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

Fanghao Chen

Peking University

Ruichi Xiong

University of Toronto

Xiaobo Zhang

Peking University & IFPRI

October 25, 2021

Abstract

This paper studies the importance of international lineage network in facilitating the entry of foreign firms in developing countries where immature market deters foreign investments. We build a unique dataset on the universe of foreign firms in China that are controlled by Huaren (overseas Chinese) whose surnames are used to infer their lineage group. To identify the causal effects of lineage connection, we implement a Triple Difference-in-Difference (DID) design exploiting variations from both the staggered timings of opening shocks during the Reform Era (1981-1996) and differential surname distributions across prefectures, which allows us to control for unobserved factors at various levels. We find that, following the prefecture-level opening shocks, lineage connection facilitates the entry of Huaren foreign firms, through reducing information frictions and strengthening contract enforcement. Such effects are not likely to be driven by selective entry of more able Huaren entrepreneurs, and are robust to various robustness checks. We also find that the lineage-driven FDI during early development stages functions as industrial seeds and led to persistently more foreign investments in the long run.

1 Introduction

Despite the broad positive spillovers associated with foreign direct investments (FDIs) (Harrison and Rodríguez-Clare, 2010), developing countries often face difficulties in competing for foreign capital from global markets. According to the statistics from OECD, usually more than 70% of annual world FDI flow to OCED countries rather than the vast developing world [1]. For example, the United States has been the most popular FDI destination for the last several decades. The immature market environment in developing countries, such as information frictions and incomplete institution building, is often thought to account for these differences. These innate weaknesses prohibit a developing country's from utilizing constructive forces from global markets, let alone kicking start the virtuous cycle and pulling itself out of the poverty trap.

The case of China is, however, exceptional in this regard. 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, even before its formal accession to WTO in 2002. This fact challenges the conventional wisdom. Many anecdotes attribute this achievement to overseas Chinese or Huaren[2], who though as emigrants are believed to have strong lineage connection with their ancestral hometowns[3]. They pioneered in an immature Chinese market where policies and institutions were uncertain and insecure, used informal networks to overcome entry barriers, as well as planted the seed investment for later industrial growth of China (Vogel, 1990). Yet no study so far, to the best of our knowledge, has ever systematically documented the importance of Huaren investments,

---

[1]https://data.oecd.org/fdi/fdi-flows.htm

[2]Huaren is the Mandarin pronunciation for the Chinese word "华人" referring to overseas Chinese. Throughout this paper, Huaren is defined as ethnic Chinese that hold citizenship of foreign countries or Hongkong, Macau, and Taiwan. Huaren and overseas Chinese are interchangeably used thereafter.

[3]In China, lineage refers to the group of descendants of one common patrilineal ancestor across 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 lineage group, as an informal institution that is parallel to the formal institution such as state (Greif and Tabellini, 2017). Traditional activities including compiling genealogy books and worshiping ancestors at lineage temple during family reunions last even to date (Clark, 2015). Many accounts suggest that the same practices are still honored by overseas Chinese even if they emigrated and settled down in a foreign country (Kuhn, 2008).

not to mention empirically examined the causal relationship between Huaren's lineage connection and foreign investments in China.

We contribute to a better understanding of the role played by Huaren in facilitating foreign investments to China, during a period before China's formal accession to WTO, usually referred to as the Reform Era (1981-1996). We choose the Reform Era as our research period because the market environment was under-developed by then, and this period witnessed Huaren investment as the dominant form of all foreign investments. Focusing on the Reform Era enables us to fully uncover the effects of lineage connection on the entry of foreign firms, and to explore the underlying mechanisms. Furthermore, while there might be concerns that foreign investments can be ephemeral, and reliance on use of informal institution could potentially crowd out local development, we provide evidence that these lineage-driven Huaren investments are persistent and conducive to the local development in the long run.

The analysis proceeds in several steps. In the first step, we introduce the new dataset 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 the foreign firms according to the official ownership code assigned by the SAIC. We next determine from the surname and the ID of the foreign firm's legal representative whether a foreign firm is controlled by Huaren. The above procedure leads to our final sample of Huaren foreign firms. Then we can use the observed surnames of the Huaren legal representatives and link them to the geographic distributions of surnames constructed from population census to measure each Huaren foreign firm's lineage connection with different prefectures of China.

Second, 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

4

entry of connected Huaren foreign firms, creating a spurious correlation. For example, in 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* (SEZ thereafter) (Wang, 2013). Moreover, entrepreneurial culture and tacit knowledge for conducting international business can be passed from generation to generation within certain lineage groups. To overcome these concerns, we adopt a Triple DID strategy that exploits variations from both China's staggered opening to foreign capital and predetermined surname distributions across prefectures. With this framework, we are able to control for fixed effects at surname-prefecture, surname-year, and prefecture-year level 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 local opening shocks on the entry of Huaren foreign firms at surname-prefecture-cohort level, measured by the following variables: the number of entrants, the survival-adjusted number of entrants and the number of survived firms in 2014. We perform a formal event study to show that our empirical design satisfies the common trend assumption. Besides, our findings remain robust to various checks such as examining the extensive margin of firm entry, using alternative lineage connection measures, clustering standard errors at a different level, excluding FDI-intensive prefectures or emigration-intensive provinces. We also conduct a permutation by reshuffling the lineage connection measures within and across prefectures, and a placebo test using China's WTO accession (the formalization of institution) as the opening shock. As expected, we find no effects of lineage connection in both cases.

However, if Huaren entrepreneurs with higher latent ability can better take advantage of their inherent lineage connections, our results might suffer from positive selection bias. To mitigate this concern, we check if lineage connection contributes to the survival ratio and average survived registered capital in 2014 of Huaren entrants conditional on their entry. These two outcomes are believed to indicate long-term firm growth and are af-

5

fected by the latent entrepreneurial ability of Huaren entrepreneurs. We do not find any significant effect of lineage connection on these two outcomes, therefore we conclude that the selection on entrepreneurial ability are unlikely to bias our results. The results also suggest that lineage connections are more salient on reducing entry barriers, rather than firm's later growth.

The third step sheds light on the mechanisms through which lineage connection helps overcome the entry barriers for Huaren foreign firms. We find that the effects of lineage connection are stronger in prefectures with higher social fragmentation or better telecommunication infrastructure. We also find that Huaren investors with higher lineage connection are more likely to start a business as an imitator of existing industries, a headquarter firm (relative to a branch firm) and in labor-intensive and supplier-intensive industries. These findings reveal the role of lineage connection as a kind of social network in facilitating information access and strengthening contract enforcement (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 present-day economic development. 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 long-term effects of these lineage-driven FDI on the presence of foreign firms by not only Huaren but also non-Huaren in terms of both counts and registered capital in 2014. This suggests that the benefits of introducing FDI as industrial seeds seem to outweigh its costs during the take-off stage of China and might account for policy makers' enthusiasm toward promoting inward FDI in world-wide developing countries.

Existing Literature. Our paper adds to a large literature on the role of social affinities in facilitating trade and investment flows across countries and regions pioneered by Rauch (2001). Following works have considered ethnic ties (Rauch and Trindade, 2002), bilateral 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). We contribute to this literature by quantifying and causally estimating a new type of social affinity that functions between a location and a surname in China—the lineage connection which is sustained and carved into the second-nature of Chinese (Greif, 1993). 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. Ma (2018) presents evidence on the effect of cultural similarity on attracting FDI, employing the geographical borders of Chinese linguistic dialect zones. Our paper, for the first time in this literature, studies surname-based Chinese lineage connection. Furthermore, we incorporate both short- and long-run effects and uncover the underlying mechanism, using innovative research designs.

This paper also complements the literature on the long term implications of international migration. Overseas Chinese networks are believed to be vital to the development of China while empirical evidences are scant if not absent. 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 using a surveyed sample of firms in a particular industry while we draw upon the universe of Huaren foreign firms in China. We show that an unexpected benefit of emigration is to facilitate foreign investments through migrant networks that provide seed investments for long-tem development of the home country. In addition, our usage of surnames to uncover 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 and 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) and institutional quality (Du, Lu and Tao, 2008). However, 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 paint a systematic picture of Huaren foreign firms across Chinese prefectures and document that a substantial portion of foreign firms in China founded since the 1980s are run by overseas Chinese. Moreover, we identify causally the effect of lineage connection on Huaren foreign firms and uncover the underlying mechanisms. Last, we estimate causally the seeding effects of lineage-driven Huaren FDI on long-term presence of foreign firms in China. These findings establish the importance of Huaren in China's development process and have profound policy implications for the other developing countries.

Finally, our research advances an emerging literature using firm registration data of China. Among others, Dai et al. (2019) and Bai et al. (2020) explore the role of community networks and state power underlying the rapid growth of China's private sector. While most existing studies pay attention to the private domestic firms, we focus on foreign firms, a nowadays relatively small but historically large part of the Chinese economy that this line of literature has not 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 elaborates our identification strategy and presents the empirical findings. Section 5 then explores the underlying mechanisms which could account for our findings. Section 6 further estimates the long-run effects of early lineage-driven Huaren FDI on future foreign investments. Section 7 concludes.

2 Historical Background

Ever since its dramatic turn toward market economy in 1978, China has been swiftly transformed from a completely closed state to a hub for foreign firms within thirty years. After 1978, when Deng Xiaoping initiated the great Reform and Opening-up, China started to expose stepwise its territories to international trade and foreign investments (Branstetter and Lardy, 2006). Table 1 summarizes the staggered opening process during the Reform Era. In 1979, the *Law on Joint Ventures* was passed. It provided for the first time in thirty years a legal framework under which foreign firms were allowed to operate in mainland China. In the following 1980, four SEZs including Shenzhen, Zhuhai, Shantou and 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 *Open Coastal City* for attracting FDI. Then Yingkou, Weihai and Hainan province were also added to the list of places open to foreign capital sequentially in 1985, 1987 and 1988. Although the economic and political crisis in 1989 temporarily added considerable uncertainties in China's transition, Deng's visit to southern China in 1992 restored people's confidence in China's resolution to embrace market economy and globalization. At the same time, the visit also marked the start of China's comprehensive reform and national opening to foreign capital.

The impacts of the stepwise rolling out of opening policies in Table 1 is evidently shown in Figure 1. It plots the entry and survival of foreign companies in mainland China by entry cohort using firm registration data from SAIC database. We can see firstly that both quantity and quality of foreign firm entries measured by long-term survival rises steadily

Table 1: China's Opening Process during the Reform Era

Year	Open Policy	Open Regions
1980	Special Economic Zone	Shenzhen, Zhuhai, Shantou, Xiamen
1984	Open Coastal City	Dalian, Qinhuangdao, Tianjin, Yantai, Qingdao, Lianyungang, Nantong, Shanghai, Ningbo, Wenzhou, Fuzhou, Guangzhou, Zhanjiang, Beihai
1985	Open Coastal City	Yingkou
1987	Open Coastal City	Weihai
1988	Special Economic Zone	Hainan Province
1990	Special Economic Zone	Shanghai Pudong District
1992	Deng Xiaoping's South Tour	All other regions

Source: https://en.wikipedia.org/wiki/Chinese_economic_reform

from 1980 to 1991 along with the opening of SEZ and *Open Coastal City*. And then right after Deng Xiaoping's south tour in 1992, foreign investors intimidated previously by the uncertainties on China's future transition to open market economy immediately entered. This is reflected in the dramatic spike of foreign firm entries in 1992 and 1993 in Figure 1. China's accession to WTO has been documented to play an particularly important role in making the country hub for foreign investments by reducing information friction and building pro-market institutions (Branstetter and Lardy, 2006). The takeoff of foreign firm entries during 2000-2005 in Figure 1 echoes this major historical event in China's great economic reform.

We noticed that the entries of foreign firms were already very active even before China's joining WTO. In 1999, right before the accession, China actually had already become one of the most popular destinations for FDI, only second to the United States (Huang, 2003). This achievement is unusual as it challenges the conventional wisdom that the commonly observed barriers such as information frictions and poor institutional environment in developing countries would deter foreign investments. Moreover, the open policies sum-

10
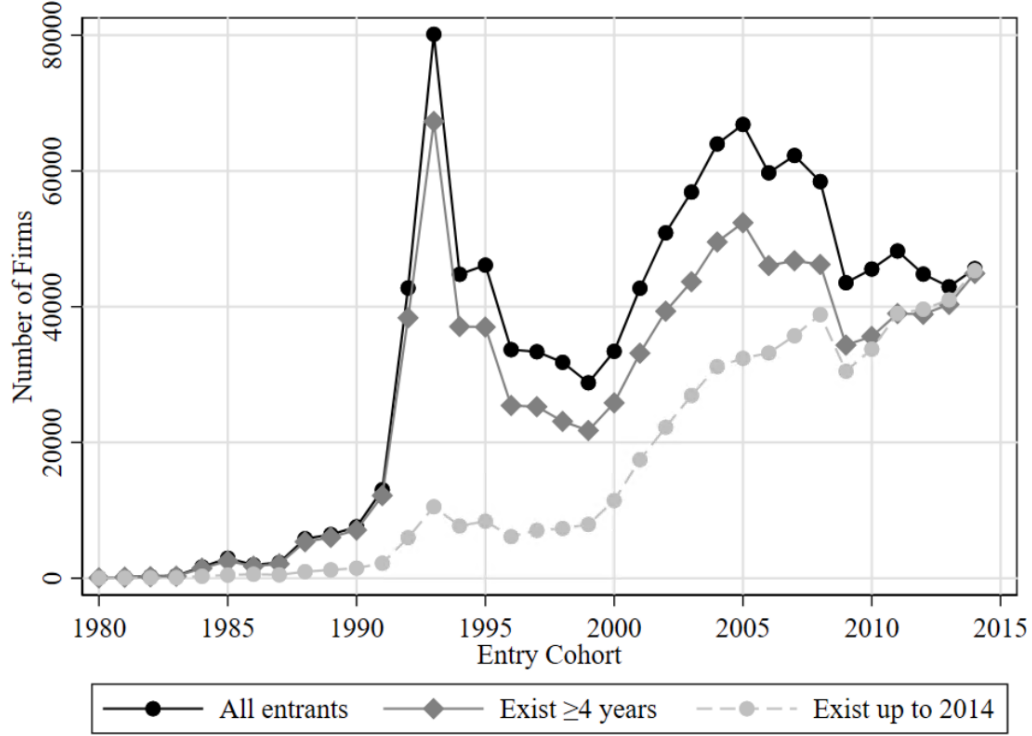
Figure 1: Entry and Survival of Foreign Firms by Cohort

marized in Table 1 can not fully explain this fact as those policies are commonly used by developing countries who mostly failed to attract foreign investments. Hence there must be complementary factor to open policies for accounting for the trend shown in Figure 1. Furthermore, the complementary factor's impacts on foreign firm entries should become weaker in recent years when the barriers to enter China has been gradually lowered with the maturing of the market economy.

We peer into the early stage of China's development before joining WTO and suggest the widespread networks with overseas Chinese originating from historical emigration as key contributor in overcoming barriers and facilitating foreign firm entries in China. As descriptive evidence in support of our proposition, Figure 2 decomposes foreign firm entries into Huaren entries and Non-Huaren Entries. During our sampling period 1980-2014, the share of Huaren entries is almost always higher than 60% and declines from above 90% in 1980s to about 60% in 2010s. These two observations point to firstly the
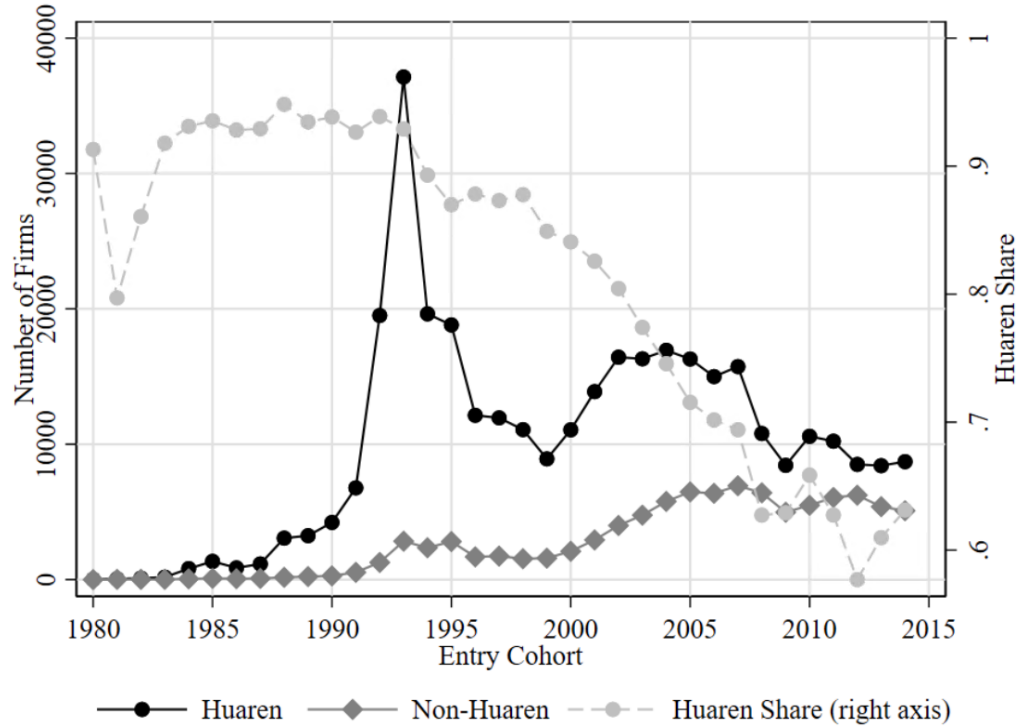
Figure 2: Breakdown of Foreign Firms: Huaren versus Non-Huaren

importance of foreign investments by Huaren through out China's economic development process since 1978 and secondly Huaren investor's lost of comparative advantage in entering China in recent years. The influence of lineage connection embedded in Huaren investor's identity on foreign firm entry therefore is significant during the Reform Era and then becomes weaker in recent year. That's to say lineage connection could be the the complementary factor we seek for to understand why open policies work in China even before the accession to WTO.

China has a extremely long history of emigration which can be dated back to at least 1300s A.D. (Kuhn, 2008). According to the Global Migrant Origin Database, the stock of overseas Chinese has reached 5.8 million in 2007, making China the world's 6th largest country in exporting migrants[4]. Among those overseas Chinese, a lot of them have become successful entrepreneurs in the host countries. According to the estimate by *The*

---

[4]www.sussex.ac.uk/Units/SCMR/drc/about/index.html

Economist in 2019, more than three quarters of the wealth owned by the South East Asian billionaire are controlled by Huaren[5]. And despite 30 years' seal-off since 1949 when China turned closed, the bond between overseas Chinese and their ancestral hometowns has never been truly removed. The overseas Chinese, though physically abroad, are emotionally attached to their ancestral hometowns. Many anecdotal facts demonstrate that in history overseas Chinese remain socially in touch with their lineage group members, through mail communication and sending remittance along with the mails (Tan, 2006). These behaviors, as Greif and Tabellini (2017) illustrated, reflect some persistent and fundamental values that are rooted in the Chinese culture. For centuries, people from the same lineage groups exchange information, smooth contract enforcement, and fight against weather crises by helping each other out.

Therefore, when China opened its door in 1978 and strove to get foreign investments, the widespread networks with overseas Chinese across China became a kind of "natural" resources. Local governments in the opened regions realized the complementarity between the opening policies and the local lineage connection with overseas Chinese. They hence enacted policies and coordinated through specialized organizations to encourage overseas Chinese businessmen to invest back in their ancestral hometowns in pursuit of capital, technological capacity and export channels (Huang, 2003). Having been isolated from the world market for three decades, China was at a huge disadvantage to engage in international trade. However, these transitional policies can set the local lineage network which remains effective even after 30 years' disconnection in motion. With the help of the international ethnic ties, China was able to overcome common barriers deterring foreign investments in under-developed countries and make its share in international trade and FDI inflows rocket up even before its accession to WTO. Furthermore these pioneering foreign entrants attracted to China, because of their lineage connection's impacts, are believed to plant the seed investments for China's later industrial growth and bring about a

---

[5]"Chinese Diaspora Inc: High-Wire Act", *The Economist*, May 30th, 2020

myriad of economic benefits in the long run.

3 Data Description

To empirically estimate the effect of lineage connection on foreign firms entries, we iden-
tify foreign firms controlled by Huaren and then obtain a sample at surname-prefecture-
year level from multiple large-scale micro datasets described below on foreign firm entries
and population. For further identifying the underlying mechanisms and evaluating the
long-term seeding effect of early foreign firm entries, a number of industry and prefecture
characteristics are collected from other datasets or constructed from our micro datasets.

3.1 Data Sources

The main micro dataset we use is firm registration data maintained by SAIC and tracks the
universe of firms ever registered in China up until 2014. For each firm, we observe its entry
date, exit date (left blank if the firm still survives in 2014), latest-reported registered capital
by the end of 2014, 4-digit industry, county-level location, ownership and a complete list of
immediate shareholders and registered personnel including usually board members and
senior executives.

    The second micro dataset we use is the 2005 population Census which reports indi-
vidual surnames of a 0.2% representative sample, from which we recover the geographic
distributions of surnames across prefectures. We then leverage the geographic variation
of surnames to construct our measure of lineage connection for each surname-prefecture
pair and social fragmentation index for each prefecture.

    Outcome variables on economic development at prefecture level are retrieved from CS-
MAR database and various China City Statistical Yearbooks. These variables include GDP
per capita, number of registered patents, average wage, export and import. We get the ca-
pacity of telephone exchange at province-year level 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 while the contract intensity measure we use is calculated based on 2002 Input-Output Table of China following the methodology of Levchenko (2007).

3.2 Identify Huaren Firm

We aim at identifying the role played by surname-based lineage connection in facilitating foreign firm entries. For this goal, our focus is put on foreign firms controlled by Huaren whose surnames are used to measure these firms' lineage connection. In this section, we describe how we proceed in steps to get the universe of Huaren firms.

By our definition, Huaren firms have to be foreign firms firstly. To construct the universe of all foreign firms in China from firm registration database, we need to have a definition of foreign firms. Two approaches coexist in literature for defining foreign firms. One approach uses the administrative ownership codes of firms assigned by government agencies while the other approach uses shareholder information. In this paper we opt for using the administrative ownership code. It is often the case that firm's immediate shareholder structure does not coincide with its structure of ultimate control. For example, investors can exert control over a firm through holding shells (Bai et al., 2020). Then if we use the shareholder information to determine the foreign firm status, we have to decide which shareholders from those investors who have either direct or indirect control over the firm should be used. The relationship between controlling rights over the firm and the shareholder structure of the firm is not the focus of this paper. We therefore wish to avoid this complicated choice which could bias our results by taking the conservative approach of using administrative code readily available in the dataset and independent with our research design. 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 Buy-Outs (MBO) rather than being sold 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 potential change of ownership type would affect our identification of foreign firms. Therefore we conclude it is better to select foreign firms according to the ownership code that the SAIC assigns to each firm. This process gives us a comprehensive database on the universe of foreign firms in China.

With the universe of foreign firms at hand, we identify Huaren firm as foreign firm controlled by Huaren. For achieving this goal, the most natural choice is to make use of the shareholder information and identify Huaren firm as firm with Huaren controlling shareholder. Unluckily, most of the shareholders of foreign firms are unsurprisingly foreign entities registered outside China and we can not trace the shareholders of those foreign entities (Bai et al., 2020). Hence we turn to use the information of legal representative of the firm. According to the corporate law in China, legal representative takes the major legal responsibility of the firm as legal entity. Then legal representative should be very likely to be the one with controlling right over the firm. And existed literature even identify legal representative as entrepreneur (Dai et al., 2019). The key advantage of using SAIC database is that we are able to observe names of each firm's registered personnel and their ID types reported to the SAIC. We develop an algorithm as described in Appendix A to extract the surnames of each foreign firm's registered personnel and identify whether they are ethnic Chinese. If we observe one personnel as ethnic Chinese and the personnel also holds a non-mainland Chinese ID[6], he or she is defined as "Huaren", i.e. overseas Chinese. And Huaren firms are then identified as foreign firms whose legal representatives are Huaren[7]

---

[6]Non-mainland ID can be either passport of a foreign country or travel permit for Hongkong, Macau, Taiwan (HMT) residents

[7]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. Firstly, we show in Appendix B, that 97% of all foreign firms have unique legal representative within the firm while they might have none or multiple directors, CEOs or top-ranked executives. Secondly, we do not observe the surnames of most shareholders of foreign firms as 92% of immediate shareholders hold equity stakes of foreign firms through legal persons rather than natural persons. Even a backing-out through ownership

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 individuals from mainland China. Such firms may be related to what Huang (2003) calls "roundtrip FDI". This type of FDI refers to capital that a domestic firm firstly exports to overseas places (majorly Hongkong) and then imports back into mainland China as a new legal entity. Since the establishments of such foreign firms are most likely for the tax breaks and preferential treatment granted toward foreign firms during the Reform Era, taking this group of firms into account would enormously exaggerate the presence of Huaren firms in China and potentially bias our estimates in empirical analysis[8]. Moreover, our main dataset also excludes foreign firms represented by non-Huaren since we want to keep our sample as comparable as possible[9]. Lastly, since the legal representative is not necessarily the major shareholder of the firm, it is possible that a company classified as Huaren firm is actually not owned by Huaren. However, the choice of appointing Huaren legal representative still reveals the motive of the foreign firm to take advantage of lineage connection. And to be noticed again that ultimate owners cannot be traced as foreign entities registered outside China do not report information to SAIC.

---

chain to identify each foreign firm's ultimate owners who will natural persons will not work because the SAIC does not have information on the owners of foreign legal entities that are registered abroad (Bai et al., 2020). Thirdly, Appendix B reports that the chance a legal representative overlaps with a top-ranked executive within a foreign firm is 93%. This means that a legal representative is highly likely to be the person in charge of the foreign firm.

[8]We cannot, however, rule the possibility that foreigners delegate the firm control rights to individuals in mainland China. And this exactly reflects the influence of lineage connection. We show in appendix that our main estimates are robust to including those foreign firms with people from mainland China as legal representatives.

[9]Including non-Huaren foreign firms would mechanically increase the size of the control group with no lineage connection. Yet, most of our findings are robust to including non-Chinese foreign firms. See Section 4.4.

3.3 Measure Lineage Connection.

We already identified Chinese surnames of the legal representatives in the process of identifying Huaren firms. Then we can measure lineage connections of these Huaren firms based on the surnames of their legal representatives with the 2005 population Census. Specifically, we define the lineage connection m_{sp} between a surname s and prefecture p[10] as following:

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

where E_{sp} denotes the size of population with surname s in prefecture p. To be noticed that 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[11]. So we explicitly rule out the impact of surname size in determining surname's lineage connections with prefectures across all our specifications.

The novelty of our approach lies in the usage of surnames instead of migration patterns to proxy lineage connections. And we contribute to literature in providing this new surname-based lineage connection measure for two purposes. Firstly, while there are relatively rich data on immigration into developed countries such as US (Burchardi, Chaney and Hassan, 2019; Sequeira, Nunn and Qian, 2020), so far the data on emigration, especially at sub-national level from developing countries like China, are extremely rare to our knowledge. Our surname-based measure makes best use of the fact that lineage institution

---

[10]There are three reasons for why we choose prefecture (instead of county) as the geographic unit for our empirical analysis. Firstly, measurement noises because of the presence of rare surnames are large at small counties. Secondly, boundaries of counties may change over time while that of prefectures are fairly stable. Thirdly, we have much more variation in terms of lineage connection across prefectures than across provinces.

[11]If we instead use E_{sp} as measure of lineage connection, we expect spurious correlation between lineage connection and Huaren firm entries even without the functioning of lineage connection as populous surnames like "陈" ("Chen", "Chan", "Tan") can be dominant both domestically and abroad so that Huaren foreign firms mechanically have legal representatives with such surnames. Moreover, Appendix C indicates that our lineage connection is actually mildly negatively correlated with prefecture population at surname-prefecture level.

in China are operating within surname (Clark, 2015). Secondly, the surname-based measure we construct is predetermined and time-invariant while emigration flows if available are endogenously determined in equilibrium over time. Adopting our measure therefore significantly mitigates potential endogeneity concerns.
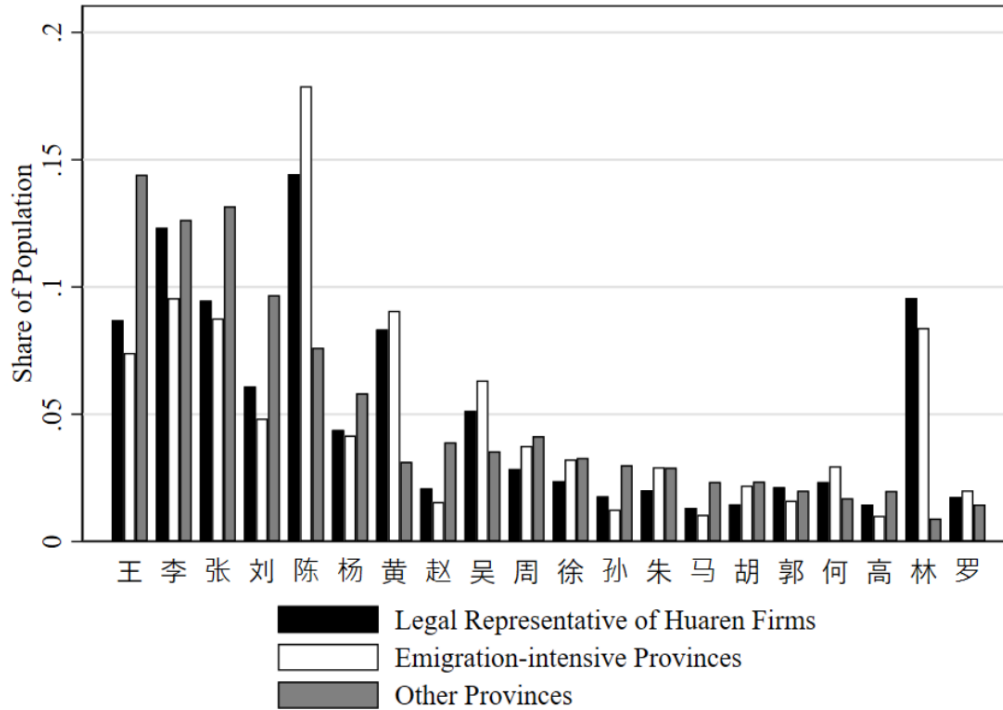
Figure 3: Surname Distributions in Different Population

Figure 3 visualizes the geographic distributions of the 20 most populous surnames in China. We draw the distributions separately for emigration-intensive provinces (including Guangdong, Fujian and Zhejiang), other provinces and all legal representatives of Huaren firms registered from 1980 to 2014. Informative disparities and similarities emerge from comparison among the three distributions. Overall, the surname distribution over legal representatives of Huaren firms are closer to that over the population of emigration-intensive provinces than the population of other provinces. For example, "陈" ("Chen", "Chan", "Tan")[12] is the most common surname both among Huaren legal repre-

---

[12] Due to Chinese-English translation, there could be more than one English spelling for the same Chinese

19

sentatives and in emigration-intensive provinces but ranks only fifth in the total population and has a modest share in the population of other provinces. This simple comparison suggests that the variations of surname distribution across regions indeed reflect historical emigration and can be used as predetermined variation for identifying the effects of lineage connection.

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[13]. While this assumption may seem arguable at first glance, we find strong support for it from both the economics and history literature. Clark (2015) documents that geographic distribution of certain elite surnames is rather stable throughout history even after warfare and revolutions. Bai (2020) compares the distribution of surnames across prefectures using 2005 population Census with using Harvard's China Biographical Database (CBDB) and confirms that the two distributions are highly correlated.

3.4   Summary Statistics

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. 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 construct three measures of Huaren firm entries at each surname-prefecture-year level: number of new entrants, survival-adjusted number of entrants (defined as those lasting for more than 4 years) and the long-term survived number of entrants at the end of 2014.

---

surname. Figure 6 illustrates the complexity of spelling-character mapping between the two languages with "陈" which is the most common Huaren surname in our data as an example. In our study, we aggregate Huaren surnames to the Chinese character level with some probabilistic assignment when in rare cases the Chinese character of the surname cannot be observed. See Appendix A for details.

[13]Internal migration might threat the stability of surname distributions. From 1949 to 1978, however, there is almost no internal migration due to stringent Hukou restriction. Even after China gradually loosened its grip on internal migration since 1984, it only formalized free mobility of internal migrants until early 2000s (Tombe and Zhu, 2019).

We exclude four provinces including Xinjiang, Tibet, Ningxia and Inner Mongolia where majority of population are ethnic minority since surname-based lineage mostly operate among ethnic Han Chinese. We further exclude four mega cities in China including Beijing, Shanghai, Guangzhou and Shenzhen because their outlying economic performance massively attract internal migrants and could potential confound our results by driving up simultaneously 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 the variation of the presence of Huaren firms is mostly accounted by the extensive margin. Only 5.2% of surname-prefecture-year triples have Huaren firm entry while the magnitude is even smaller (0.9%) for Huaren firm survival rate in 2014. Panel B suggests our lineage connection measure varies widely with a mean of 0.005 and a standard deviation of 0.011. In Panel E, we can also find that Huaren entrants tend to be headquarter firms rather than branch firms, labor-intensive rather than capital-intensive and followers rather than leaders in new industries[14]. These characteristics are largely in line with the anecdotal descriptions on Huaren firms (Huang, 2003). Besides, only 57.8% of Huaren firms during this period originated from HMT, suggesting that HMT entrants cannot be used as a precise proxy for Huaren entrants as done by literature[15].

---

[14]A pioneer firm is defined as a first-ever foreign 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 supplier-intensive if the contract intensity measure, proposed by Levchenko (2007) and calculated using Chinese Input-Output Table in 2002, is above the median level across all 2-digit industries.

[15]See Appendix E for detailed discussion on the strengths of our dataset relative to those used by prior literature and cross-validation of our dataset with other data sources.

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
Survival-adjusted Number of 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
Average Registered capital in 2014 Conditional On Entry	27,846	806.588	3967.006	0	98643
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 Exchange 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
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: Huaren Firm					
Pioneer Firm	157,875	0.155	0.362	0	1
Branch Firm	157,875	0.131	0.337	0	1
HMT Firm	157,875	0.578	0.494	0	1
Labor Intensive	92,951	0.685	0.465	0	1
Supplier-Intensive	138,925	0.684	0.465	0	1

Note: In panel B, the measures of lineage connection constructed SAIC is calculated based on the surnames of registered 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 in 2014 of non-Huaren foreign firm using SAIC database.

4 Lineage Connection and Foreign Firm Entries

4.1 Identification Strategy

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

$$Y_{spc} = \eta_{sp} + \eta_{sc} + \eta_{pc} + \beta \times Open_{pc} \times m_{sp} + \lambda \times S_{spc} + \epsilon_{spc} \qquad (2)$$

where Y_{spc} represents the outcome variables. We are primarily interested in how lineage connection affects the number of Huaren foreign entrants. Nevertheless, entrants vary by their quality and the welfare implication of firm entry might be undermined if entrants with low quality are selected. Therefore, we supplement number of entrants with another two measures to address this concern: the survival-adjusted number of entrants (defined as number of entrants that lasted for more than 4 years) and the number of survived firms by 2014. The usage of the latter measure makes sure that our results do not come from cherry-picking the particular 4 year threshold.

Under all specifications, we control for surname-prefecture fixed effects η_{sp}, surname-cohort fixed effects η_{sc} and prefecture-cohort fixed effects η_{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. To be noticed that, when we use number of survived firms in 2014 as outcome, these fixed effects help us control for any consequences brought by the fact that firms enter in different year have different probability to survive in 2014. We also include S_{spc} indicating the number of incumbent firms one year prior to the entry cohort in order to capture potential agglomeration spillovers from incumbent firms.

The main independent variable is the interaction term $Open_{pc} \times m_{sp}$. $Open_{pc}$ is a time-variant measure of prefecture's opening status to foreign capital that varies across regions.

$Open_{pc}$ take value 1 if the prefecture p started to be open since year c and 0 otherwise[16]. m_{sp} is the measure of lineage connection between surnames and prefectures and is defined in Equation 1. We mainly use the 2005 population census of China 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[17]. 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 estimates the heterogeneous effects of lineage connection across surnames following the staggered opening of Chinese prefectures to foreign capital during the Reform Era. Our empirical design is therefore essentially a staggered Triple DID strategy with differential treatment intensity across surnames. We are in fact, by employing the design, doing the following three-fold comparisons: (1) compare surnames within prefecture before and after the opening of prefecture to get trends; (2) compare the trends across opened and closed prefectures to get the diff-in-diff at prefecture level; (3) compare the double differences further across surnames with different lineage connections to get the Triple DID estimates. Therefore the identification assumption for the causal interpretation of our estimates should be the following: there exists no differential pre-trends in firm entries of surname-prefecture-cohorts with different lineage connections before the opening shock. We argue this identification assumption should be very likely to hold for several reasons. Firstly, our measure of lineage connection is predetermined. Secondly, the timing of the opening of prefectures during the Reform Era comes largely as shocks. Thirdly, our empirical design as Triple DID really eliminates a lot of factors at different levels which could contaminate our estimates.

We further provide empirical tests to corroborate the validity of our identification strategy. And one test is checking the parallel trends. To perform the test, we employ an event-

---

[16]See Table 1 for a full depiction of the stepwise opening of Chinese prefectures during the Reform Era. To be noticed that, the case that one prefecture became open and then reversed back to be closed in following years does not exist.

[17]See Section 4.4.

study framework as following:

$$Y_{spc} = \eta_{sp} + \eta_{sc} + \eta_{pc} + \sum_{\tau=-4}^{4} \beta^{\tau} \times Open_{pc}^{\tau} \times m_{sp} + \lambda \times S_{spc} + \epsilon_{spc} \qquad (3)$$

where $\tau = c - c_p$ referring the time window relative to the opening of the prefecture p. So by employing this specification, we can examine both the pre-trends four years before the opening of prefectures and dynamic effects four years afterwards. $Open_{pc}^{\tau}$ equals 1 if year c is τ years after the opening of prefecture p and 0 otherwise. The omitted benchmark group is $\tau = -1$. Hence all estimates of β^{τ} should be interpreted as relative to one year prior to the opening shock. For our identification assumption not to be violated, we expect β^{τ} for all $\tau < 0$ to be statistically indifferent from zero.

To further shed light on the underlying mechanisms, we disaggregate further Huaren entrant counts at prefecture-surname-cohort-type level where type indicates the fourth dimension. This helps tease out underlying channels through which lineage connection works on foreign firm entry from differential responses to the lineage connection along the additional dimension. In analogy to Equation 2, we estimate

$$Y_{spc} = \eta_{sp} + \eta_{sc} + \eta_{pc} + \eta_x + \beta_1 \times Open_{pc} \times m_{sp} + \beta_2 \times Open_{pc} \times m_{sp} \times type_x + \lambda \times S_{spc} + \epsilon_{spc} \quad (4)$$

where $type_x$ is a dummy variable for firm type x. The sign of the estimate of β_2 determines whether the characteristic of type x entrants complements or substitutes the effects of lineage connection.

4.2   Baseline Results

We first estimate Equation 2 to evaluate how much lineage connection affects the entry and survival of Huaren firms following the prefecture opening. Results are presented in Table

3. Across all columns, we control for surname-prefecture, surname-cohort and prefecture-cohort fixed effects as well as the number of incumbent firms at surname-prefecture-cohort level. In column (1), result suggests that a surname with 0.1 higher lineage connection m_{sp} can get approximately 0.1767 more Huaren entrants in one year than other surnames if prefecture p started to be open to foreign investments in year c. And this effect is economically large as 0.1767 is approximately three times of the mean number of Huaren entrants across all surname-prefecture-cohort triplets. Column (2) confirms that not all entrants are short-lived by getting a slightly smaller yet still significant estimate with number of survival-adjusted entrants as outcome. Even if we use survived number of entrants in 2014 as outcome we still arrive at significant estimates in column (3), though much smaller. Columns (1) to (3) altogether suggest that lineage connection facilitates the entry of Huaren foreign firms no matter whether and how we adjust the quality of the entries by survival. And this result indicates that there are significant and economically large effect of lineage connection on foreign firm entries which persist (though becomes smaller) over time.

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

	Number of Entrants	Survival-adjusted Number of Entrants	Number of Survived Firms in 2014
	(1)	(2)	(3)
Mean of Dep. Var.	0.060	0.052	0.009
Open × Lineage Connection	1.767***	1.517***	0.218*
	(0.574)	(0.527)	(0.399)
Adj.R^2	0.570	0.555	0.399
N	1,344,421	1,344,421	1,344,421
Number of Incumbent Firms	Y	Y	Y
Surname-Prefecture FE	Y	Y	Y
Surname-Cohort FE	Y	Y	Y
Prefecture-Cohort FE	Y	Y	Y

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

As a crucial test for the identification assumption, we also employ the event-study spec-

ification in Equation 3 and show in Figure 4 that the entry and survival of Huaren foreign firms across surnames and prefectures had similar trends prior to the prefecture opening. This results supports our identification strategy and causal interpretation of the results. The dynamic effects of lineage connection in Figure 4 also evidently show that lineage connection has only short-term effects on Huaren firm entries.
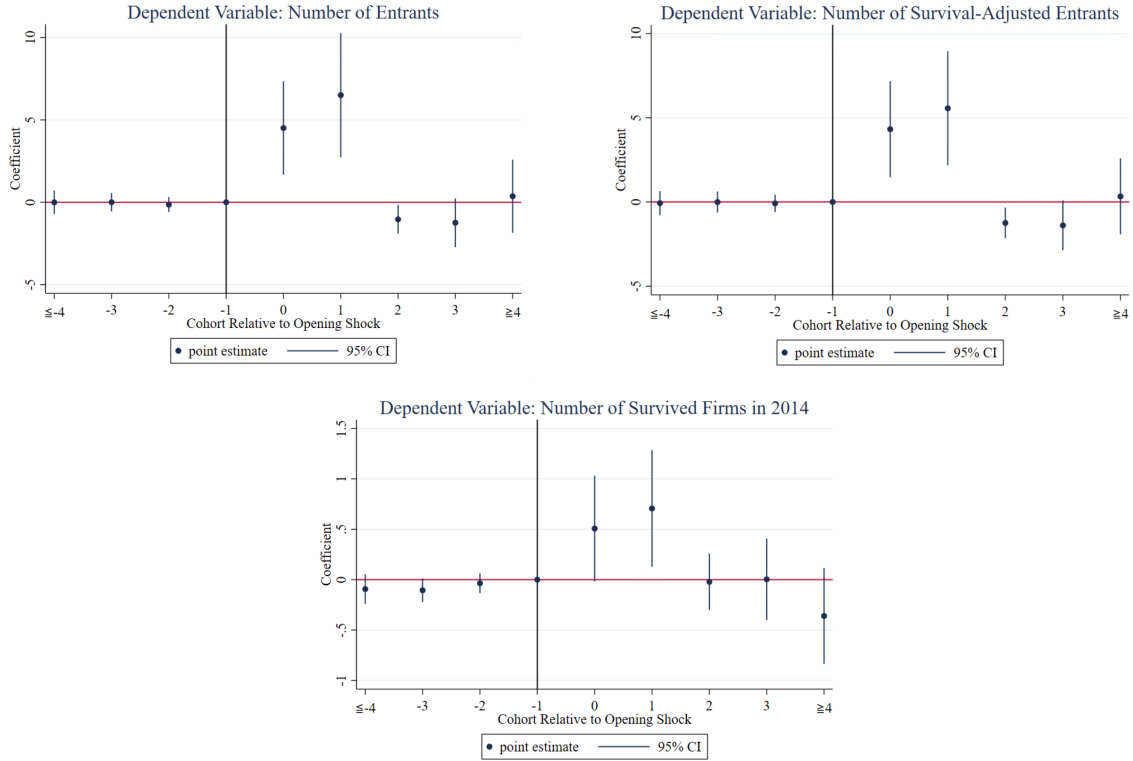
Figure 4: Event Study

4.3   Selection on Entrepreneur Quality

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. For ruling out this plausible explanation, Table 4 additionally looks at the effect of lineage connection on survival ratio and average regis-

27

tered capital of Huaren entrants in 2014 conditional on entry in order to identify if there are any lineage-driven selection of foreign entrants on inherent quality reflected in later survival and firm growth. Although we find effects of lineage connection on the counts of survived firms, we did not discover any benefit of lineage connection on survival probability and registered capital in 2014 of Huaren firms conditional on their establishments.

Table 4: Lineage Connection and Entry and Survival of Huaren Foreign Firms

	Survival Ratio in 2014 Conditional on Entry	Average Registered Capital in 2014 Conditional on Entry
	(1)	(2)
Open × Lineage Connection	-0.333	-7040.914
	(0.825)	(7984.062)
Adj.R^2	0.364	0.054
N	20,633	20633
Number of Incumbent Firms	Y	Y
Surname-Prefecture FE	Y	Y
Surname-Cohort FE	Y	Y
Prefecture-Cohort FE	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10% respectively. Standard errors are clustered at surname-prefecture level.

This result not only mitigates the concern of overestimating the effects of lineage connection on foreign firm entry but also, combined with the dynamic effects in Figure 4, suggests that the mechanism through which lineage takes effect is mainly lowering entry barriers. The result of Table 4 informs us that entrants driven by high lineage connection are not different in their underlying quality in terms of survival and growth of firms. Therefore this result motivates us to interpret the role played by lineage connection as lowering entry barriers in the short run regardless of Huaren firm's latent quality[18].

---

[18]In stylized industrial equilibrium models such as Melitz (2003), lowering fixed cost of entry would select entrants with lower productivity. However, survival and growth of firm are not necessarily monotone in productivity. Then reduced entry barrier won't imply a lower average survival probability nor average registered capital in 2014. In fact, survival and registered capital in 2014 are more related to firm growth which is rooted in unobserved entrepreneur quality (Guzman and Stern, 2020). And our results indicates that the operation of lineage in affecting foreign firm entry is independent with firm growth measured by survival and average registered capital in 2014 or in another word latent quality of entrants. We could further conduct a test on selection of low productivity firm if we have a productivity measure upon entry of the

4.4 Robustness Checks

There are many concerns on our main results about the effects of lineage connection on reducing entry barriers for foreign firms. We'd like, by taking full advantage of our unique data, to conduct a series of robustness checks to ensure that our main results do not suffer from those concerns.

Alternative Dependent Variables. We notice that the entry and survival of Huaren foreign firms are pretty scarce at surname-prefecture-cohort level (see Table 2). Therefore we want to at first make sure our results still hold if we consider only the extensive margin at surname-prefecture-cohort level. We then replace our firm count measures with dummies indicating whether there is at least one entry in a surname-prefecture-cohort triplet. Columns (1) to (3) of Table 5 show our findings that lineage connection significantly increases the entry and survival-adjusted entry of Huaren foreign firms at the extensive margin as well. We are also interested in the stability of our findings in cases where the measure of firm entry is adjusted by firm size and quality. Column (4) provides a test by regressing the inverse hyperbolic sine of total survived registered capital in 2014 as dependent variable, and still arrives at economically and statistically significant estimates.

Alternative Lineage Connection Measure. The second potential challenge to our results is the measurement error in our lineage connection. 2005 Census is the only data source where we could observe nationally representative distributions of surnames across prefectures. Since our sample period starts from 1980, it is possible that the internal migration between 1980 and 2005 could lead to changes in the spatial distribution of surnames and Huaren entrants at the same time. However, We firstly as argued previously gain confidence in our measure from past literature showing that the spatial distribution of surnames does not vary much throughout the history even after warfare and revolutions

---

firm in the data. However, what we have is registered capital of firms in 2014 which is still better understood as proxy for firm growth instead of static productivity. So unfortunately, we cannot examine whether the static selection emphasized in trade literature takes place here and this is not our focus either.

Table 5: Robustness Checks I: Alternative Dependent Variables

	At Least One Entrant	At Least One Survived Entrant in 4 Years	At Least One Survived Entrant in 2014	Arcsinh (Survived Registered Capital in 2014)
	(1)	(2)	(3)	(4)
Open $\times$ 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 FE	Y	Y	Y	Y
Surname-Cohort FE	Y	Y	Y	Y
Prefecture-Cohort FE	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10% respectively. Standard errors are clustered at surname-prefecture level.

(Clark, 2015) and recent work by Bai (2020) confirming the stability of spatial distribution of surnames by comparing the surname distributions constructed from 2005 population Census and the Harvard China Biographical Database (CBDB). We then further provide additional checks on our results taking advantage of the surname information available in our SAIC database. We construct from the SAIC database an alternative lineage connection measure using surnames of registered personnel who have worked in *domestic firms* established before 1992 to mitigate concerns on internal migration.

The results on entry outcomes using this alternative lineage measure are presented in column (1) of Table 6. We still get significant estimates for our main entry outcome variables as before and they remain economically large relative to the mean. Notice 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 should put more faith in our original estimates relative to the new one using the alternative measure of lineage connection. We can, however, still reach the same conclusion as before that lineage connection exerts positive impacts on the entry of Huaren foreign firms with and without being adjusted by survival.

Alternative Standard Error. Our baseline results come from regressions which

Table 6: Robustness Checks II: Alternative Setups

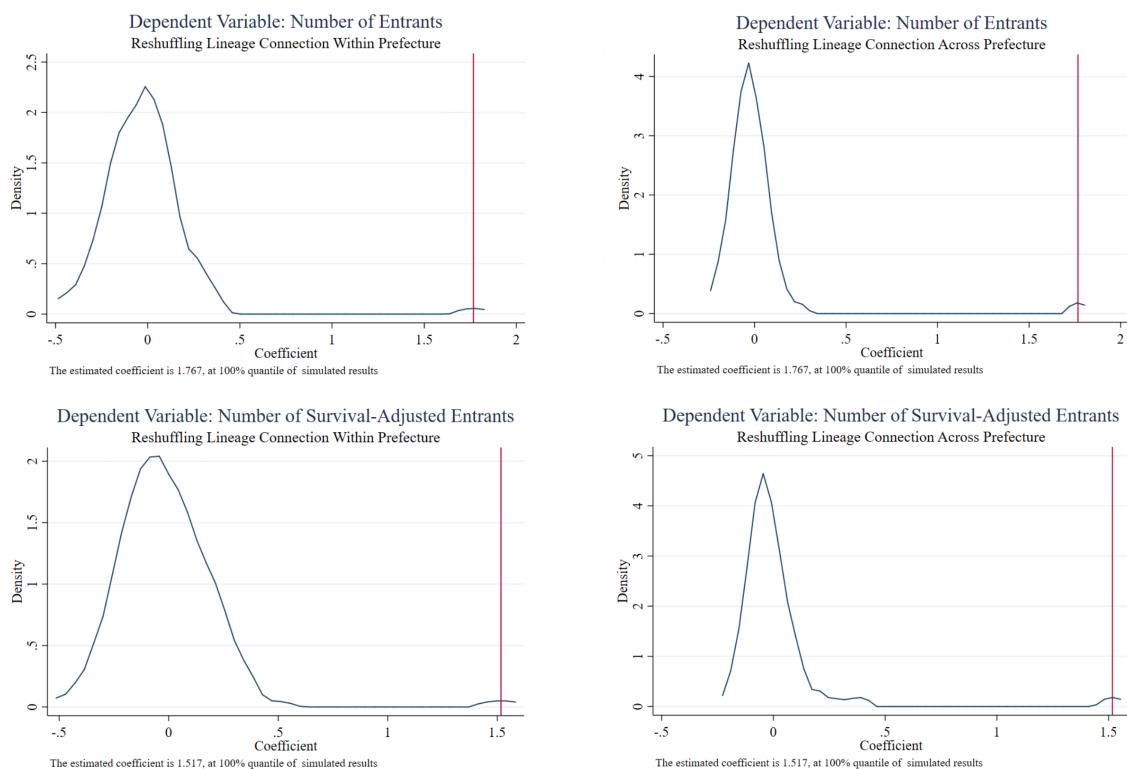|	Open × Lineage Connection			
	Alternative Lineage Connection Measure	Alternative Standard Error	Include Potential Roundtrip Huaren Entrants	WTO Accession Placebo Test
	(1)	(2)	(3)	(4)
Number of Entrants	2.874***	1.767***	3.821***	0.980
	(0.946)	(0.493)	(0.820)	(0.842)
Survival-Adjusted Number of Entrants	2.511***	1.517***	3.253***	0.739
	(0.903)	(0.456)	(0.743)	(0.726)
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FE	Y	Y	Y	Y
Surname-Cohort FE	Y	Y	Y	Y
Prefecture-Cohort FE	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10% respectively. Standard errors are clustered at surname-prefecture level except in column (2). Survival-adjusted Number of Entrants is the number of entrants that survive for more than 4 years (included). Potential roundtrip Huaren entrants are defined as entered foreign firms represented by citizens of People's republic of China. Column (4) uses *Open* status before 1992 as treatment while the WTO accession in 2002 as shock for our placebo test.

cluster standard errors at surname-prefecture level to account for serial heteroskedasticity within surname-prefecture pair. However, the error terms could be correlated at a higher level. The results in column (2) of Table 6 show that the significance of our estimates does not change much even if we cluster our results at prefecture level.

Including Roundtrip Huaren Entrants. The construction of our sample for baseline results in Table 3 is very cautious on the definition of Huaren foreign firms as policy environment was complicated and rapid-changing during the Reform Era. We took the most conservative approach by excluding those firms with foreign firm administrative code while represented by individuals with national identity card of China. In this way, we hope to eliminate the influence of the the prevalent fake foreign firms who constitute part of roundtrip FDI during the Reform Era. However, though unlikely, it is still possible that excluding those suspicious foreign firms might contribute to our results. For robustness, we add back those firms into our sample and show in column (3) of Table 5 that the estimates go up quite a lot. This further supports our sampling choice as those potentially fake Huaren foreign firms dramatically exaggerate the presence of Huaren foreign firms

and therefore drive up the estimated effects of lineage connection on foreign firm entries according to the results.

WTO Accession Placebo Test: Our understanding of the main results as lineage connection's capability to reduce entry barriers for Huaren foreign firms can be further tested by extending our sampling period beyond the Reform Era and using later WTO accession in 2002 as a placebo. If the lineage connection only works when barriers deterring entry of foreign firms exist, we should not see such differential effects of lineage connection across pre-1992 opened prefectures and closed prefectures before and after 2002 as China has spent decades to develop its market institutions for catering the demand of WTO and being accepted as member. Indeed as expected, column (4) of Table 5 shows that there are no significant effects of lineage connection on number of entrants and survival-adjusted number of entrants before and after the 2002 WTO accession.

Figure 5: Permutation Test

Permutation Tests. To further check the extent to which our results are influenced

by any errors in the lineage connection measures, we perform two permutation tests by randomly shuffling our lineage connection measures within and across prefectures. The left panel of Figure 5 plots the distributions of the estimates for our two main outcome variables from 100 simulations using reshuffled lineage connections across surnames within same prefecture while figure the right panel plots estimates in 100 simulations where lineage connections are reshuffled across prefectures within same surname. We see that all distributions are centered around zero and the actual estimates tend to lie at the far right end of the distributions. Taken together, these results suggest that the positive and significant effects of the lineage connection on entry and survival of Huaren foreign firms are not driven by unobserved measurement noises on lineage connection.

Table 7: Robustness Checks III: Subsample Regressions

	Open × Lineage Connection			
	Excluding Emigration-intensive Provinces	Excluding Emigration-intensive Surnames	Within Emigration-intensive Provinces	Excluding FDI-intensive Prefectures
	(1)	(2)	(3)	(4)
Number of Entrants	0.642***	1.733***	6.001***	1.534***
	(0.233)	(0.332)	(2.476)	(0.440)
Survival-Adjusted Number of Entrants	0.482***	1.480***	5.152***	1.308***
	(0.167)	(0.292)	(2.186)	(0.380)
Number of Incumbent Firms	Y	Y	Y	Y
Surname-Prefecture FE	Y	Y	Y	Y
Surname-Cohort FE	Y	Y	Y	Y
Prefecture-Cohort FE	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10% respectively. Standard errors are clustered at surname-prefecture level except in column (2). Survival-adjusted Number of Entrants is the number 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.

Subsample Regressions. Another challenge to the interpretation of our results is the possibility that variations from certain regions or surnames could completely drive our findings. For example, emigration-intensive provinces like Guangdong, Fujian and Zhejiang disproportionately exported more emigrants and attracted more foreign firms at the

same time than other places. And some surnames, whose population historically resided in emigration-intensive provinces, have persistently higher presence in both FDI-intensive regions and among Huaren entrepreneurs due to surname-specific expertise. Although these concerns do 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 remain positive and statistically significant in columns of Table 6 where we use different subsample regressions to test the robustness of our baseline findings. Column (1) and Column (3) together suggest that lineage connection takes effects not only within but also outside of emigration-intensive provinces, though unsurprisingly the magnitude of the coefficients within emigration-intensive provinces are indeed much larger. Likewise, column (2) verifies the robustness of our results to the exclusion of emigration-intensive surnames. And finally column (4) overrules the possibility that FDI-intensive prefectures as outliers in our sample drive all our observed patterns by excluding them from our regression sample.

5 Mechanisms

Until this section, we have documented, for the first time in literature, a large causal effect of surname-base lineage connection on foreign investments in terms of firm entries. We now turn to explore the underlying mechanisms through which the lineage connection exerts the observed effects on the entry of Huaren foreign firms. Social ties can work plausibly through demand side and supply side to benefit movement of economic factors between the two nodes of the tie.

As argued by Burchardi, Chaney and Hassan (2019), on the demand side, in the context of our study, lineage connection could facilitate entry of Huaren foreign firms as tastes could persist over generations and Huaren entrepreneurs might just work abroad in industries catering the demand of their ancestral hometown. However, compared to Burchardi,

Chaney and Hassan (2019) where variation is at origin country level, we compare surnames within prefectures and tastes by different surname groups are extremely unlikely to vary so much within prefecture for accounting for our results. On the supply side, as summarized in Rauch (2001), any kind of social networks in business plays two possible major roles: sharing information and enforcing contracts. So better lineage connection could either provide information advantage or social collateral for enforcing contracts for Huaren entrepreneurs. We discover evidence in support of both of them in the following empirical analysis testing auxiliary predictions of the two distinct mechanisms.

5.1 Overcoming Information Frictions

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 to outsiders 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 mails even during the time when communication infrastructure in China was in its 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. Hence lineage connection could mitigate the information frictions and creates an information advantage for Huaren entrepreneurs with better lineage connection than the others. However, information transmission requires conditions which were met in various degrees at different locations. So we test the information channel through looking at heterogeneous effects of lineage connection on Huaren entries across places with different predetermined circumstances.

Firstly, if Huaren entrepreneurs with higher lineage connection have better information access than others, those information in places with superior communication infrastructure would be more efficiently transmitted to Huaren entrepreneurs. In another word,

Table 8: Overcoming Information Frictions

	Number of Entrants	Survival-adjusted Number of Entrants
	(1)	(2)
Panel A: Communication Infrastructure		
Open × Lineage Connection	-2.507	-1.814
	(1.908)	(1.846)
Open × Lineage Connection × Telephone Exchange Capacity	1.074***	0.868**
	(0.409)	(0.386)
Adj.R^2	0.454	0.437
N	666,397	666,397
Panel B: Social Fragmentation		
Open × Lineage Connection	-9.356***	-7.201***
	(2.936)	(2.899)
Open × Lineage Connection × Surname Fragmentation	11.963***	9.365***
	(3.336)	(2.899)
Adj.R^2	0.409	0.395
N	1,186,664	1,186,664
Panel C: Pioneer Firm		
Open × Lineage Connection	7.424**	6.390**
	(3.063)	(2.712)
Open × Lineage Connection × Pioneer Firm	-13.080**	-11.263**
	(5.650)	(4.974)
Adj.R^2	0.465	0.458
N	2,690,048	2,690,048
Number of Incumbent Firms	Y	Y
Surname-Prefecture FEs	Y	Y
Surname-Cohort FEs	Y	Y
Prefecture-Cohort FES	Y	Y
Additional Controls	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level. 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 foreign entrant in a 2-digit industry in a prefecture) in Panel C.

we conjecture that the effect of lineage connection on Huaren firm entry would be higher in places of lower information transmission cost if lineage connection increases the supply of information to Huaren entrepreneurs. To empirically test this auxiliary prediction of information channel, we examine the heterogeneous effects of lineage connection on Huaren foreign entries across province-years with differential preexisted telephone exchange capacity. Back in 1980s and 1990s, the telephone exchange capacity determined whether Huaren entrepreneurs could communicate easily with local people they were connected to. For achieving the goal, we include an additional term in our baseline regression which interacts our main independent variable with telephone exchange capacity at province-year level. Panel A of Table 8 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 the expectation that communication infrastructure are complementary with lineage connection since Huaren can use the communication infrastructure to access the more information supplied by Huaren entrepreneurs' local contacts following the opening reform.

Secondly, in places with higher social fragmentation, information search cost could be larger according to Padró i Miquel, Qian and Yao (2012) as information sources are fragmented. Then for less connected Huaren entrepreneurs, searching for local information would be harder in such places. Therefore lineage connection would have higher effects on Huaren entries in places with higher social fragmentation. In another word, higher social capital brought by lineage connection could lower the cost of searching in environment with more fragmented information sources. Consistent with this prediction, we indeed discover in Panel B of Table 8 stronger effects of lineage connection in prefectures with higher social fragmentation as measured by one minus the HHI index of surname concentration. This result further supports the information channel for explaining the large causal effects we identified of lineage connection on Huaren foreign firm entry.

Lastly, for entrepreneurs, local information would be more useful and abundant if

there are already entrants in the same industry at that location. In another word, with pioneer firm defined as the first foreign entrant in a 2-digit industry in a prefecture, the effects of lineage connection on Huaren entries would be lower for pioneering firm who won't have much useful same-industry information to access through their lineage connection. For testing this prediction, we further organize our data to distinguish pioneering firms from imitating firms when conducting aggregation at surname-prefecture-cohort-type level as Equation 4 in which type is referred to pioneer dummy here. Panel C of Table 8 demonstrates that Huaren firms of stronger lineage connection tend to enter into existing industries as imitators rather than new industries as pioneers following opening reform. This suggests that information on proven business opportunities could pass through lineage group, therefore inducing more imitators rather then pioneers 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). To be noticed that, the effect of lineage connection on pioneering firms is actually negative. That's to say higher lineage connection even redistribute Huaren entrepreneurs from being pioneering to imitating. However, this does not mean lineage connection does not help motivate entry of pioneers. Most of the pioneering foreign firms are still Huaren firms as seen in Figure 2. And as pioneers induce imitators, these early pioneering Huaren entrants could have long-term effects on future local economic development identified in the next section.

5.2 Strengthening Contract Enforcement

Weak contract enforcement has also long been seen as key barrier deterring foreign investments toward developing countries where there is a lack of formal contracting institution. And informal institution such as lineage in Chinese culture could be useful strengthen contract enforcement under such circumstances as argued by Greif (1993) and Greif and Tabellini (2017). If lineage connection indeed helps address the problem of contract en-

forcement, we conjecture the effects would be higher in situations where demand for contract enforcement is higher. In the following discussion, we present how we test this auxiliary prediction from three aspects and show the results which all support the contract channel in Table 9.

When foreign entrepreneurs consider entering China, they face this decision whether to setup the firm as headquarter in China or branch of headquarter in their home country. The former choice needs more commitment than the letter as maintaining a headquarter in foreign entrepreneur's home country enable revolving part of the problems faced by the firm there instead of in a foreign country. Therefore entering as a headquarter firm needs better contract enforcement than branch firm. Then from the perspective of contract channel, we conjecture lineage connection would have larger effect on Huaren firm entry as headquarter firm. Panel A of Table 9 finds that Huaren foreign firms with higher lineage connection are less likely to enter and survive as a branch firm. This results is consistent with the auxiliary prediction implied by contract channel and therefore suggests that lineage connection does help strengthen contract enforcement.

Doing overseas business also involves negotiating over two major kinds of contracts: one is with local employees and the other with local suppliers. If foreign entrepreneur needs to hire a lot of local employees or contract with a lot of local suppliers, the demand for better contract enforcement would be very high. We hypothesize then lineage connection may facilitate contract enforcement more for Huaren entrants in industries with either higher labor intensity or intermediate intensity. To formally test the hypothesis, we further breakdown entrants into labor-intensive industries and non-labor-intensive industries (for manufacturing subsamples only), supplier-intensive industries and non-supplier-intensive industries. Results in Panel B and Panel C indicate that Huaren foreign firms with stronger lineage connection do have comparative advantage in labor-intensive and supplier-intensive industries. These results suggest that lineage connection helps mediate access to labor and intermediate inputs through smoothing contract enforcement.

Table 9: Strengthening Contract Enforcement

	Number of Entrants	Survival-adjusted Number of Entrants
	(1)	(2)
Panel A: Branch Firm		
Open × Lineage Connection	8.795***	7.633**
	(3.363)	(2.997)
Open × Lineage Connection × Branch Firm	-15.823***	-13.749***
	(6.255)	(5.547)
Adj.R^2	0.452	0.444
N	2,690,048	2,690,048
Panel B: Labor Intensive		
Open × Lineage Connection	-1.864*	-1.650*
	(1.043)	(0.928)
Open × Lineage Connection × Labor Intensive	4.990**	4.433**
	(2.466)	(2.226)
Adj.R^2	0.665	0.645
N	2,690,048	2,690,048
Panel C: Supplier Intensive		
Open × Lineage Connection	-2.331*	-2.030**
	(1.156)	(1.010)
Open × Lineage Connection × Supplier Intensive	6.196**	5.397**
	(2.736)	(2.427)
Adj.R^2	0.688	0.673
N	2,690,048	2,690,048
Number of Incumbent Firms	Y	Y
Surname-Prefecture FEs	Y	Y
Surname-Cohort FEs	Y	Y
Prefecture-Cohort FES	Y	Y
Additional Controls	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level. 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. Supplier-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).

6 Seeding Effect on Long Run Foreign Investments

We have estimated the causal effects of lineage connection on the entry of Huaren foreign firms during the take-off stage of Chinese economy. However, the short-run surge of entries during the Reform Era may or may not have meaningful impacts on Chinese economy. This section, taking one step further, asks whether those industrial "seeds" planted in the form of the early Huaren FDI had long-lasting effects on future economic development. For achieving this goal, we employ the lineage-driven FDI predicted by our baseline regression as instruments for the early Huaren FDI and therefore are able to estimate causally the persistent effects of early Huaren FDI on future foreign investments.

6.1 Isolating Lineage-driven FDI

Motivated by a new strategy in Sequeira, Nunn and Qian (2020), We draw results from our baseline regression in Equation 2 (the "zero stage regression") and isolate lineage-driven FDI that arises from the interaction of surname distribution and the gradual opening of Chinese prefectures during the Reform Era. 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 through channels other than early Huaren FDI inflows to the prefecture. So the exclusion restriction for using linage-driven FDI as instrument for early Huaren FDI would be very likely to hold.

The IV regression have three stages as Sequeira, Nunn and Qian (2020). In zero stage, we estimate our baseline regression in Equation 2 with number of entrants as outcome and get estimates $\hat{\beta}$. With the estimates, we are able to get the predicted value $\hat{\beta} \times Opening_{pc} \times m_{sp}$ as what we call lineage-driven FDI at surname-prefecture-cohort level. We then aggregate them over all surnames and cohorts to get a predicted value at prefecture level:

$$FDI_p^l = \sum_s \sum_{c=1981}^{1996} \hat{\beta} \times Opening_{pc} \times m_{sp} \tag{5}$$

This value represents the lineage-driven FDI for each prefecture in year 1996. We also have observed stock of Huaren entrants for each prefecture in year 1996. Then the first stage regression where we use lineage-driven FDI as IV for observed Huaren FDI can be run as following:

$$FDI_p = \alpha + \theta \times FDI_p^l + \lambda \times X_p + \epsilon_p \tag{6}$$

where FDI_p is the observed number of Huaren entrants in 1996 and X_p is a set of controls at prefecture level. And with predicted prefecture-level number of Huaren entrants from this first-stage regression, we can finish our 2SLS by running the second-stage regression:

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

where π_p is our long-term outcome of interest.

Compared to past literature on identifying FDI spillovers in China, our paper makes several significant contributions. Firstly, we extend the research periods to the early 1980s when the spillover effects of early FDI in China which we call the "industrial seeds" tend to be strongest. In contrast, most existed studies only focus on late 1990s to mid 2000s due to data limitations. Secondly, our paper draws upon the most comprehensive dataset on FDI covering all sectors while past literature mainly use the above-scale manufacturers[19]. Last but not the least, our paper establishes the causal link between early Huaren FDI and future foreign investments in the long run. This fills the gap between anecdotal evidences on the importance of overseas Chinese in China's growing to be hub for global FDI and no formal empirical analysis (even description) examining the role played by Huaren investors in the process of China's Great Reform and Opening-Up.

---

[19]See Appendix E for a detailed discussion.

6.2 Identifying Long Run Effects on Foreign Investments

As FDI literature has pointed out, FDI inflows into developing countries do not necessarily promise economic gains in the long run. Indeed, FDI has been found to be 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 growth of domestic sectors (Harrison and Rodríguez-Clare, 2010; Lu, Tao and Zhu, 2017). In this subsection, we examine whether the lineage-driven FDI which is the dominant form of inward FDI during the period when China's market economy was largely immature has any long term effect on the future foreign investments.

For our purpose, we implement the methodology described in Section 6.1. Firstly, we obtain our instrument from the predicted number of entrants calculated based on previous results of baseline regression. We then sum up the number of Huaren foreign firms survived in 1996 to construct lineage-driven FDI at the prefecture level $F\hat{D}I_p$ and use it as dependent variable in the first stage (Equation 6). And we perform second-stage regressions for a series of long-run measures of foreign investments including non-Huaren foreign firm stocks, all foreign firm stocks and these survived firms' total registered capital in 2014. In all specifications, we additionally control for initial economic development measured by GDP per capita in 1996, years since the Reform and Opening-Up, distance to sea, social fragmentation index, SEZ dummy, *Open Coastal City* dummy and province fixed effects.

The results are presented in Table 10. We find that prefectures with more historical lineage-driven FDI today has higher foreign investments from both non-Huaren and Huaren in terms of both survived number of firms and registered capital in 2014. And we call this seeding effect of early Huaren FDI on future foreign investments. This result further supports our assertion of the influential role played by overseas Chinese in attracting foreign investment not only in early stage of the great Reform and Opening-Up but also

Table 10: Seeding Effect of Huaren Entry on Future Foreign Investment

	Non-Huaren Firm Stocks in 2014	All Foreign Firm Stocks in 2014	Registered Capital of Non-Huaren Firms in 2014 (10^6 CNY)	Registered Capital of All Foreign Firms in 2014 (10^6 CNY)
	(1)	(2)	(3)	(4)
	Panel A: Foreign Investment in 2014			
Huaren Firm Stocks in 1996	0.696***	2.593***	15.276***	63.256***
	(0.273)	(0.915)	(6.239)	(25.460)
N	231	231	231	231
	Panel B: First Stage Dependent Variable: Huaren Firm Stocks in 1996			
Lineage Driven Huaren Firm Stocks in 1996		30.953***		
		(14.391)		
Log(GDP) in 1996	Y	Y	Y	Y
Years since Opening	Y	Y	Y	Y
Distance to Sea	Y	Y	Y	Y
Social Fragmentation	Y	Y	Y	Y
SEZ dummy	Y	Y	Y	Y
Open Costal City dummy	Y	Y	Y	Y
Province Fixed Effects	Y	Y	Y	Y

Note: ***, **, * denote significance level at 1%, 5%, and 10%, respectively. Standard errors are clustered at surname-prefecture level.

later period when China has already become the hub for global FDI through the seeding effects. We further decompose the seeding effect into two parts: spillover effects on non-Huaren foreign investments and within-lineage persistent effects on Huaren investments. As expected, the estimates of spillover effects in column (1) and (3) are always smaller than the estimates of within-lineage effects in column (2) and (4) of Panel A. The significant positive spillover effect on non-Huaren foreign investments in 2014 is especially surprising as we usually label foreign investments by Huaren as of low-quality projects seeking for short-term rents during China's transition to market economy in anecdotal evidences and therefore expect the lineage-driven Huaren FDI would be detrimental to the entry of other foreign investments. One plausible explanation for this result, combined with our finding on the effects of lineage connection and the underlying mechanisms, is the better information access and contracting environment brought by these Huaren firms spillovers to non-Huaren foreign firms.

7 Conclusions

The lineage connections between Huaren foreign firms and their ancestral hometowns are believed to be one of the most important drivers for China's remarkable industrial growth over the last forty years. However, only anecdotal evidences exist while systematic description and formal empirical analysis are absent. Our paper, for the first time in literature, examines the effects of lineage connection on foreign investments. For achieving this goal, we proceed in two steps. Firstly, we identify the effects of lineage connection on Huaren firm entries during the Reform Era when attracting foreign investments was difficult and foreign investments by overseas Chinese was dominant. Then we, taking advantage of the lineage-driven FDI identified in the first step to instrument early Huaren FDI, estimate the long-run effects on foreign investments by non-Huaren and all investors.

By defining Huaren firm as foreign firms with overseas Chinese as legal representative, We documents the prevalence of Huaren foreign firms during the take-off period of Chinese economy and causally identifies the effect of lineage connection on the entry of Huaren foreign firms leveraging a unique dataset containing the universe of Huaren foreign firms and the surnames of their Huaren legal representatives. We also show that the effects are primarily driven by reductions in the information frictions and contracting costs which lower the entry barriers for the connected Huaren foreign firms. This is done by examining the heterogeneous effects of lineage connection across prefectures with different telephone exchange capacity, social fragmentation, across firms which are pioneers and followers, branches and headquarters and across industries which are labor intensive or not and supplier intensive or not. We also find that these early lineage-driven FDI exerts positive spillovers on non-Huaren foreign investments in the long run. Our empirical evidences therefore suggest that the benefits of introducing FDI as industrial seeds seem to outweigh its costs during the take-off stage of an developing economy.

This paper contributes to an improved understanding of the FDI policy for the devel-

oping countries. While opening-up has numerous potential benefits for developing countries, it is often hard for them to kick start the virtuous cycle. Being lack of information access and facing with severe contracting friction, developing countries often struggle with setting up exporting channels, attracting foreign direct investments, let alone fully enjoy the benefits of opening. Findings of this paper may provide a practical solution: 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 indeed realized 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 being one of the largest, even before the accession to WTO.

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.

Branstetter, Lee, and Nicholas Lardy. 2006. "China's embrace of globalization." National Bureau of Economic Reseach.

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.

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.

Guzman, Jorge, and Scott Stern. 2020. "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." *American Economic Journal: Economic Policy*, 12(4): 212–43.

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.

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.

Melitz, Marc J. 2003. "The impact of trade on intra-industry reallocations and aggregate industry productivity." *econometrica*, 71(6): 1695–1725.

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 registered 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 6 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 6.

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 ethnic 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) Chinese 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 ethnic 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.
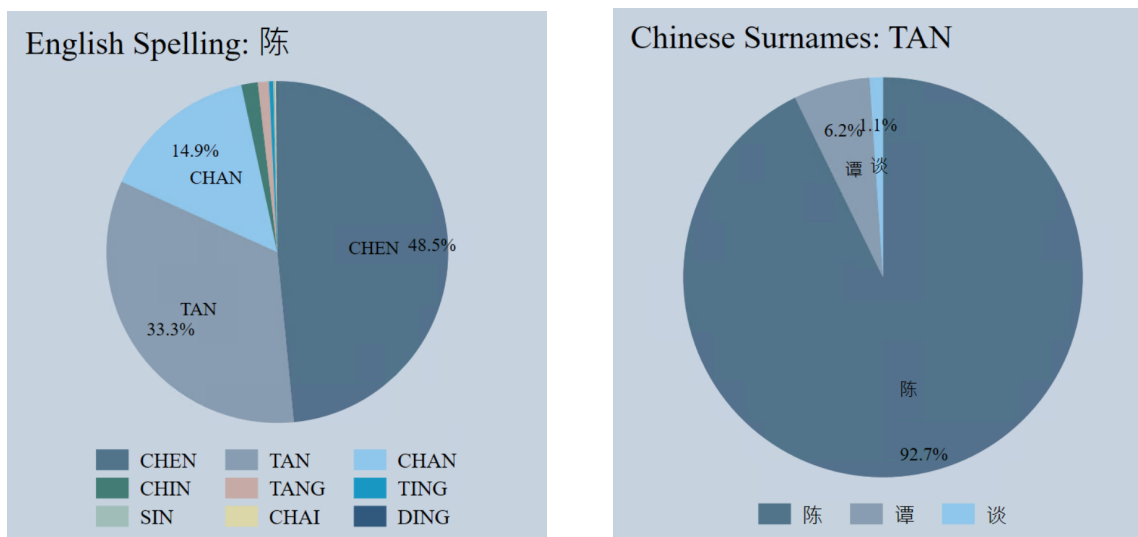
Figure 6: An Example of Spelling-Character Mapping between English and Chinese

B   Legal Representative and Executive Position

Table 11: 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 registered 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 12: 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 registered personnel working in foreign firms ever existed from 1985 to 2014.

C Lineage Connection vs. Population Size
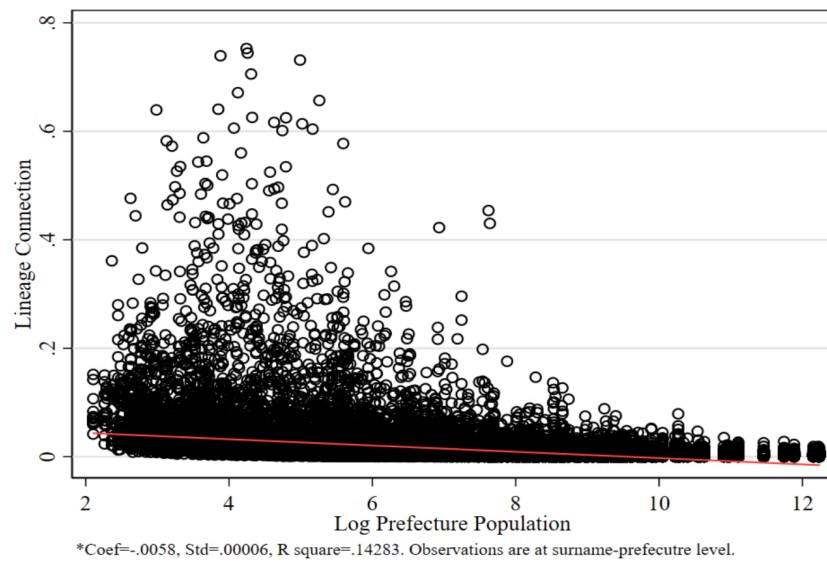
Figure 7: Lineage Connection Measure vs. Population Size

D   Registered Capital vs. Other Economic Outcomes

Table 13: 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.

E   Comparison with Other Data Sources

In this section, we illustrate the advantage of our comprehensive foreign firm dataset and cross-validate our data with other data sources.
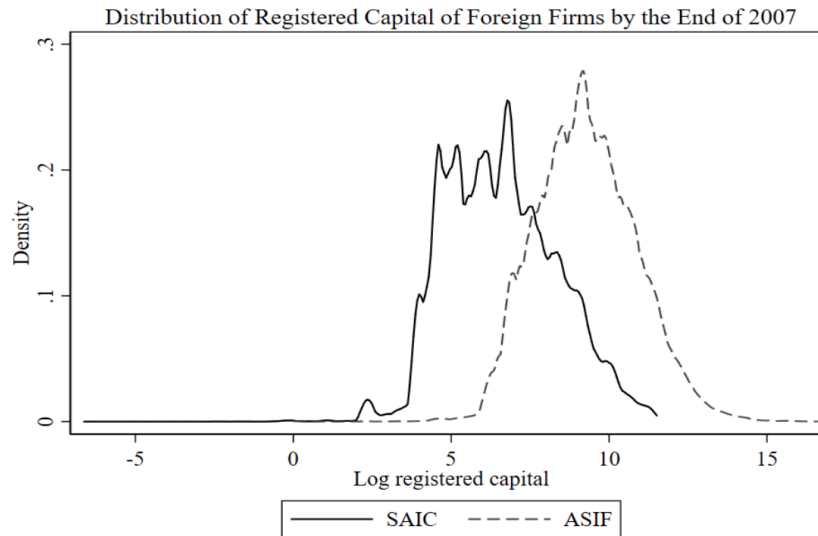

Figure 8: SAIC v.s. ASIF: Firm Size

Foreign firms covered by ASIF vs. by SAIC. Another frequently-used firm-level dataset is the Annual Survey of Industrial Firms (ASIF). We 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 for foreign firms of all sizes while ASIF only covers large firms with sales volume above 5 million CNY. Figure 8 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 9 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 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 10.
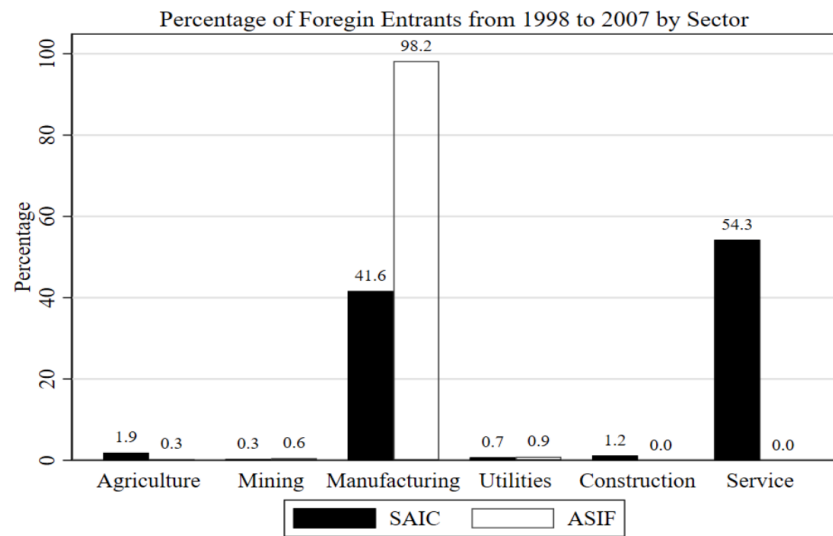
**Figure 9:** SAIC v.s. ASIF: Sector
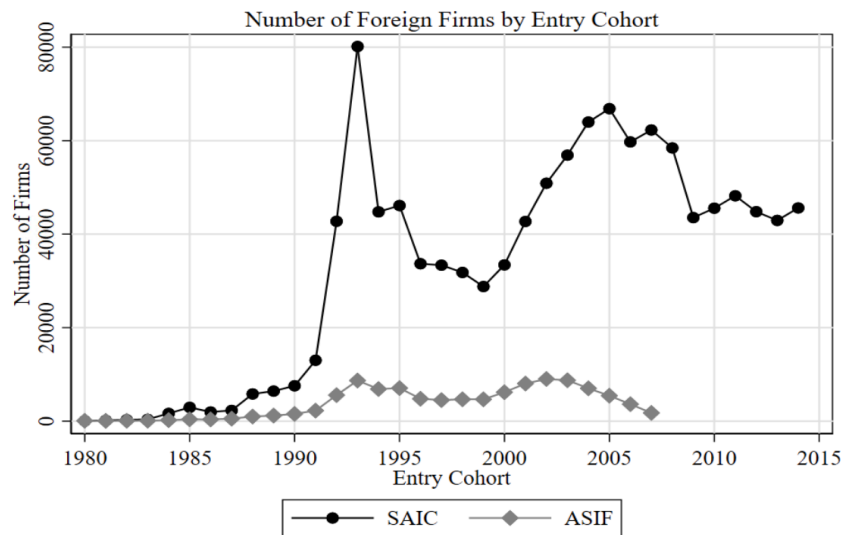
Figure 10: SAIC v.s. ASIF: Counts of Entrant

Huaren foreign firms vs. HMT firms.. Due to data limitation, prior literature 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 invested in mainland China through Hong Kong as a conduit; Second, residents in HMT only make up a fraction of total overseas Chinese. As can be seen from Figure 11, 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. This is consistent with the declining importance of Hongkong.
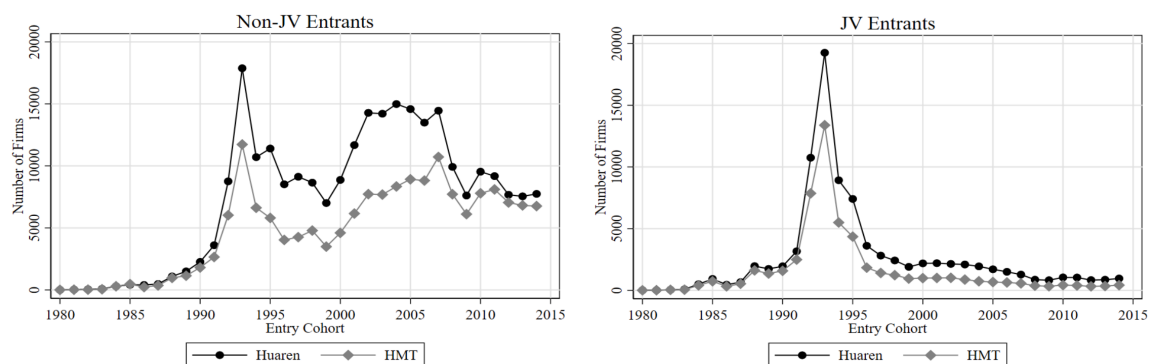

Figure 11: Huaren Foreign Firms vs. HMT firms
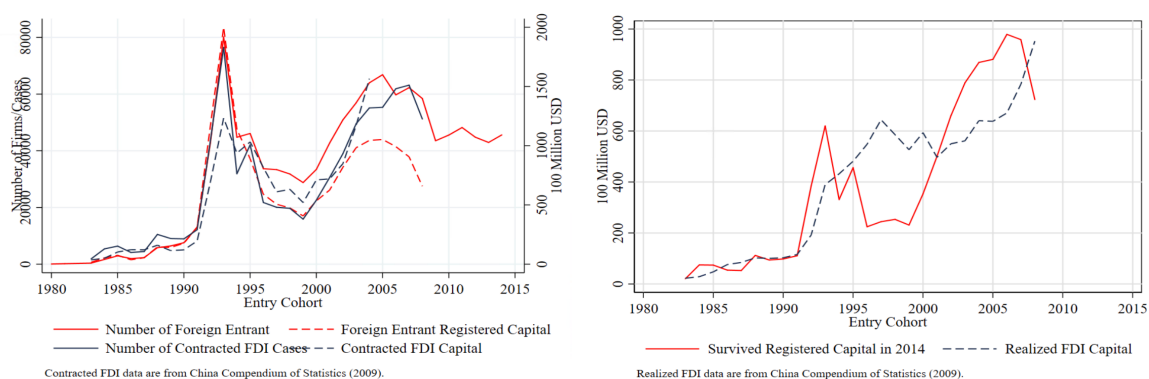
Figure 12: Foreign firms vs. official statistics of FDI inflows

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 12 suggests that the contracted FDI, measured in number of cases or USD, is highly correlated with our measure of foreign firm entries. And the realized FDI measured in USD is highly correlated with our measure of survived registered capital of foreign firms by 2014.
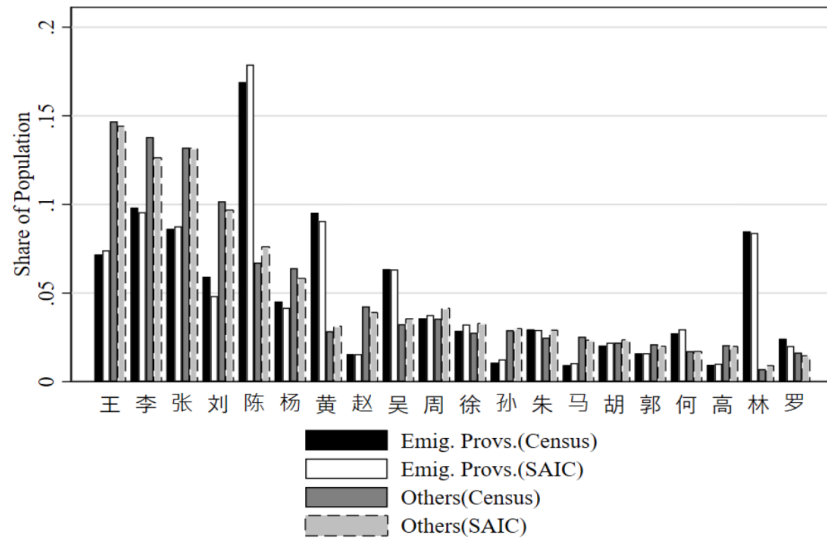
Figure 13: Surname Distribution: Census vs. SAIC

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

---

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

*Coef=.756, Std=.007, R square=.326. Observations are at surname-prefecture level.
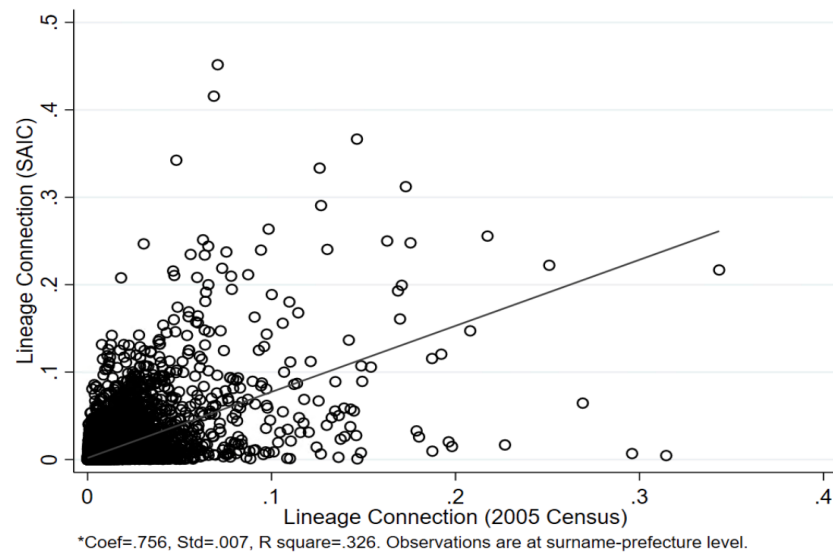
Figure 14: Lineage Connection Measures: 2005 Census vs. SAIC