineage group, inducing imitators to enter. Since we focus on the Reform Era when most industries in China just began to take root, our findings are consistent with observations that entry of imitators is most exuberant during early stage of an industry life cycle (Gort and Klepper, 1982).

Table 6: Overcoming Information Frictions

	Number of Entrants	Survival-adjusted Number of Entrants	Number of Survived Firms in 2014	Survival Ratio Conditional on Entry in 2014
	(1)	(2)	(3)	(4)
Panel A: Combining Soft and Hard Infrastructure				
Openness × Lineage Connection	-2.507	-1.814	-0.355	0.223
	(1.908)	(1.846)	(0.765)	(1.404)
Openness × Lineage Connection × Telephone Exchanger Capacity	1.074***	0.868**	0.161	-0.246
	(0.409)	(0.386)	(0.184)	(0.193)
Adj.R^2	0.454	0.437	0.244	0.317
N	666,397	666,397	666,397	8,825
Panel B: Dealing with fragmented Information Sources				
Openness × Lineage Connection	-9.356***	-7.201***	-0.407	8.830***
	(2.936)	(2.899)	(1.110)	(3.015)
Openness × Lineage Connection × Surname Fragmentation	11.963***	9.365***	0.719	-10.014
	(3.336)	(2.899)	(0.399)	(3.900)
Adj.R^2	0.409	0.395	0.208	0.295
N	1,186,664	1,186,664	1,186,664	12,149
Panel C: Discovering Business Opportunities				
Openness × Lineage Connection	7.424**	6.390**	0.702**	-0.568
	(3.063)	(2.712)	(0.282)	(0.823)
Openness × Lineage Connection × Pioneer Firm	-13.080**	-11.263**	-1.186**	-0.073
	(5.650)	(4.974)	(0.477)	(0.404)
Adj.R^2	0.465	0.458	0.321	0.554
N	2,690,048	2,690,048	2,690,048	24,936
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included). Regressions are conducted at surname-prefecture-cohort-type level and additional controls refer to a dummy for pioneering firm (first entrant in a 2-digit industry in a prefecture) in Panel C.

Securing Property Right and Facilitating Contract Enforcement. Lineage connection, as a form of informal institution, can potentially fill the absence of formal institution in two aspects: securing property right and facilitating contract enforcement. Panel A of Table 7 finds that Huaren foreign firms with higher lineage connection are less likely to

enter and survive as a branch firm. This reveals that they may be less concerned with property rights since entering as branches are more flexible and associated with weaker commitments to a particular location.

Doing overseas business involves negotiating over two major kinds of contracts, that is, one with local employees, and the other with local suppliers. We hypothesize that lineage connection may facilitate contract enforcement with respect to both parties. To formally test the hypothesis, we further breakdown entrants into labor-intensive industries and non-labor-intensive industries (for manufacturing subsamples only), intermediate-goods-contract-intensive industries and non-intermediate-contract-goods-intensive industries. Results in Panel B and Panel C indicate that Huaren foreign firms with stronger lineage connection have comparative advantage in labor-intensive and intermediate-goods-contract-intensive industries, which suggest lineage connection helps mediate access to labor and intermediate inputs through smoothing contract enforcement.

5 Lineage-driven FDI Spillovers

The previous sections have established the causality between Huaren's lineage connection and the entry of Huaren foreign firms during the take-off stage of Chinese economy. This section, taking one further step, asks whether these industrial "seeds"—the lineage-driven FDI—had a long-lasting effect on present-day economic development.

5.1 Isolating Lineage-driven FDI

There is a large literature documenting the spillovers from the presence of FDI on local economy (Harrison and Rodríguez-Clare, 2010). While it may be difficult to come up with causal estimates of these spillovers due to the endogenous choice of foreign firms, several attempts have been to identify the FDI spillovers in China. In particular, Lu, Tao and Zhu (2017) exploits sector-level variations from a policy that deregulates inward FDI in 2002, and finds that policy-induced FDI competed down the productivity of domestic firms at sector-level in China. Ma (2018) adopts a regression discontinuity design (RDD) around the geographical borders of Chinese linguistic dialect zones to isolate exogenous change of culture-specific FDI, identifying positive spillovers of FDI on domestic firms.

We advance this literature by employing a new strategy that is motivated by Sequeira, Nunn and Qian (2020). Specifically, We draw results from estimating our baseline specification as in Equation 2 (the "zero stage regression") and isolate lineage-driven FDI that

Table 7: Securing Property Right and Facilitating Contract Enforcement

	Number of Entrants	Survival-adjusted Number of Entrants	Number of Survived Firms in 2014	Survival Ratio Conditional on Entry in 2014
	(1)	(2)	(3)	(4)
Panel A: Securing Property Right				
Openness × Lineage Connection	8.795***	7.633**	1.034***	-0.258
	(3.363)	(2.997)	(0.379)	(0.833)
Openness × Lineage Connection × Branch Firm	-15.823***	-13.749***	-1.850***	-0.961
	(6.255)	(5.547)	(0.676)	(0.936)
Adj.R^2	0.452	0.444	0.300	0.566
N	2,690,048	2,690,048	2,690,048	23,274
Panel B: Facilitating Labor Contracts				
Openness × Lineage Connection	-1.864*	-1.650*	-0.193*	-0.922
	(1.043)	(0.928)	(0.099)	(1.284)
Openness × Lineage Connection × Labor Intensive	4.990**	4.433**	0.535**	-0.031
	(2.466)	(2.226)	(0.257)	(0.328)
Adj.R^2	0.665	0.645	0.345	0.567
N	2,690,048	2,690,048	2,690,048	18,799
Panel C: Facilitating Input Contracts				
Openness × Lineage Connection	-2.331*	-2.030**	-0.256**	-0.161
	(1.156)	(1.010)	(0.104)	(0.814)
Openness × Lineage Connection × Intermediate Goods Contract Intensive	6.196**	5.397**	0.693***	-0.619**
	(2.736)	(2.427)	(0.258)	(0.292)
Adj.R^2	0.688	0.673	0.376	0.545
N	2,690,048	2,690,048	2,690,048	25,245
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included). Regressions are all conducted at surname-prefecture-cohort-type level and additional controls refer to a dummy for each specific type in each panel. Labor-intensive industries are those whose capital to labor ratio lies below the median level across all two-digit manufacturing industries according to 2004 Industrial Census. Intermediate-goods-contract-intensive industries are those whose sector-level HHI of input values from all upstream sectors according to Input-Output Table of China in 2002 is below the median level, following the methodology of Levchenko (2007).

arises from the interaction of surname distribution and the gradual openness across Chinese prefectures. The benefit of combining the two sources of variation is that the interaction between the two produces variation that is unlikely to affect our contemporary outcomes of interest other than through historical FDI inflows to the prefecture. The empirical analysis takes following steps:

$$\widetilde{Entry_{spc}} = \hat{\beta} \times Openness_{pc} \times m_{sp} \tag{5}$$

$$\overline{Entry_p} = \sum_{s,c=1981/1996} \widetilde{Entry_{spc}} \tag{6}$$

$$\hat{FDI}_p = \alpha \times \overline{Entry_p} + \lambda \times X_p + \epsilon_p \tag{7}$$

$$\pi_p = \alpha \times \hat{FDI}_p + \lambda \times X_p + \epsilon_p \tag{8}$$

First, $\hat{\beta}$ in Equation 5 is estimated from our baseline results using number of Huaren entrants as dependent variable, according to Equation 2. Then we calculate the predicted number of Huaren entrants $\widetilde{Entry_{spc}}$ for each surname-prefecture-cohort triplets, and aggregate them from 1981 to 1996 to the prefecture level to get $\overline{Entry_p}$ as in Equation 6. The predicted number of Huaren entrants, by construction, is orthogonal to geographic characteristics, entrepreneurial culture, place-based policies, surname-specific advantages, and so on. We therefore can use it as a valid Instrument to predict lineage-driven FDI $\hat{FDI}_p$, as in Equation 7. We finally assess the long run spillovers of lineage-driven FDI on a variety of economic outcomes π_p in the second stage in Equation 8, controlling for a bunch of prefecture-level characteristics.

Compared to past literature on identifying FDI spillovers in China, our paper make several significant contributions: First, we extend the research periods to the early 1980s when the spillover effects of early FDI in China—the "industrial seeds"—tended to be strongest. In contrast, most existing studies only focus on late 1990s to mid 2000s, due to data limitations. Second, our paper draws upon the most comprehensive dataset on FDI covering all sectors while past literature mainly use above-scale manufacturers[13]. Last but not least, our paper establishes the causal link between FDI inflow and economic development across space in the long run, which helps account for divergent growth paths across Chinese regions.

---

[13]See Appendix E for a detailed discussion.

5.2 Identifying Long Run Spillovers

As FDI literature has pointed out, FDI inflows into developing countries unnecessarily promise economic gains in the long run. Indeed, FDI has been found successful in generating growths, importing technology, creating jobs and raising wages. But FDI also adds to the competition faced by the domestic firms, and thus may discourage the domestic sectors to develop (Harrison and Rodríguez-Clare, 2010; Lu, Tao and Zhu, 2017). In this subsection, we examine whether the lineage-driven FDI—the dominant form of inward FDI during the period when China's institutional environment was largely immature—has any long term effect on the current economic development.

To provide answers, we implement the methodology described in Section 5.1. We first calculate the our instrument—the predicted number of entrants $\overline{Entry_p}$ based on previous estimation results. Next, We sum up the registered capital of Huaren foreign firms of the 2014 cross-section, to construct lineage-driven FDI at the prefecture level $\hat{FDI}_p$, and use it as dependent variable in the first stage (Equation 7). And then we perform second-stage regressions for a series of long-run economic indicators, including GDP per capita, number of patents, average wage, non-Huaren FDI, import and export. In all specifications, we additionally control for initial economic development as proxied by GDP per capita in 1996, years since openness reform, distance to sea, social fragmentation, SEZ dummy, Coastal Open City dummy, province fixed effects.

The results are presented in Table 8. We find that prefectures with more historical lineage-driven FDI today has higher economic output, more innovations, higher wage level, more import and export. These findings replicate the findings of previous literature that FDIs are successful in generating output, diffusing technology, and facilitating international trade. Notably, column (5) provides striking new evidence that even non-Huaren FDI inflows are higher in prefectures with more historical lineage-driven FDI. While the underlying mechanisms are yet to be understood, this finding is interesting because lineage-FDI, supposedly to rely on an informal network, can be beneficial to the building of formal institution in the long run (Du, Lu and Tao, 2008).

6 Conclusions

The lineage connections between Huaren foreign firms and their ancestral hometowns are believed to be one of the important drivers for China's remarkable industrial growth over the last forty years. Our paper documents the prevalence of Huaren foreign firms during the take-off period of Chinese economy and causally identifies the effect of lin-

Table 8: Long Run Spillovers of Lineage-driven FDI

	Log GDP per capita (2014)	Log Number of Patents (2014)	Log Average Wage (2014)	Log Non-Huaren FDI (2014)	Log Import (2017)	Log Export (2017)
	(1)	(2)	(3)	(4)	(5)	(6)
Log Huaren FDI	0.704***	0.937***	0.029*	1.139***	1.047***	0.962***
	(0.138)	(0.184)	(0.015)	(0.258)	(0.246)	(0.225)
Anderson-Rubin Wald Test F Statistics	45.543***	34.207***	3.413*	15.111***	12.675***	14.482***
N	242	242	238	226	241	241
Controls	Y	Y	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Robust standard errors are shown in parenthesis. Non-Huaren FDI are constructed summping up the registered capital of foreign firms active in 2014 whose legal representatives are not Huaren. Controls include log GDP per capita in 1996, years since openness reform, distance to sea, social fragmentation, SEZ dummy, Coastal Open City dummy, province FEs.

eage connection on the entry and survival of foreign firms, leveraging a unique dataset containing the universe of Huaren foreign firms and the surnames of the Huaren legal representatives'. We show that the effects are primarily driven by reduction in the information frictions and contracting costs, thus lowering the entry barriers for the connected Huaren foreign firms.

This paper contributes to an improved understanding of the FDI policy for the developing countries. While opening up has numerous potential benefits for developing countries, it is often hard for them to kick start the virtuous cycle. Lacking market information and having an immature domestic environment fraught with information frictions, non-trade barriers, corruption and bureaucrat misconduct, developing countries often struggle setting up exporting channels, attracting foreign direct investments, let alone fully enjoy the benefits of openness. Findings of this paper may provide a practical solution: to mobilize the diaspora networks to overcome the frictions in information access and imperfections of market institution, during the early stage of development. We illustrate this point using the case of China: Many local governments in China realized during very early periods the value of the region's lineage connection with overseas Chinese. They enacted policies, and coordinated through specialized organizations to encourage overseas Chinese businessmen to invest back in their ancestral hometowns. With the help of these international ethnic ties, China' share in international trade volumes and FDI inflows rocket up from literally zero to world-leading levels, even before accession to WTO.

We also find that these early lineage-driven FDI exerts positive spillovers on economic development and even facilitates the transition toward formal institution in the long run. Although the sign and magnitude of FDI spillovers are generally inconclusive in different contexts, our empirical evidence suggest that the benefits of introducing FDI as industrial seeds seem to outweigh its costs during the take-off stage of an developing economy.

References

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do immigrants assimilate more slowly today than in the past?" *American Economic Review: Insights*, 2(1): 125–141.

Amiti, Mary, and Beata Smarzynska Javorcik. 2008. "Trade costs and location of foreign firms in China." *Journal of Development Economics*, 85(1-2): 129–149.

Bai, Chong-En, Chang-Tai Hsieh, Zheng Michael Song, and Xin Wang. 2020. "Special Deals from Special Investors: The Rise of State-Connected Private Owners in China." National Bureau of Economic Research.

Bai, Ying. 2020. "Surname Distance and Technology Diffusion: The Case of the Adoption of Maize in Late Imperial China." *Manuscript.*

Bazzi, Samuel, Martin Fiszbein, and Mesay Gerresilasse. 2020. "Frontier culture: The root and persistence of "rugged individualism" in the United States." *Econometrica*, 88(6): 2329–2368.

Bellemare, Marc F., and Casey J. Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics*, 82(1): 0305–9049.

Branstetter, Lee, and Nicholas Lardy. 2006. "China's embrace of globalization." 12373.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan. 2019. "Migrants, Ancestors, and Foreign Investments." *The Review of Economic Studies*, 86(4): 1448–1486.

Cheng, Leonard K, and Yum K Kwan. 2000. "What are the determinants of the location of foreign direct investment? The Chinese experience." *Journal of international economics*, 51(2): 379–400.

Chen, Yuyu, Mitsuru Igami, Masayuki Sawada, and Mo Xiao. 2020. "Privatization and productivity in china." *Manuscript, available at SSRN 2695933.*

Clark, Gregory. 2015. *The son also rises: Surnames and the history of social mobility.* Princeton University Press.

Dai, Ruochen, Dilip Mookherjee, Kaivan Munshi, and Xiaobo Zhang. 2019. "The community origins of private enterprise in China." Boston University-Department of Economics.

Du, Julan, Yi Lu, and Zhigang Tao. 2008. "Economic institutions and FDI location choice: Evidence from US multinationals in China." *Journal of comparative Economics*, 36(3): 412–429.

Ellison, Glenn, and Edward L Glaeser. 1997. "Geographic concentration in US manufacturing industries: a dartboard approach." *Journal of political economy*, 105(5): 889–927.

Gort, Michael, and Steven Klepper. 1982. "Time paths in the diffusion of product innovations." *The Economic Journal*, 92: 630–653.

Greif, Avner. 1993. "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition." *The American economic review*, 525–548.

Greif, Avner, and Guido Tabellini. 2017. "The clan and the corporation: Sustaining cooperation in China and Europe." *Journal of Comparative Economics*, 45: 1–35.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural biases in economic exchange?" *The Quarterly Journal of Economics*, 124(3): 1095–1131.

Harrison, Ann, and Andrés Rodríguez-Clare. 2010. "Trade, foreign investment, and industrial policy for developing countries." *Handbook of Development Economic*, 5: 4039–4214.

Huang, Yasheng. 2003. *Selling China: Foreign direct investment during the reform era.* Cambridge University Press.

Huang, Yasheng, Li Jin, and Yi Qian. 2013. "Does ethnicity pay? Evidence from overseas Chinese FDI in China." *Review of Economics and Statistics*, 95(3): 868–883.

Javorcik, Beata S, Çağlar Özden, Mariana Spatareanu, and Cristina Neagu. 2011. "Migrant networks and foreign direct investment." *Journal of Development economics*, 94(2): 231–241.

Jiang, Kun, Wolfgang Keller, Larry D Qiu, and William Ridley. 2018. "International joint ventures and internal vs. external technology transfer: Evidence from China." National Bureau of Economic Research.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou. 2020. "Social proximity to capital: Implications for investors and firms." National Bureau of Economic Research.

Kuhn, Philip A. 2008. *Chinese among others: Emigration in modern times.* Rowman & Littlefield.

Levchenko, Anderi A. 2007. "Institutional quality and international trade." *Review of Economic Studies*, 74: 1–35.

Lin, Ping, Zhuomin Liu, and Yifan Zhang. 2009. "Do Chinese domestic firms benefit from FDI inflow?: Evidence of horizontal and vertical spillovers." *China economic review*, 20(4): 677–691.

Lucas, Robert E. 1990. "Why doesn't capital flow from rich to poor countries?" *The American Economic Review*, 80(2): 92–96.

Lu, Yi, Zhigang Tao, and Lianming Zhu. 2017. "Identifying FDI spillovers." *Journal of International Economics*, 107: 75–90.

Ma, Sen. 2018. "The Effects of Cultural Similarity on Foreign Direct Investment and Productivity of Domestic Firms: Identification from Borders of Chinese Dialect Zones." *Manuscript.*

Melitz, Jacques, and Farid Toubal. 2014. "Native language, spoken language, translation and trade." *Journal of International Economics*, 93(2): 351–363.

Nanda, Ramana, and Tarun Khanna. 2010. "Diasporas and domestic entrepreneurs: Evidence from the Indian software industry." *Journal of Economics & Management Strategy*, 19(4): 991–1012.

Padró i Miquel, Gerard, Nancy Qian, and Yang Yao. 2012. "Social fragmentation, public goods and elections: Evidence from China." National Bureau of Economic Research.

Parsons, Christopher, and Pierre-Louis Vézina. 2018. "Migrant networks and trade: The Vietnamese boat people as a natural experiment." *The Economic Journal*, 128(612): F210–F234.

Rauch, James E. 2001. "Business and social networks in international trade." *Journal of Economic Literature*, 39(4): 1177–1203.

Rauch, James E, and Vitor Trindade. 2002. "Ethnic Chinese networks in international trade." *Review of Economics and Statistics*, 84(1): 116–130.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian. 2020. "Immigrants and the Making of America." *The Review of Economic Studies*, 87(1): 382–419.

Tan, Chee-Beng. 2006. *Chinese transnational networks.* Routledge.

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, migration, and productivity: A quantitative analysis of china." *American Economic Review*, 109(5): 1843–72.

Vogel, Ezra F. 1990. *One step ahead in China: Guangdong under reform.* Vol. 2, Harvard University Press.

Wang, Jin. 2013. "The economic impact of Special Economic Zones: Evidence from China." *Journal of Development Economics*, 101: 133–147.

Xu, Chenggang, and Xiaobo Zhang. 2009. *The evolution of Chinese entrepreneurial firms: Township-village enterprises revisited.* Vol. 854, Intl Food Policy Res Inst.

Appendix

A   Algorithm to Identify Huaren and Their Surnames

This section introduces the algorithm we use to identify Huaren and their Chinese surnames among the personnel working in foreign companies. We extract from SAIC database all the personnel working in foreign firms according to the firm's ownership code, with following variables available for each person: names, ID types, registry address, executive position, a dummy for the legal representative status.

Before formally executing our algorithm, we perform a preparatory cleaning procedure to leave out all symbols (such as comma, periods, semicolon) and word content that is unrelated to names (such as titles, "Mr.", "Dr.", "appointed by the parent company"), since many name entries are unstructured in the raw data we acquire. This step breaks down the raw variable "name" into strings of pure Chinese or English characters. Thus name strings can be further categorized into three kinds: names written in pure Chinese (李小龙), names written in pure English (Bruce Lee), and names written in both Chinese and English (李小龙 Bruce Lee).

It is fairly straightforward to identify the Chinese surnames for name strings written in pure Chinese and in both Chinese and English, because both groups contain Chinese content of names. But it is relatively hard to identify surnames for the names written in English because English spellings of Chinese surnames unnecessarily map one-to-one into Chinese characters. See Figure 7 for an example. To lessen this issue, we construct an English-Chinese mapping based on the third group of strings, the names written in both Chinese and English. We then break down each English spelling into Chinese surnames in proportion to its relative presence in mixed entries. For example, a "TAN" is equivalent to 0.927 "陈", 0.062 "谭" and 0.011 "谈" as in Figure 7.

Below we provide detailed introductions to the algorithm we execute to identify Huaren and their surnames for each types of strings.

Names Written in Pure Chinese. In the subgroup of name strings that are written in pure Chinese, what is noticeable is that foreigners, typically Japanese and Korean, are also possible to register their names using Chinese characters. Also, some westerners also use transliterated names, which further complicates the analysis. In light of these concerns, we apply the following procedures to identify ehtnic Chinese and their surnames:

1. For each string of name entry: if the length of the string is greater than 4 Chinese characters, we tag the string as non-ethnic Chinese, since usual (ethnic Han) Chi-

nese names hardly contain more than 4 Chinese characters; If the length of the string is exactly 4 Chinese characters, we check whether the first two characters of string matches with double-character Chinese surnames (such as 欧阳, 司马). If the match is successful, we tag the name as ethnic Chinese, otherwise non-ethnic Chinese; If the length of the string is less than 4 Chinese characters, we match the first Chinese character with the Chinese surname dictionary. If the match is successful, we tag name as ethnic Chinese. Otherwise non-ethnic Chinese; This step produces a tag of ethnic Chinese status.

2. For each string of name entry, we match the first one/two/three Chinese characters with the most common 1000 Japanese surnames obtained from Wikipedia. If the following three conditions are satisfied simultaneously, we tag the string as Japanese, otherwise non-Japanese: the match with Japanese surname dictionary is successful, the ID type of the personnel is foreign passport, and the registry address starts with typical locations in Japan; This step produces a tag of Japanese status.

3. For each string of name entry, we match the first Chinese characters with the most common 100 Korean surnames obtained from Wikipedia. If the following three conditions are satisfied simultaneously, we tag the string as Korean, otherwise non-Korean: the match with Korean surname dictionary is successful, the ID type of the personnel is foreign passport, and the registry address starts with typical locations in South Korea; This step produces a tag of Korean status.

4. For each name entry, if every string is tagged ethnic Chinese, non-Japanese, and non-Korean, we deem the person an ethnic Chinese.

5. If a person is deemed an ethnic Chinese, we extract the leftest character of the first string as the surname for the ethnic Chinese, if the length of first string of name entry less than 4; we extract leftest two characters of the first string as the surname for the ethnic Chinese ,if the length of first string of name entry is 4.

Names Written in Pure English. In the subgroup of name strings that are written in pure English, what is noticeable is that surnames can be placed in either the leftest string or rightest string. And some irregular entries that fail to insert blank space between surnames and given names disable us to match name strings directly with Chinese surname dictionary. In light of these concerns, we apply the following procedures to identify ethnic Chinese and their surnames:

1. Divide the name entries into two groups: one with multiple strings (with blank space in the name entry), and another with single string (without blank space in the name entry).

2. For the group with multiple strings, we match the leftest and rightest character with the Chinese surname dictionary. If the match is successful for either the left character or the right character, we tag the name as ehtnic Chinese. If only one of them is matched, the successfully-matched surname is chosen to be the surname of the ethnic Chinese. If both characters are successfully matched, we keep the leftest character as surname by default (in our database, surnames are more likely to be identified at the leftest position). Otherwise, a name entry is deemed non-ethnic Chinese.

3. For the group with single string, we manually determine ethnic Chinese status and surnames.

4. We break down each English-spelt surnames into Chinese characters in proportion to its relative presence in the mixed entries.

Names Written in Both Chinese and English. The group of names written in both Chinese and English serve as a "bridge" between English spellings and Chinese characters. We construct a English-Chinese mapping based on the group of names written in both Chinese and English, which enables us to break down each English-spelt Chinese surnames into Chinese characters proportionately.

1. We break down the name into two groups: the part of Chinese strings, and the part of English strings.

2. For the part of Chinese strings, we apply the same procedure as for the names written in pure Chinese. This step produces a temporary ethnic Chinese tag and extracts surname of ethnic Chinese written in Chinese character.

3. For the part of English strings, we apply the same procedure as for names written in pure English. This step produces another temporary ethnic Chinese tag and extract leftest and rightest English string.

4. If temporary ethnic Chinese tags from both parts are negative, we tag the person as non-ethnic Chinese. If not, we continue the followings: We match sequentially the leftest and rightest English string with the identified Chinese character from Chinese strings, according to the Chinese surname dictionary. If either match is successful,

we tag this person ethnic Chinese, deem the identified Chinese character as ethnic Chinese's surname, and the matched English string as the legitimate spelling for the Chinese character for this surname. If both matches are successful, we keep the leftest spelling as default.

5. This procedures produces not only the surnames for each ethnic Chinese, but also an English-Chinese mapping that can be used to break down English-spelt surnames into Chinese characters.

Determining Huaren. The previous steps identify whether a person is ethnic Chinese or not. We further separate mainlander and Huaren according to the ID type associated with each personnel. A Huaren is an ethnic Chinese that holds a non-mainland Chinese ID—including passport of a foreign country, travel permit for Hongkong, Macau, Taiwan (HMT) residents.
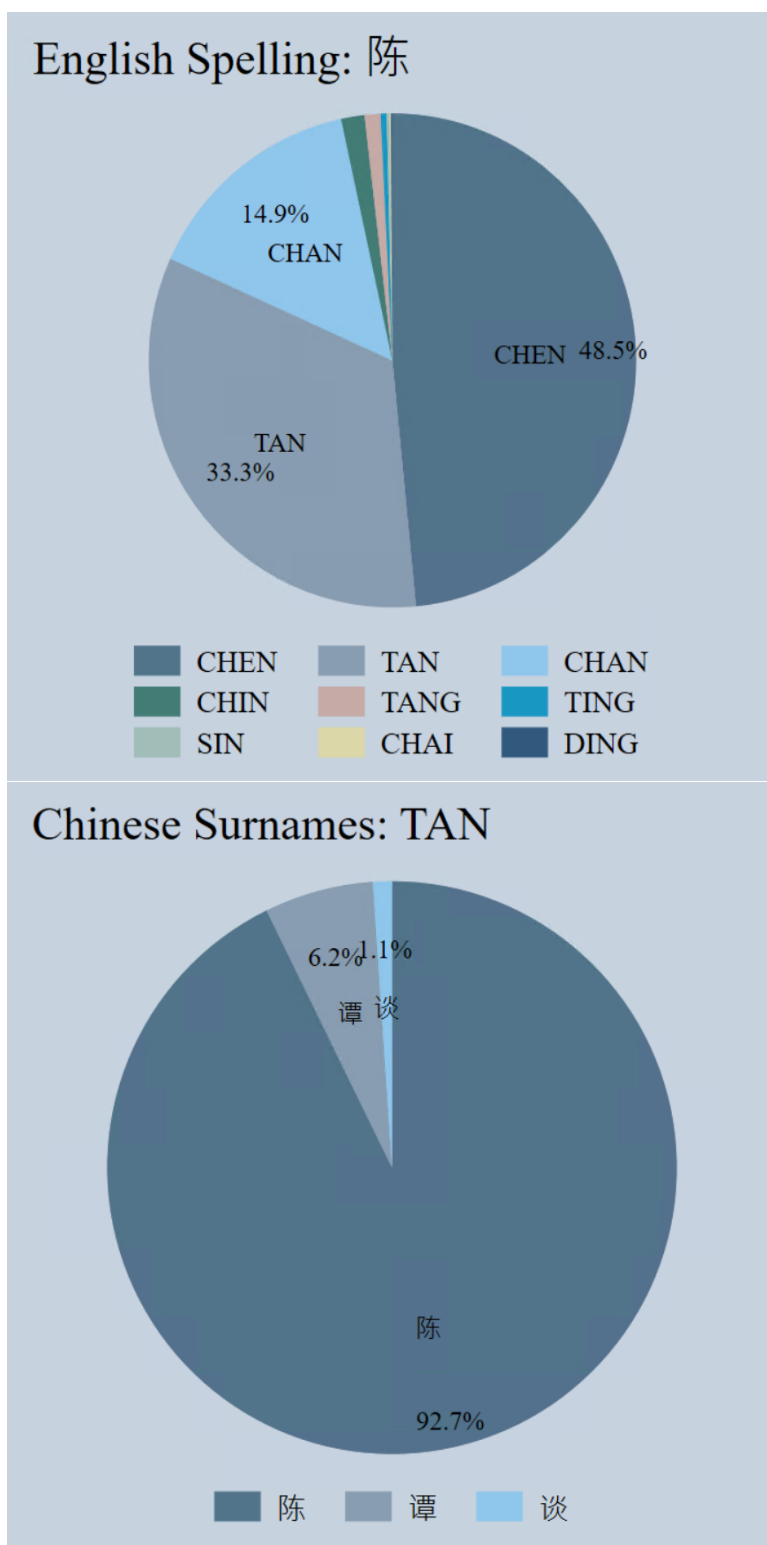
Figure 7: An Example of Spelling-Character Mapping between English and Chinese

B   Legal Representative and Executive Position

Table 9: Legal Representative v.s. Top Executive Position in Foreign Firms

	Top Executive	Not Top Executive	Total
Legal Representative	1334367	94956	1429323
Not Legal Representative	458908	1065118	1524026
Total	1793275	1160074	2953349

Note: The samples are all personnel working in foreign firms ever existed from 1985 to 2014. Conditional on being a legal representative, a person has 93.35%(=1334367/1429323) chance of holding a top executive position within a foreign firm. Conversely, the chance is reduced to 74.41% =1334367/1793275) for one's being a legal representative given that he or she holds a top executive position.

Table 10: Personnel Structures of Foreign Firms

	Percentage
Has a legal representative	96.84%
Has a chairman on the board	23.90%
Has a CEO	20.57%
Has more than one legal representative	1.35%
Has more than one chairman	4.78%
Has more than one CEO	1.24%

Note: The samples are all personnel working in foreign firms ever existed from 1985 to 2014.

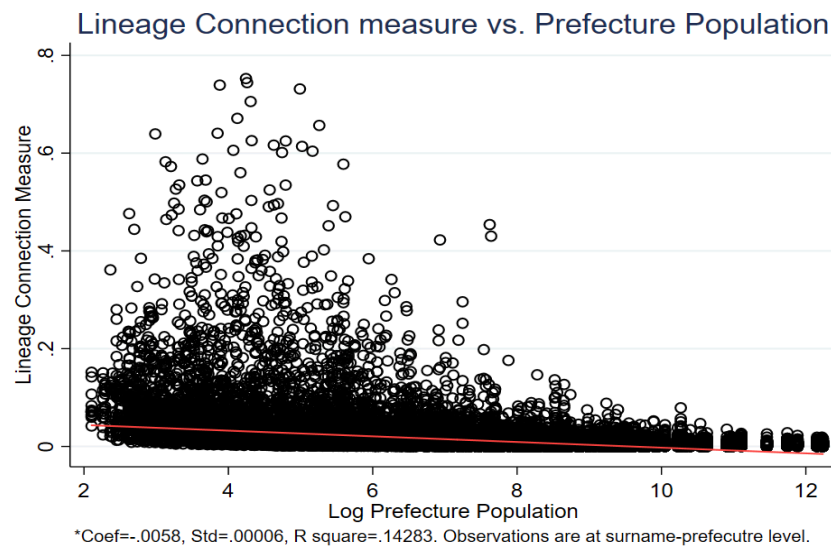# C   Lineage Connection Measure vs. Population Size

Figure 8: Concentration of Surnames

D Heterogeneous Strengths of Lineage Connections

While we treat the strengths of lineage connection homogeneous across surnames in our baseline regressions in Table 3, this section considers the potential heterogeneous strength of ties across surnames. For example, surnames in locations that are more exposed to historical clan culture can exhibit stronger preference for tribalism. Specifically, We measure the strength of lineage connection of a specific surname with the EG index of this surname following Ellison and Glaeser (1997). The rationale is that the EG index of a surname measures the extent to which a person of this surname benefits from co-locating with his or her kin surname.

Table 11: Heterogeneous Strengths of Lineage Connection Across Surnames

	Number of Entrants	Number of Survival-adjusted Entrants	Number of Survived Firms in 2014	Survival Ratio Conditional on Entry in 2014
	(1)	(2)	(3)	(4)
Openness × Lineage Connection	1.548***	1.322***	0.248***	-0.002
	(0.328)	(0.291)	(0.094)	(0.816)
Openness × Lineage Connection × EG Index of Surname	47.592	42.237	3.819	-2.656
	(37.101)	(32.752)	(3.910)	(14.423)
Adj.R^2	0.570	0.555	0.399	0.364
N	1,013,245	1,013,245	1,013,245	9,879
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FEs	Y	Y	Y	Y
Surname-Cohort FEs	Y	Y	Y	Y
Prefecture-Cohort FES	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level and shown in parenthesis. Survival-adjusted Entrants are subset of entrants that survive for more than 4 years (included). EG index of surname is constructed based on population census 2005 following Ellison and Glaeser (1997).

E   Comparison with Other Data Sources

In this subsection, we illustrate the strength of our comprehensive foreign firm dataset and cross-validate our data with official data sources.
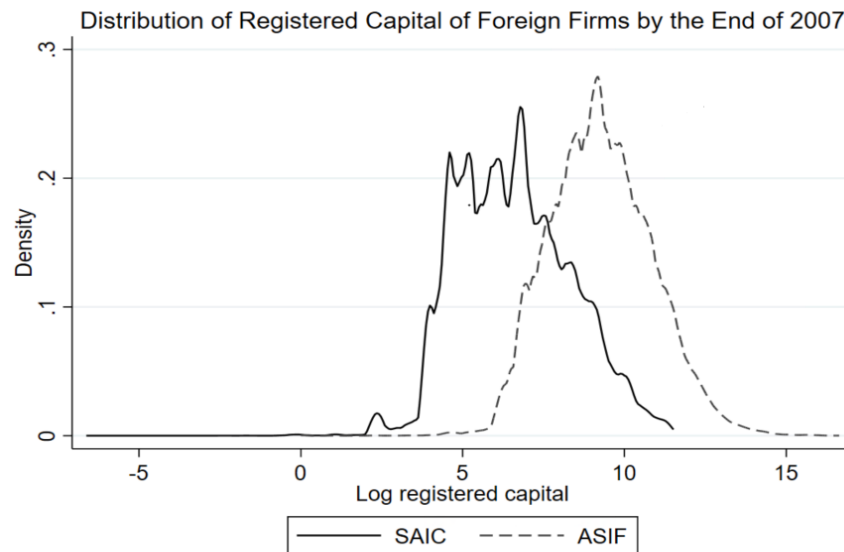

Figure 9: SAIC v.s. ASIF: Firm Size

Foreign firms covered by ASIF vs. by SAIC. Another frequently-used firm-level dataset is Annual Survey of Industrial Firms (ASIF). Our show with empirical evidence that our SAIC database offers additional strengths to study foreign firms in China, compared with ASIF: First, the SAIC database is representative of foreign firm of all size, while ASIF only covers large firms with sales volume above 5 million CNY. Figure 9 plots the distribution of registered capital of foreign firms by the end of 2007, constructed by SAIC and ASIF respectively. It is clear that SAIC has better representation for smaller firms. Second, Figure 10 also depicts that while 98% of firms included in ASIF are manufacturing firms, our full-sample SAIC database suggests that manufacturing firms only account for 41.3% of the population of foreign firms. Third, the two above biases of ASIF in sample coverage result in the large discrepancy between the two databases in terms of the number of entrants over time, as shown in Figure 11.

Huaren foreign firms vs. HMT firms.. Due to data limitation, prior literatures often rely on HMT firms to proxy Huaren foreign firms (Lin, Liu and Zhang, 2009; Huang, Jin and Qian, 2013). There are at least two biases for such approach. First, a lot of Europe-based and America-based multinationals invest in mainland China through Hong Kong
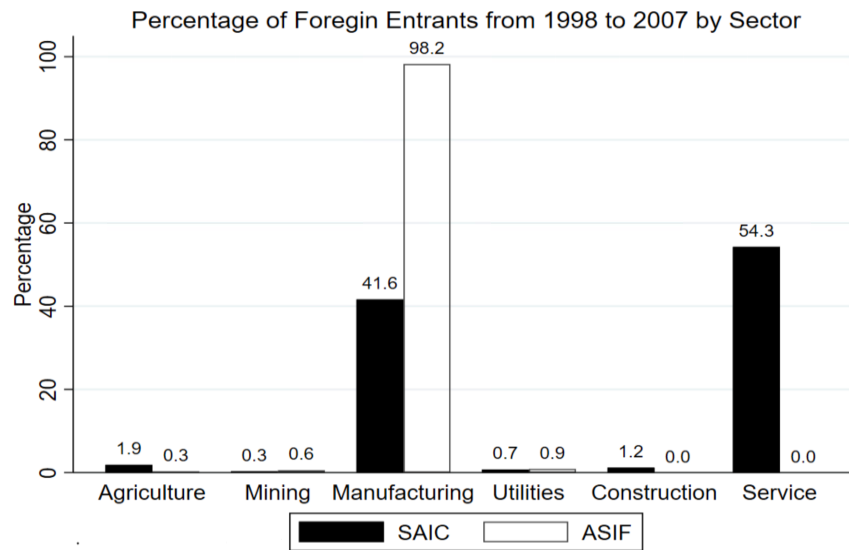
Percentage of Foregin Entrants from 1998 to 2007 by Sector

Figure 10: SAIC v.s. ASIF: Sector

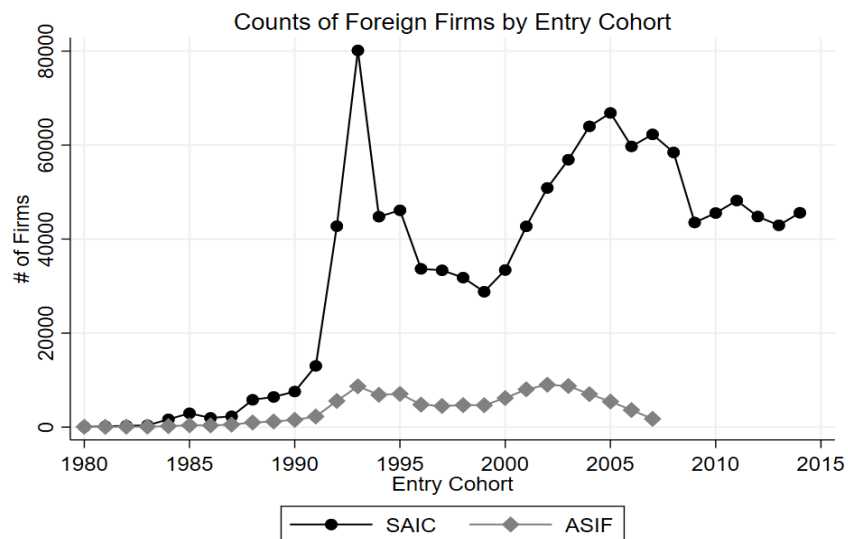Counts of Foreign Firms by Entry Cohort

Figure 11: SAIC v.s. ASIF: Counts of Entrant

as a conduit; Second, residents in HMT only make up a fraction of total overseas Chinese. As can be seen from Figure 12, our estimated number of Huaren foreign firm entries are always greater than HMT firms over time in terms of both JV and non-JV, and the difference seems to widen during the process of China's gradual accession to WTO from 1995 to 2007.
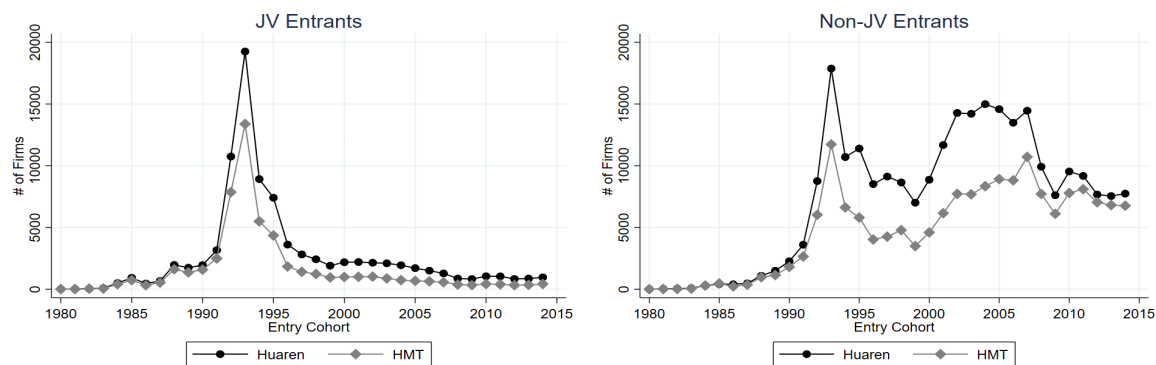
Figure 12: Huaren foreign firms vs. HMT firms

Foreign firms vs. FDI. The notion of foreign firm is related to, but not perfectly comparable to foreign direct investment (FDI). To be exact, FDI is a monetary concept. Besides, FDI includes both the initial and follow-up investment from a foreign-owned entity. We do, however, cross-validate our estimates of foreign firms both in counts and volumes with official FDI data supplied by the Ministry of Commerce of China. Figure 13 suggests that the contracted FDI, measured in number of cases or USD, is highly correlated with our estimates of foreign firm entries, and the realized FDI measured in USD is highly correlated with our estimates of long term foreign firm survival capital by 2014.
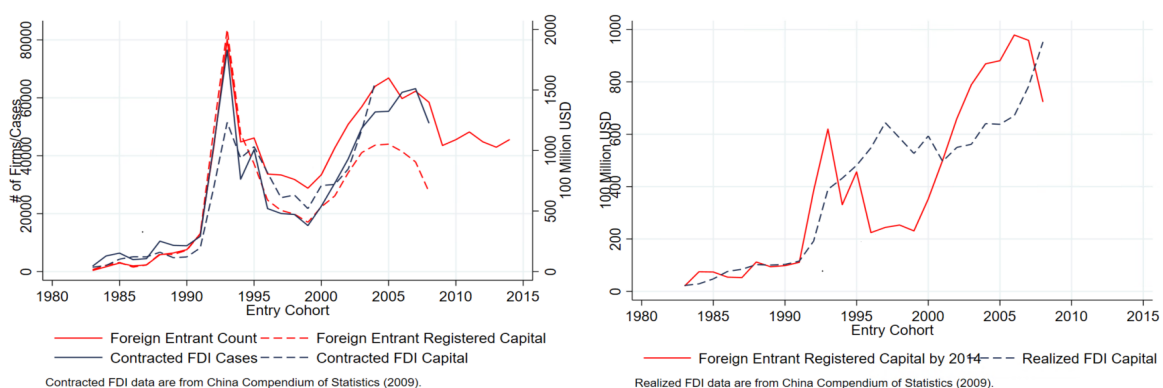

Figure 13: Foreign firms vs. official statistics of FDI inflows

Huaren foreign firms in Fujian Province during the 1980s. Fujian province is a typical emigration-intensive province. Official data on the aggregate number of Huaren foreign

firms from 1979 to 1988 as documented by the Chronicle for overseas Chinese of Fujian Province (1989) provides an additional check on our estimates. Figure 14 compares the official number and our estimated number for the wholly-owned and JVs of Huaren foreign firms. While we are aware of the gap between the estimates of two sources, we are reassuring that both estimates share very similar time trend.
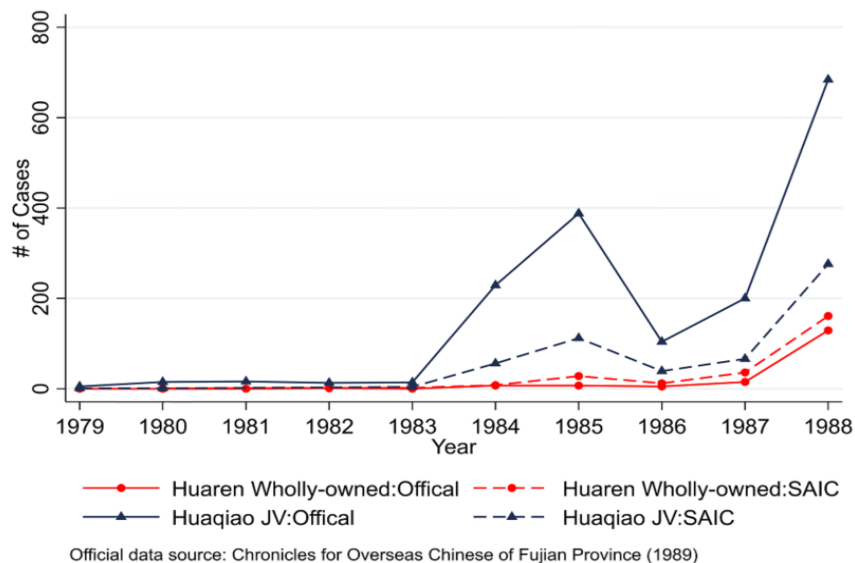

Figure 14: Comparison with Official Aggregate Data in Fujian Province

Lineage Connection measures based on 2005 Census vs. based on SAIC. We construct an alternative lineage connection measure from the list of personnel that worked in a firm established before 1992[14] using the SAIC database. The distribution for top 20 surnames constructed from two data sources are highly similar, and overall there is significant correlation between the two measures (see Figure 16).

---

[14]Xu and Zhang (2009) documents that there was almost no internal migration before the early 1990s in China. Private firms are rather scarce. Most employees and managers worked in township-village enterprises (TVEs) and have lived in the same community for generations.
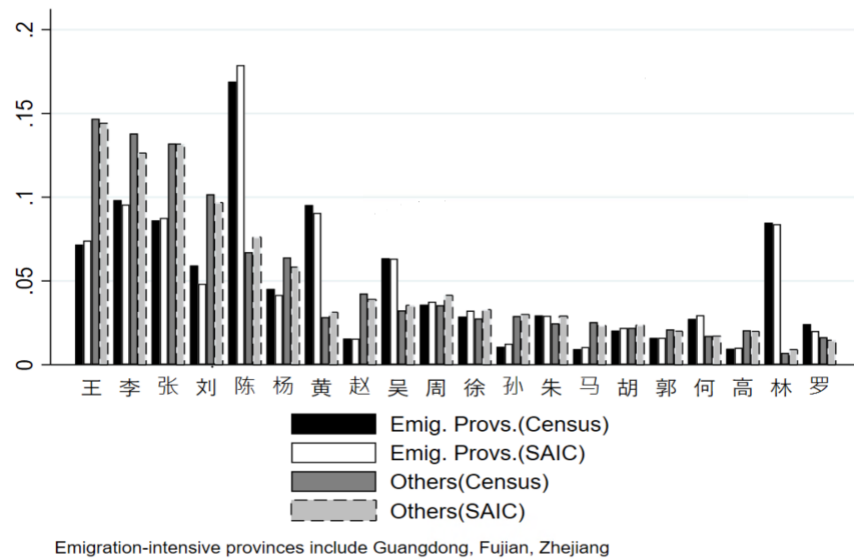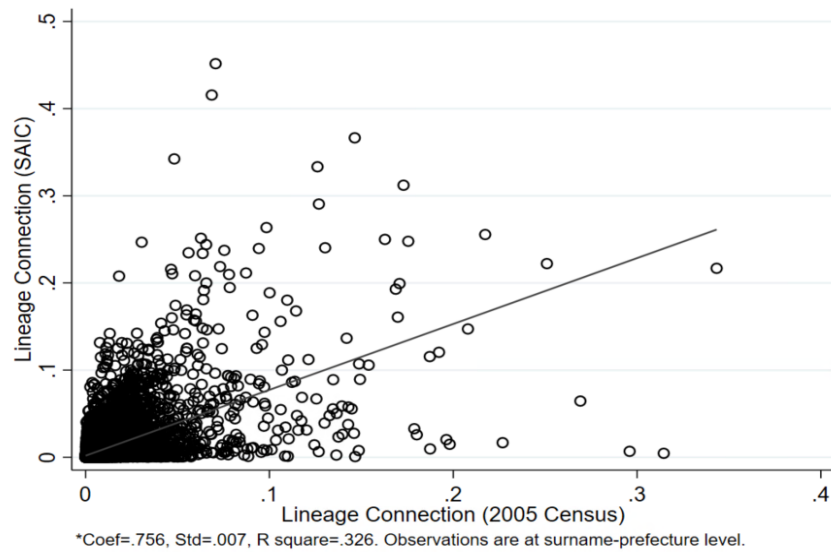
Figure 15: Surname Distribution: Census vs. SAIC

Figure 16: Lineage Connection Measures: 2005 Census vs. SAIC

F   Registered Capital vs. Other Economic Outcomes

Table 12: Registered Capital vs. Other Economic Outcomes

	Log Registered Capital	
	(1)	(2)
Log Employment	0.025***	0.018***
	(0.003)	(0.004)
Log Assets	0.976***	0.416***
	(0.003)	(0.004)
Log Sales	-0.124***	-0.005
	(0.003)	(0.003)
R^2	0.710	0.946
N	15065	124964
Year Fixed Effects	Y	Y
Industry Fixed Effects	Y	N
Firm Fixed Effects	N	Y

Note: Samples are foreign firms according to ownership code from Annual Survey of Industrial Firms (ASIF) ,1998-2007. Industry is at 2-digit level. ***,**,* denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at firm level and shown in parenthesis.