

# **Familiar Strangers: Lineage Connection and Diaspora Direct Investments in China**

**Fanghao Chen**

Peking University

**Ruichi Xiong**

University of Toronto

**Xiaobo Zhang**

Peking University & IFPRI

March 3, 2022

<sup>1</sup>Fanghao Chen: [fhchen2017@nsd.pku.edu.cn](mailto:fhchen2017@nsd.pku.edu.cn), National School of Development at Peking University. Ruichi Xiong: [ruichi.xiong@rotman.utoronto.ca](mailto:ruichi.xiong@rotman.utoronto.ca), Rotman School of Management at University of Toronto. Xiaobo Zhang: [x.zhang@gsm.pku.edu.cn](mailto:x.zhang@gsm.pku.edu.cn), Guanghua School of Management at Peking University and International Food Policy Research Institute. Ruichi Xiong and Fanghao Chen acknowledge the financial support from China Research Initiative Grant at Rotman School of Management at University of Toronto. We thank Nathaniel Baum-Snow, April Franco, Kailin Gao, Ruixin Wang, Shuo Chen as well as seminar participants at University of Toronto, Sun Yat-Sen University, Harbin Institute of Technology (Shenzhen), Jiaxing University, University of Nottingham Ningbo, Peking University, Southwestern University of Finance and Economics and European Winter Meetings of the Econometric Society for helpful comments and discussions. All errors remain our own.

## **Abstract**

This paper studies the importance of international lineage network in facilitating the entry of foreign firms in developing countries where immature market environment deters foreign investments. We build a unique dataset on the universe of diaspora foreign firms in China that are controlled by 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 diaspora firms, through information sharing and contract enforcement. Our findings are robust to various robustness checks and we find no such effects following China's accession to WTO as placebo shocks because market institutions had been much improved. We also find that the lineage-driven diaspora firm entrants during early development stages functioned as industrial seeds, as they generate large and long-term positive spillover effects on non-diaspora firm entry and investments.

*Key Words:* International Migration, Diaspora, Lineage Network, Foreign Direct Investment, FDI Spillover, Seeding Effect, China

*JEL Codes:* F22, F23, L14

# 1 Introduction

Despite the broad positive spillovers associated with foreign direct investments (FDI) (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<sup>1</sup>. 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 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 diaspora firms controlled by overseas Chinese<sup>2</sup>, who though as emigrants are believed to have strong lineage connection with their ancestral hometowns<sup>3</sup>. They pioneered in an immature Chinese market where policies and institutions were uncertain and insecure, used informal networks to overcome entry barriers and conduct diaspora direct investment (DDI)<sup>4</sup>, as well as planted the seed investment for later industrial

---

<sup>1</sup><https://data.oecd.org/fdi/fdi-flows.htm>

<sup>2</sup>Throughout this paper, overseas Chinese is defined as ethnic Chinese that hold citizenship of foreign countries or Hong Kong, Macau, and Taiwan. And diaspora firms are defined as foreign firms with legal representative who are overseas Chinese.

<sup>3</sup>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 worshipping ancestors at lineage temples 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).

<sup>4</sup>In this paper, we mainly measure DDI by the number of new entrants of overseas-Chinese-run enter-

growth of China (Vogel, 1990; Lever-tracy, Ip and Tracy, 1996). Yet no study so far, to the best of our knowledge, has ever systematically documented the importance of DDI, not to mention empirically examined the causal relationship between lineage connection and DDI in China. We contribute to a better understanding of the role played by overseas Chinese in facilitating foreign investments into China, during a period even 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 largely underdeveloped, and this period witnessed DDI as the dominant form of FDI. Focusing on the Reform Era enables us to fully uncover the effects of lineage connection on DDI, and to explore the underlying mechanisms. And we focus on one particular form of DDI: the entry of diaspora firms.

The analysis proceeds in several steps. In the first step, we use firm registration data to screen out all the diaspora firms controlled by overseas Chinese according to the official ownership classification assigned by State Administration of Industry and Commerce of China (SAIC) and reported names of registered personnel of firms. Specifically, we identify foreign firms by ownership code and determine from the surname and the ID of the foreign firm's legal representative whether a foreign firm is controlled by overseas Chinese. The above procedure leads to our final sample of diaspora firms. Then we can use the observed surnames of the overseas Chinese legal representatives to measure diaspora firms' lineage connections with different prefectures of China. Our measure between surnames and prefectures characterize the probability that one overseas Chinese has ancestral hometown in any prefecture.

Second, we implement our main empirical design to identify the effect of lineage con-

---

prises or diaspora firms each year. There are two reasons: First, compared with the overall DDI (including loans, donations, and capital investments), the DDI driven by the establishment of firms is a more direct and constructive force in driving local industrial growth; and second, compared with the stock of diaspora firms, the number of new entrants can better measure the marginal response of entrepreneurs toward the changes in the economic environment, and in the early stage of China's Reform and Opening up, the entry of firms had been the major driver of economic growth(Wei and Zhang, 2011).

nection on diaspora firm entry. 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, fixed effects at surname-prefecture, surname-year, and prefecture-year level can be controlled 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 number of diaspora entrants and the number of the survival-adjusted diaspora entrants. Event studies show that our empirical design satisfies the common trend assumption. Besides, our findings remain robust to various alternative settings. And reassuringly, we find no effects of lineage connection there in permutation tests shuffling our lineage connection measure. Last, for ruling out potential selection on firm latent quality, we confirm that lineage connection does not contribute to the survival ratio and average survived registered capital in 2014 of diaspora entrants conditional on their entry. This also suggests lineage mainly takes effect through reducing entry barriers.

The third step sheds light on the mechanisms through which lineage connection helps overcome the entry barriers for diaspora firms. Our first finding is that using WTO accession in 2002 as opening shock gives us no effect on entry of diaspora firms. This indicates that lineage only works in immature market environment. We then find that the effects of lineage connection are stronger in prefectures with higher social fragmentation or better telecommunication infrastructure. We also find that diaspora firms with higher lineage connection are more likely to start a business as an imitator of existing industries. These three results all support information sharing mechanism. Then effects are also found to be stronger in headquarter firms (relative to a branch firms) and in labor-intensive and supplier-intensive industries. These findings reveal the role of lineage connection in enforcing contracts.

The final question is whether the lineage-driven DDIs during the early stage of China's opening-up have spillover effects on non-diaspora firm entry and investments and how

large they are. We show that the spillovers happen in the form of seeding effects in the long run through event studies. We then implement a strategy similar to [Sequeira, Nunn and Qian \(2020\)](#) to isolate lineage-driven DDI across prefectures that is independent of various confounding factors and get large long-term effects of these lineage-driven DDI on the presence of non-diaspora foreign firms and domestic private firms in terms of both counts and registered capital in 2014. This suggests that the benefits of introducing DDI 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 DDI 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 the causal effect of 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 as identification variation. Our paper, for the first time in this literature, studies surname-based lineage connection.

This paper also complements the literature on the long term implications of international migration. Existing researches primarily focus on examining the effect of immigration on the developed countries (see for example, [Sequeira, Nunn and Qian \(2020\)](#)). Researches studying the effect of emigration on the origin (usually developing) countries,

on the other hand, 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 diaspora firms in China. Overseas Chinese networks are believed to be vital to the development of China while empirical evidences are scant if not absent. We show that an unexpected benefit of emigration is to facilitate foreign investments through migrant networks that provide seed investments for long-term development of the home country. In addition, our usage of surnames to uncover the ancestral origins of overseas Chinese 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 surnames in the context of Chinese lineage groups. Our findings also contribute to the vast debate on Brain Drain v.s. Brain Gain roles played by international migration. The results we found point to one particular channel through which international migration can lead to Brain Gain in the spirit of [Saxenian \(2007\)](#).

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. First, we paint a systematic picture of diaspora 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 the causal effect of lineage connection on diaspora firms and uncover the underlying mechanisms. Besides, we also provide evidence in support of the seeding effects of lineage-driven DDI on long-term presence of non-diaspora firms in China as a new FDI spillovers other than the traditional vertical/horizontal ones across sectors intensively studied in the context of China (A recent example is [Lu, Tao and Zhu \(2017\)](#)). These findings establish the importance of overseas Chinese in China's development process and have profound policy implications for the other developing countries.

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 DDI on future non-diaspora investments. Section 7 concludes.

## 2 Historical Background

Ever since its dramatic turn toward market economy in 1978, China has swiftly transformed from a completely closed state to a hub for foreign firms. After 1978, when Deng Xiaoping initiated the great Reform and Opening-up, China started to gradually open its regions to international trade and foreign investments ([Lever-tracy, Ip and Tracy, 1996](#); [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 year 1980, four SEZs including Shenzhen, Zhuhai, Shantou and Xiamen were established where foreign corporations enjoyed broad autonomy and preferential tax treatment. 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 open regions in 1985, 1987 and 1988 respectively. Although the economic and political crisis in 1989 temporarily brought considerable uncertainties to China’s transition, Deng’s visit to the 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 capitals.

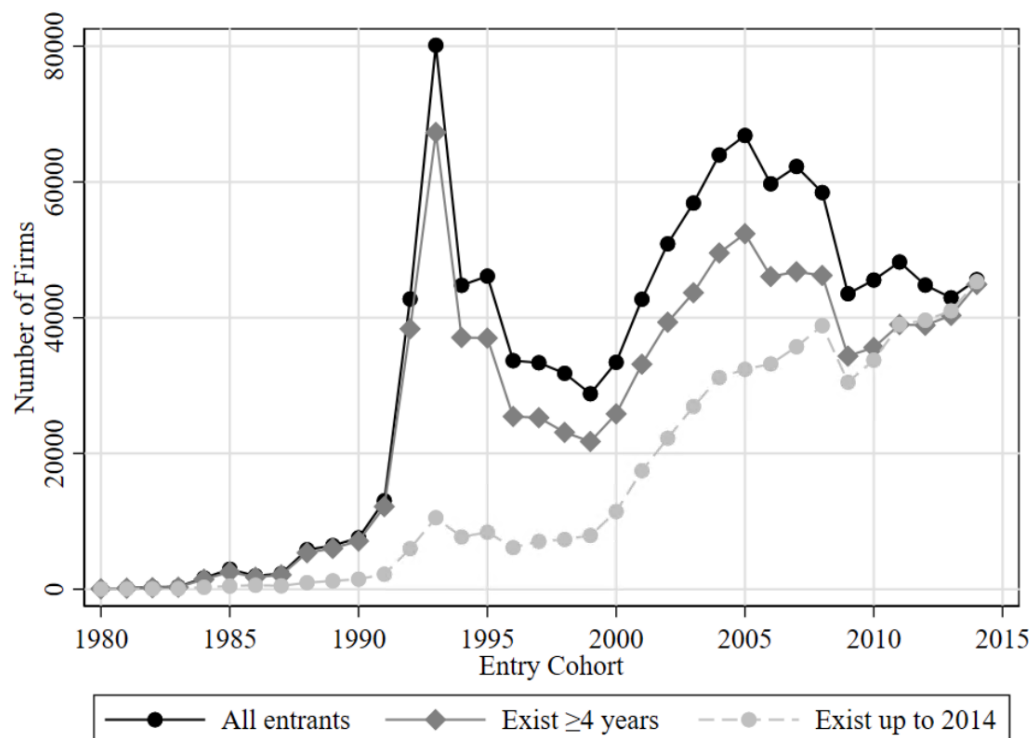
**Table 1:** China’s Opening Process during the Reform Era

Year	Opening 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](https://en.wikipedia.org/wiki/Chinese_economic_reform)

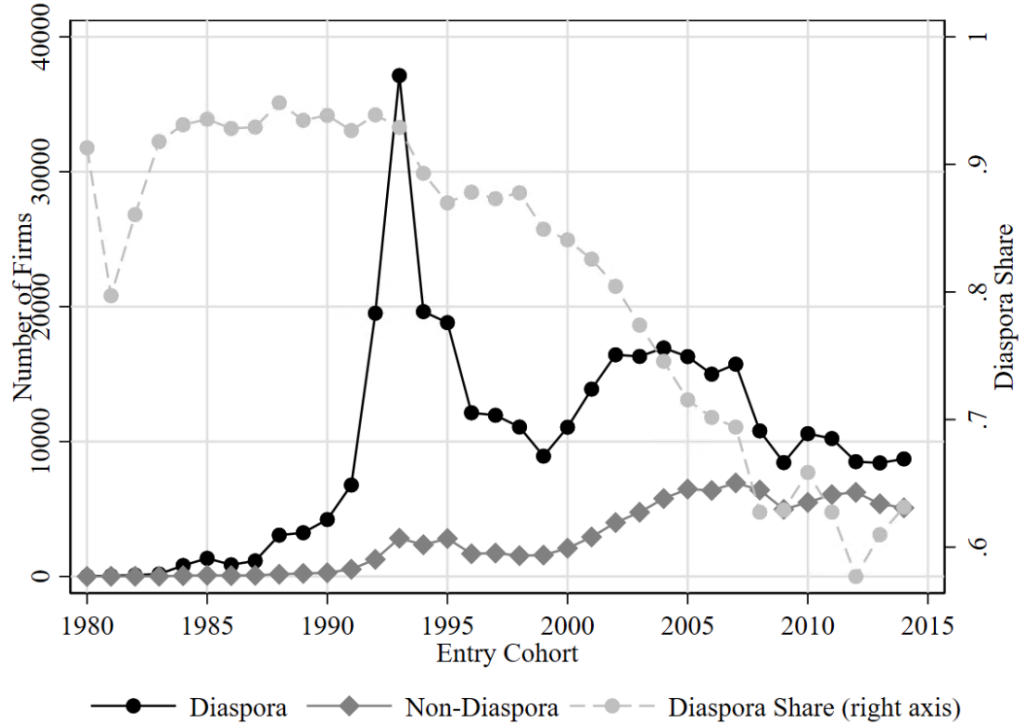
The impacts of the staggered rolling out of opening policies are vividly illustrated by Figure 1, which plots the entry and survival of foreign companies in mainland China by entry cohort using firm registration data from SAIC database. We can see that the number (and the survival-adjusted number) of foreign entrants rose steadily from 1980 to 1991 along with the opening of SEZs and *Open Coastal Cities*. And then in 1992, Deng Xiaoping’s southern tour resolved the uncertainties that once intimidated the foreign investors, and sparked the enthusiasm for entry into Chinese market. This is reflected in the dramatic spike in the number of foreign entrants in 1992 and 1993 as shown in Figure 1. Another historic event is China’s formal accession to WTO in the early 2000s, which has been

widely documented and believed to finalize the pro-market institutions of China toward foreign capitals (Branstetter and Lardy, 2006). The re-surge of foreign entrants during 2000-2005 in Figure 1 echoes this major historical event in China's great economic reform.



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

We noticed that the entry of foreign firms was already very active even before China joined WTO. By 1999, China had already become one of the most popular destinations for FDI flows, 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 opening policies cannot fully explain this fact as those policies are commonly practiced by other developing countries who however often failed to attract foreign investments. Some other complementary factors must have been in play.



**Figure 2: Foreign Entrants Decomposition: Diaspora v.s. Non-Diaspora**

Peering into the early stage of China's development before joining WTO, we claim that China's widespread networks with overseas Chinese must have been critical for the entry of foreign firms. To provide a descriptive evidence in support of our proposition, Figure 2 decomposes foreign entrants into diaspora ones and Non-diaspora ones. From 1980 to 2014, the share of diaspora entrants remained higher than 60% but declined gradually from above 90% in 1980s to about 60% in 2010s. These two observations together reveal: first, the dominant role of diaspora investments among all foreign investments throughout China's economic development process; and second, diaspora investment's relative importance during the early periods of opening-up. In other words, lineage connection could be the complementary factor we seek to understand why opening reforms were effective in China even before the market environment matured.

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 producing diasporas<sup>5</sup>. Among the vast number of overseas Chinese, many of them have become successful entrepreneurs in the host countries. According to the estimate by *The Economist* in 2019, more than three quarters of the wealth owned by the South East Asian billionaire are controlled by Chinese diasporas<sup>6</sup>. 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 cut off. The overseas Chinese, though physically abroad, are emotionally attached to their ancestral hometowns. Many anecdotal facts demonstrate that overseas Chinese remained socially in touch with their lineage group members, through sending mails and remittance to their hometowns (Tan, 2006). These pro-social behaviors, as Greif and Tabellini (2017) illustrated, reflect some persistent and fundamental values that are rooted in the Chinese culture. For centuries, Chinese people from the same lineage group exchange information, smooth contract enforcement, and weather the crises by helping each other out.

Therefore, when China reopened its door in 1978 and strove to attract foreign investments, the pervasive networks with overseas Chinese across China became a kind of social "endowments"<sup>7</sup>. Local governments in the opened regions realized the complementarity between the opening reforms and the local lineage connection with overseas Chinese. They hence enacted policies and coordinated through specialized organizations to encourage diaspora entrepreneurs 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 inter-

---

<sup>5</sup>[www.sussex.ac.uk/Units/SCMR/drc/about/index.html](http://www.sussex.ac.uk/Units/SCMR/drc/about/index.html)

<sup>6</sup>"Chinese Diaspora Inc: High-Wire Act", *The Economist*, May 30th, 2020

<sup>7</sup>One may be concerned that over time, the emotional attachment of Chinese diasporas, especially younger and foreign-born ones, with their ancestral homeland might wane, and the strength of lineage connection might be weakened. However, according to the interviews conducted by Tan (2006), many China-born diasporas indeed exhibited strong emotional attachment for their hometowns because they cared for the close relatives that still lived there; and second, many China-born parents would bring along their foreign-born children to ancestral hometowns to cultivate their self-identification as a member of the lineage group; last, many second or third generations of diaspora would take over their parents' business relationship even if they no longer consider themselves Chinese.

national trade. However, these transitional policies can mobilize the local lineage network which remains active even during China’s thirty years’ disconnection from the world market. With the help of the international ethnic ties, China was able to overcome common barriers that deter foreign investments in other under-developed countries and raise its share in international trade and FDI inflows substantially even before its accession to WTO. Furthermore, these pioneering foreign entrants, induced by lineage connection with their ancestral hometowns, are believed to have planted the seed investments for China’s later industrial growth and brought about a myriad of economic benefits in the long run.

### **3 Data Description**

This section describes the data used for empirical examinations. First, we introduce the data sources. Second, we lay out the procedures and discuss the methodology to identify diaspora firms as foreign firms that are controlled by overseas Chinese from firm registration data. Then we propose our measure of lineage connection. Last, we combine datasets aggregated from multiple sources to obtain our sample of analysis and provide summary statistics.

#### **3.1 Data Sources**

The main dataset we use is firm registration data maintained by SAIC which 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 type and a complete list of immediate shareholders and registered personnel including usually board members and senior executives.

The second dataset we use is the 2005 population Census which reports individual surnames of a 0.2% representative sample, from which we recover the geographic distri-

butions 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.

We also get the capacity 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 sample 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 Diaspora Firm**

We aim at understanding the role played by surname-based lineage connection in facilitating the entry of diaspora firms. In this section, we describe how we take steps to get the universe of diaspora firms in China .

By our definition, diaspora firms have to be foreign firms in the first place. To construct the universe of all foreign firms in China from firm registration database, we need to have a consistent definition of foreign firms. Two approaches coexist in literature to define foreign firms. One approach uses the administrative ownership code 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 a 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 have direct or indirect control over the firm. 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 irrelevant complexity and take the conservative approach by using administrative ownership 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 administrative 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 in China at hand, we proceed to identify diaspora firm. To achieve this goal, the most natural choice is to make use of the shareholder information and define a diaspora firm as a firm with a overseas Chinese 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 is highly likely to be the one with controlling right over the firm (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.1 to extract the surnames of each foreign firm’s registered personnel and identify whether they are ethnic Chinese<sup>8</sup>. If we observe one personnel as ethnic Chinese and the personnel also holds a non-mainland Chinese ID<sup>9</sup>, he or she is defined as overseas Chinese. And di-

---

<sup>8</sup>One may speculate that the names of some Chinese diasporas may be assimilated. For example, in Thailand and Indonesia, Chinese diasporas often felt the need to downplay their Chinese identity and switched to a foreign name in local language during the Cold War. But Tan (2006) suggests that Chinese diasporas tend to use their original Chinese names when dealing with Chinese business so as to make best use of their flexible cultural identity and multiple belongings.

<sup>9</sup>Non-mainland ID can be either passport of a foreign country or travel permit for Hong Kong, Macau, Taiwan (HMT) residents.

aspora firms are then identified as foreign firms whose legal representatives are overseas Chinese<sup>10</sup>.

Our sample for empirical analysis consists of all diaspora 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 Hong Kong) and then imports back into mainland China as a new legal entity. Since the founding of such foreign firms was most likely motivated by 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 diaspora firms in China and potentially bias our estimates in empirical analysis<sup>11</sup>. Moreover, our main dataset also excludes non-diaspora foreign firms since we want to keep our sample as comparable as possible<sup>12</sup>. Last, since the legal representative is not necessarily the major shareholder of the firm, it is possible that a company classified as diaspora firm is actually not owned by overseas Chinese. However, the choice of appointing overseas Chinese legal representative still reveals the motive of the foreign firm to take advantage of lineage connection and missing this part of the observations would actually make the effects of lineage connection

---

<sup>10</sup>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 registered personnel to represent the firm. Firstly, we show in Appendix [A.2](#), 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 chain to identify each foreign firm’s ultimate owners who would be 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 [A.2](#) 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.

<sup>11</sup>We cannot, however, rule out the possibility that foreign citizens delegate the firm control rights to citizens in mainland China. But delegating a firm to a local person does not undermine the value of lineage connection. Moreover, we show that our main findings are robust to including foreign firms that are legally represented by citizens in mainland China. See Section [4.4](#).

<sup>12</sup>Including non-Huaren foreign firms would mechanically increase the size of the control group with no lineage connection. Yet, our findings are also robust to including non-Huaren foreign firms.



on facilitating the entry of foreign firms underestimated<sup>13</sup>.

### 3.3 Measure Lineage Connection

For each diaspora firm, we also identify the legal representative’s surname, which can be used to infer the lineage connection endowed with the firm. Then we can leverage the geographic distributions of surnames from 2005 population Census to measure the strength of lineage connections between diaspora firms of different surnames and different locations. Specifically, we define the lineage connection  $m_{sp}$  between a surname  $s$  and prefecture  $p$ <sup>14</sup> as following:

$$m_{sp} = \frac{E_{sp}}{\sum_p E_{sp}} \quad (1)$$

where  $E_{sp}$  denotes the size of population with surname  $s$  in prefecture  $p$ . This measure has the natural interpretation that it indicates the probability one overseas Chinese with surname  $s$  can trace her origin to 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<sup>15</sup>. 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

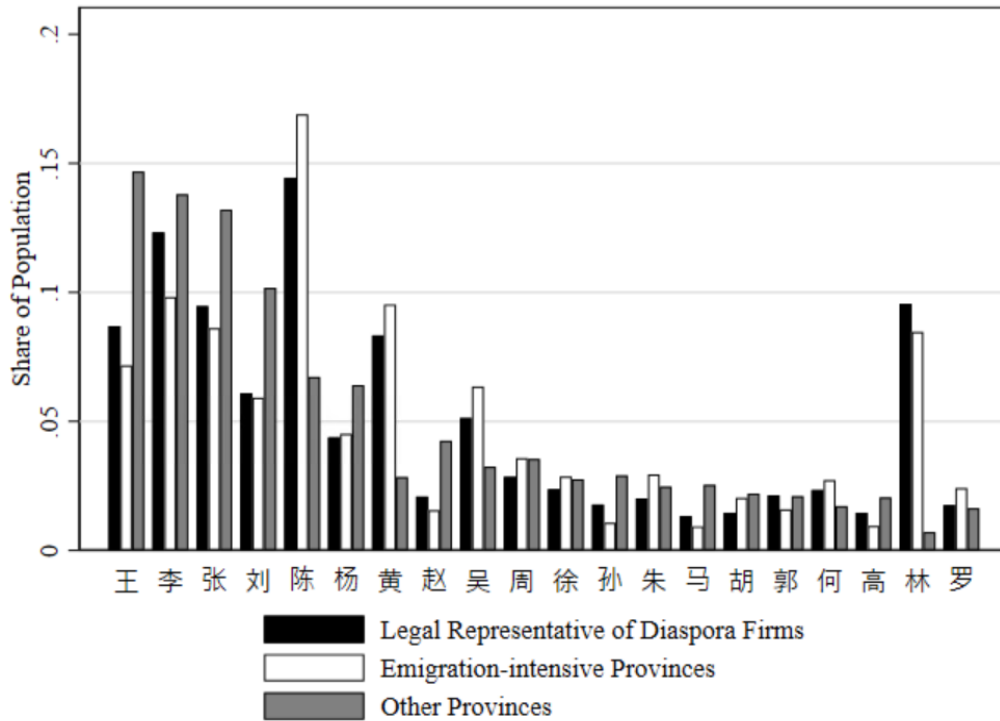
---

<sup>13</sup>There could be interesting differences between diaspora-owned foreign firms and diaspora-managed foreign firms in terms of incentives. This could be potentially fruitful avenue of research in the future.

<sup>14</sup>There are three reasons that we choose prefecture as the geographic unit for our empirical analysis. First, measurement noises in lineage connection for rare surnames are large at small counties. Second, administrative boundaries of counties may change over time in China, while those of prefectures remain fairly stable. Third, we have more variations in our measure of lineage connection across prefectures than across provinces.

<sup>15</sup>If we instead use  $E_{sp}$  as measure of lineage connection, we expect spurious correlation between lineage connection and the entry of diaspora firms even without the functioning of lineage connection, because populous surnames like “陈” (“Chen”, “Chan”, “Tan”) can be dominant both domestically and abroad. So Huaran firms of populous surnames could mechanically have more presence everywhere. Moreover, Appendix A.3 indicates that our lineage connection is actually mildly negatively correlated with prefecture population at surname-prefecture level.

to proxy lineage connections. The originality is twofold. First, while there are relatively rich data on immigration into developed countries such as the United States (Burchardi, Chaney and Hassan, 2019; Sequeira, Nunn and Qian, 2020), so far the data on emigration, especially at sub-national level from developing countries like China, are extremely rare to our knowledge. Our surname-based measure makes best use of the fact that lineage groups in China are usually operating within surnames (Clark, 2015). Second, 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 Populations**

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 diaspora 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 diaspora firms are closer to that over the population of emigration-intensive provinces than the population of other provinces. For example, “陈” (“Chen”, “Chan”, “Tan”)<sup>16</sup> is the most common surname both among overseas Chinese legal representatives 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<sup>17</sup>. 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

---

<sup>16</sup>Due to Chinese-English translation, there could be more than one English spellings for the same Chinese surname. Figure [A.1](#) illustrates the complexity of spelling-character mapping between the two languages, using “陈” —the most common overseas Chinese surnames written in Chinese characters, and “Tan”—the most common English spelling for “陈”, as an example. In this study, we aggregate the overseas Chinese surnames to the Chinese character level with probabilistic assignments in cases when we cannot uniquely identify the Chinese character from the English spelling. See Appendix [A.1](#) for details.

<sup>17</sup>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](#)).

market economy. This Reform Era also saw that diaspora firms represent a dominant share of all foreign firms (see Figure 2). Utilizing the entry and exit information in the SAIC, we construct two measures of diaspora firm entry at surname-prefecture-year level: number of new entrants and survival-adjusted number of entrants (defined as those lasting for more than 4 years). 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 diaspora firms is mostly accounted by the extensive margin. Only 5.2% of surname-prefecture-year triples have diaspora firm entry while the magnitude is even smaller (0.9%) for diaspora 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 diaspora firm entrants tend to be headquarter firms rather than branch firms, labor-intensive rather than capital-intensive and followers rather than leaders in new industries<sup>18</sup>. These characteristics are largely in line with the anecdotal descriptions on diaspora firms (Huang, 2003). Besides, only 57.8% of diaspora firms during this period originated from HMT, suggesting that HMT entrants may fail to serve as a precise proxy for diaspora firm entrants as in the literature<sup>19</sup>.

---

<sup>18</sup>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 its 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.

<sup>19</sup>Appendix A.5 discusses the strengths of our dataset relative to those used by prior literature and cross-validates our dataset with other official 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-Diaspora 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: Diaspora 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-diaspora FDI which we sum up registered capital in 2014 of non-diaspora foreign firm using SAIC database.

## 4 Lineage Connection and the Diaspora Firm Entry

### 4.1 Identification Strategy

As is obvious, unobserved factors at surname level, prefecture level or even surname-prefecture level may simultaneously affect lineage connection and the entry of connected diaspora firms, creating a spurious correlation. For example, in southeastern China, where most overseas Chinese 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. Therefore to identify the causal effect of lineage connection on the entry of diaspora foreign firms, we employ the following baseline specification:

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

where  $s$  denotes surname,  $p$  denotes prefecture, and  $c$  denotes entry cohort.  $Y_{spc}$  represents the outcome variables. We are primarily interested in how lineage connection affects the number of diaspora foreign entrants. Nevertheless, entrants were heterogeneous in their latent quality and the welfare implication of firm entry might be undermined if firms of low quality selectively entered. Therefore, we supplement another measure to mitigate this concern: the survival-adjusted number of entrants (defined as number of entrants that lasted for more than 4 years following Kerr and Nanda (2009))<sup>20</sup>.

Under all specifications, we control for surname-prefecture fixed effects  $\eta_{sp}$ , surname-cohort fixed effects  $\theta_{sc}$  and prefecture-cohort fixed effects  $\delta_{pc}$ . The broad set of fixed effects help us guard against a wide range of confounding factors including geographic advan-

---

<sup>20</sup>Robustness checks using other thresholds are shown in Figure A.11 in the Appendix.

tages, place-based policies and surname-specific expertise. It is noticeable that since all possible cohort fixed effects have been controlled, all differential probabilities to exit by entry cohorts have been accounted for. 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 and time.  $Open_{pc}$  equals 1 if the prefecture  $p$  had been open since year  $c$ , and it equals 0 otherwise<sup>21</sup>.  $m_{sp}$  is the measure of lineage connection between surname  $s$  and prefecture  $p$  and is defined by 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<sup>22</sup>. The error term  $\epsilon_{spc}$  captures all the idiosyncratic disturbances. Standard errors are clustered at surname-prefecture level.

The coefficient of interest is  $\beta$  which captures the heterogeneous effects of lineage connection across surname-prefecture pairs 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 surname-prefecture pairs. We are, in effect, performing the following three-fold comparisons: (1) surnames within prefecture before and after the opening of prefecture to get trends of surnames; (2) differential surname trends across opened and closed prefectures to get treatment effects of opening shock (double DID); (3) further differences across surname-prefecture pairs with different strengths of lineage connections to get our 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 the number of

---

<sup>21</sup>See Table 1 for a full depiction of the stepwise opening of Chinese prefectures to foreign capital during the Reform Era. As we excluded Shanghai from our data, the opening of Pudong district in 1990 is not captured in our analysis.

<sup>22</sup>See Section 4.4.

entrants for surname-prefecture pairs with different lineage connections before the opening shocks. We argue that this identification assumption is highly likely to hold for three reasons. First, our measure of lineage connection has been shown to be predetermined (Bai, 2020). Secondly, the timing of the opening of prefectures during the Reform Era comes largely as shocks since foreign capital inflows were almost prohibited before then. Third, a vast set of fixed effects have been controlled to eliminate various factors that could contaminate our causal estimates.

For providing further evidences to corroborate the validity of our identification strategy, we employ an event-study framework as following:

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

where  $\tau = c - c_p$  references the time window relative to the opening shock of the prefecture  $p$ . By employing this specification, we can examine both the dynamic effects of lineage connection four years before and after the local opening shocks.  $Open_{pc}^{\tau}$  equals 1 if year  $c$  is  $\tau$  years after the opening of prefecture  $p$  and 0 otherwise. The omitted benchmark group is  $\tau = -1$ . Hence all estimates of  $\beta^{\tau}$  should be interpreted as relative to one year prior to the opening shock. For our identification assumption not to be violated, we expect  $\beta^{\tau}$  for all  $\tau < 0$  to be statistically insignificant.

To shed light on the underlying mechanisms, we follow Kerr and Nanda (2009) and further disaggregate diaspora entrant counts at prefecture-surname-cohort-type level where type indicates a 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



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

where  $type_x$  is a dummy variable for firm type  $x$ . The sign of the estimate of  $\beta_2$  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 lineage connection affects the entry of diaspora firms following the prefecture-level opening shocks. 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 0.01 higher lineage connection  $m_{sp}$  can get induce 0.01767 more diaspora entrants in one year than other surnames following the opening shocks of prefecture  $p$ . And this effect is economically large since 0.01767 is more than one quarter of the mean number of diaspora entrants across all surname-prefecture-cohort triplets. Column (2) confirms that not all entrants are short-lived by estimating a slightly smaller yet still significant effect of lineage connection with number of survival-adjusted entrants as outcome. Columns (1) to (2) altogether suggest that lineage connection facilitates the entry of diaspora foreign firms even if we consider the heterogeneous quality of the entrants by survival.

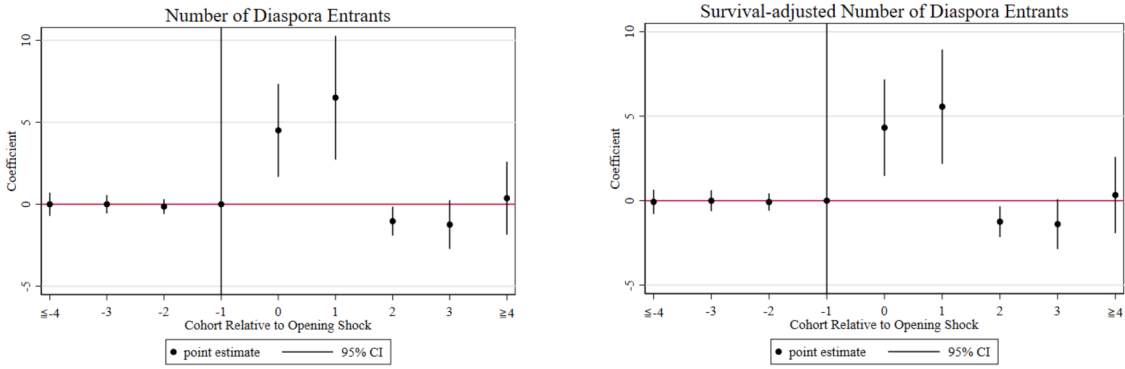
As a crucial test for the identification assumption, we also estimate an event study specification in Equation 3 and show in Figure 4 that the entry and survival of diaspora foreign firms across surname-prefecture pairs had similar trends prior to the prefecture-level opening shocks. This results supports our identification strategy and causal interpretation

**Table 3:** Lineage Connection and Entry and Survival of Diaspora Foreign Firms

	Number of Diaspora Entrants	Survival-adjusted Number of Diaspora Entrants
	(1)	(2)
Mean of Dep. Var.	0.060	0.052
Open $\times$ Lineage Connection	1.767*** (0.574)	1.517*** (0.527)
Adj. $R^2$	0.570	0.555
N	1,344,421	1,344,421
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. Survival-adjusted Number of Diaspora Entrants is the number of entrants that survive for more than 4 years (included).

of the results. The time-varying effects shown in Figure 4 also suggest that lineage connection facilitates the entry of diaspora foreign firms only in the short run.

**Figure 4:** Event Study

There are concerns on the validity of our use of the staggered opening of prefectures in China as shocks. Since most of the opening happened in 1992, the opening of the small set of coastal cities before 1988 (Shanghai excluded in the sample) might not be effective at all. If 1992 is the true shock we should use, then what we observe in Figure 4 as parallel

trends and short-run effects might in fact be pre-trends and no effects afterwards if we move the treated timing to universally 1992. For addressing this concern and providing validity to our use of the staggered opening shocks instead of the 1992 static shock, we run the same event studies as in Figure 4 and show the results in Figure A.10 in the Appendix. Immediately we notice that there are no pre-trends 4 years before the 1992 shock. And the last shock before 1992 we use in our main analysis is exactly the 1988 opening of Hainan province. This result should not be coincidence and informs us that the staggered opening of process before the 1992 full opening of China should not be ignored for estimating the effects of opening shocks.

### 4.3 Selection on Firm Quality

If only firms of high quality can take advantage of the lineage connection, the sorting of more capable diaspora firms into surname-prefecture pair with high lineage connection could possibly be conducive to both entry and survival of diaspora firms, leading to an overestimation of the effects of lineage connection. To rule out this plausible explanation, Table 4 further looks at the effect of lineage connection on survival ratio and average registered capital of diaspora entrants in 2014 conditional on entry, which are used to proxy the inherent firm quality reflected in later survival and firm growth. Although we find effects of lineage connection on the number of survived firms in Table 3, we did not discover any benefit of lineage connection on survival probability and registered capital in 2014 of diaspora firms conditional on their entry.

This result not only mitigates the concern of overestimating the effect of lineage connection on the entry of foreign firms but also, together with the dynamic effects identified in Figure 4, suggests that the mechanism through which lineage takes effect is mainly by lowering entry barriers rather than contributing to survival or long-term growth for firms. At this stage, we interpret the role played by lineage connection as lowering entry barriers

**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 $\times$ Lineage Connection	-0.333 (0.825)	-7040.914 (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.

in the short run regardless of diaspora firm's latent quality<sup>23</sup>. And selection is not a major threat to our identification strategy.

## 4.4 Robustness Checks

There still remains many concerns on our findings. In this section, we would like, by taking full advantage of our unique dataset, to test our findings with a wealth of robustness checks.

**Alternative Dependent Variables.** We notice that the entry and survival of diaspora 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 exten-

<sup>23</sup>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 the lowered 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 firm 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 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.

sive margin at surname-prefecture-cohort level. We then replace our firm count measures with dummies indicating whether there is at least one entrant 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 diaspora 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.

**Table 5:** Robustness Checks I: Alternative Dependent Variables

	At Least One Diaspora Entrant	At Least One Survived Diaspora Entrant in 4 Years	At Least One Survived Diaspora Entrant in 2014	ArcSinh (Survived Diaspora Registered Capital in 2014)
	(1)	(2)	(3)	(4)
Open $\times$ Lineage Connection	0.766*** (0.112)	0.672*** (0.102)	0.165*** (0.053)	1.229*** (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.

**Alternative Lineage Connection Measure.** The second potential challenge to our results is the measurement errors in our lineage connection. 2005 population Census is the only data source where we could observe nationally representative distributions of surnames across prefectures. Since our sample period starts from 1981, it is possible that the internal migration between 1981 and 2005 could lead to changes in the spatial distribution of surnames and diaspora entrants at the same time. However, as we argued before, past literature have shown that the spatial distribution of surnames remains fairly stable throughout the history even after warfare and revolutions (Clark, 2015) and recent work by Bai (2020) confirms 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 provide additional checks on our results by 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.

**Table 6:** Robustness Checks II: Alternative Setups

	Open $\times$ Lineage Connection		
	Alternative Lineage Connection Measure	Alternative Standard Error	Include Roundtrip Diaspora Entrants
	(1)	(2)	(3)
Number of Diaspora Entrants	2.874*** (0.946)	1.767*** (0.493)	3.821*** (0.820)
Survival-Adjusted Number of Diaspora Entrants	2.511*** (0.903)	1.517*** (0.456)	3.253*** (0.743)
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 except in column (2). Survival-adjusted Number of Diaspora Entrants is the number of entrants that survive for more than 4 years (included). Potential roundtrip diaspora entrants are defined as entered foreign firms represented by citizens of People's republic of China.

The results on firm entry using this alternative lineage measure are presented in column (1) of Table 6. We still get significant estimates for our 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 ones 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 diaspora foreign firms with and without adjustment by survival.

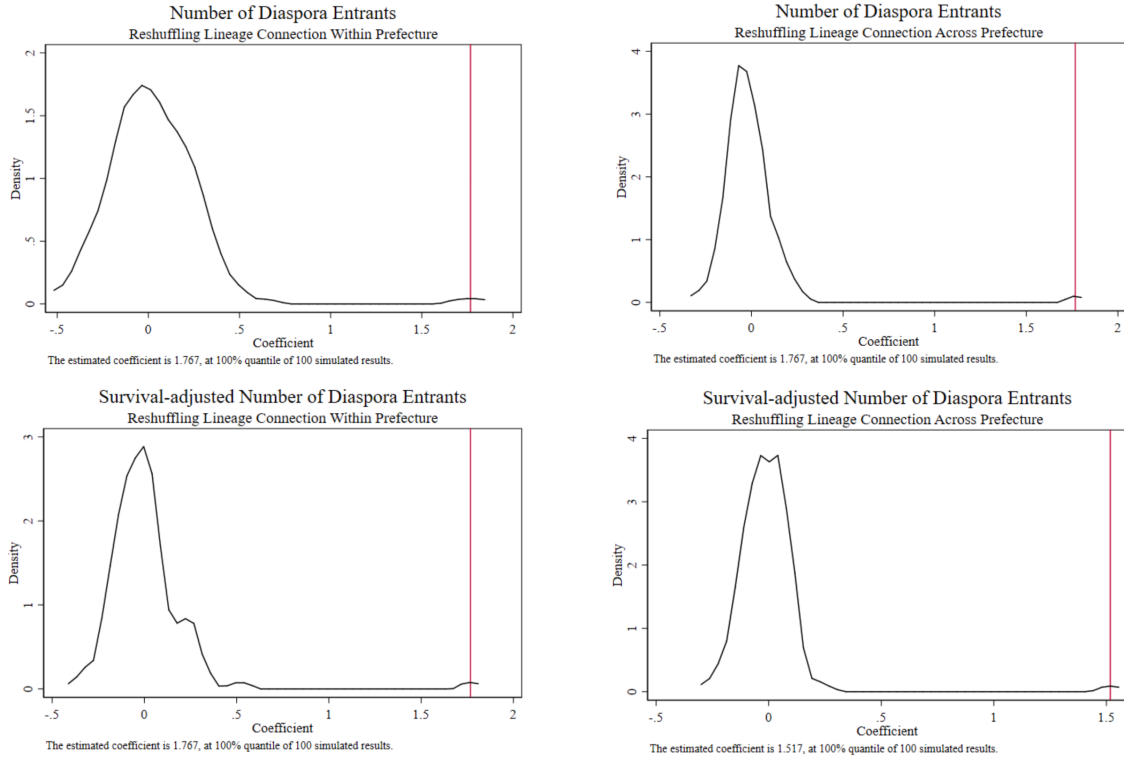
**Alternative Standard Error.** Our baseline results come from regressions which cluster standard errors at surname-prefecture level to account for serial heteroskedasticity

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 by much even if we cluster our results at prefecture level.

**Including Roundtrip Diaspora Entrants.** To construct our baseline sample, we adopt a conservative definition for diaspora foreign firm, that is, a firm represented by an ethnic Chinese with foreign passports or travel permits for HMT residents. In this way, we hope to eliminate the potential influence of the prevalent "roundtrip" foreign firms who were virtually domestic firms motivated by tax benefits during the Reform Era on our results. To show robustness, we include back those foreign firms whose representatives hold an ID card of People's Republic of China into our sample and show in column (3) of Table 6 that our estimates expand a lot. This, to some degree, supports our baseline sample construction since these potentially roundtrip foreign firms would dramatically exaggerate the presence of diaspora foreign firms and therefore drive up the estimated effects of lineage connection on the entry of foreign firms.

**Permutation Tests.** To check whether our results are influenced by any other errors in our lineage connection measure, 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. The right panel plots distributions of estimated coefficients from 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 the entry of diaspora foreign firms are not driven by unobserved measurement noises in lineage connection.

**Subsample Regressions.** Another challenge to the interpretation of our results is



**Figure 5: Permutation Tests**

**Table 7: Robustness Checks III: Subsample Regressions**

	Open $\times$ Lineage Connection			
	Excluding Emigration- intensive Provinces (1)	Excluding Emigration- intensive Surnames (2)	Within Emigration- intensive Provinces (3)	Excluding FDI- intensive Prefectures (4)
Number of Diaspora Entrants	0.642*** (0.233)	1.733*** (0.332)	6.001*** (2.476)	1.534*** (0.440)
Survival-Adjusted Number of Diaspora Entrants	0.482*** (0.167)	1.480*** (0.292)	5.152*** (2.186)	1.308*** (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 Diaspora 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 overseas Chinese legal representatives from 1981 to 2014. FDI-intensive prefectures are those whose cumulative number of hosted foreign firms rank top 20 among all prefectures during 1981 to 2014.



the possibility that variations from certain regions or surnames could largely 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 regions. And some surnames, whose population historically resided in emigration-intensive provinces, have persistently higher presence in both FDI-intensive regions and among diaspora 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 diaspora entrants and survival-adjusted number of diaspora entrants remain positive and statistically significant in columns of Table 7 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

We have documented, for the first time in literature, a large causal effect of surname-based lineage connection on foreign investments in terms of firm entry. We now turn to explore the underlying mechanisms through which the lineage connection exerts the observed effects on the entry of diaspora foreign firms. A placebo test is first used to show that lineage connection helps lower the entry barriers caused by China's immature market environment which had been eliminated later for WTO accession. Then we further identify what roles lineage connection played in facilitating the entry of diaspora firms in such

immature market environment.

In general, social ties can work plausibly through demand side and supply side to benefit movement of economic factors between the two nodes of the tie. On the demand side, as argued by [Burchardi, Chaney and Hassan \(2019\)](#), lineage connection could facilitate the entry of diaspora foreign firms as tastes could persist over generations and diaspora entrepreneurs might just work abroad for 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 rationalizing our results. On the supply side, as summarized by [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 diaspora entrepreneurs. We discover evidence in support for both of them in the following empirical analysis testing auxiliary predictions of the two distinct mechanisms.

## 5.1 WTO Accession Placebo Test

One of the main goals of this paper is to demonstrate the importance of international lineage networks in facilitating the entry of foreign firms in face of an immature market environment. A natural placebo test is see how the effects of lineage connection would change when the institutional environment has been formalized. China's accession to WTO in 2002 is the most ideal event for such a test, since China had spent decades to improve its market institutions in order to be accepted as a member of WTO. We show that indeed the heterogeneous effects of opening shock across surname-prefecture pairs with different lineage connections we found during the early Reform Era when market institutions have not been established do not appear during the later era when market environment has been built to meet the requirements of WTO if we use China's WTO accession as opening

shock.

**Table 8:** Placebo Test: WTO Accession

	Number of Diaspora Entrants	Survival-Adjusted Number of Diaspora Entrants
	(1)	(2)
WTO Accession $\times$ Lineage Connection	0.224 (0.240)	0.300 (0.217)
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. Survival-adjusted Number of Entrants is the number of entrants that survive for more than 4 years (included). The sample periods are from 1997 to 2009. Variable *WTO accession* takes value 1 for the period from 2002 onward, and 0 otherwise.

In Table 8, we run our baseline triple DID with WTO accession of China in 2002 as shock and interact the WTO accession shock with our surname-prefecture varied lineage connection measure. This regression gives us the heterogeneous effects of the WTO accession shock on (survival-adjusted) diaspora firm entry. As expected from our conceptual framework, both column (1) and column (2) of Table 8 show that there are no significant effects of lineage connection on number of diaspora entrants and number of survival-adjusted number of diaspora entrants between the surname-prefecture pairs with different lineage connection, before and after the 2002 WTO accession.

## 5.2 Sharing Information

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. Diaspora 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 lin-

age members through the social tie to the diaspora entrepreneurs abroad. Hence lineage connection could mitigate the information frictions and creates an information advantage for diaspora 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 across places with different predetermined conditions.

First, if diaspora 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 diaspora entrepreneurs. In another word, we conjecture that the effect of lineage connection on diaspora firm entry would be higher in places of lower information transmission cost if lineage connection increases the supply of information to diaspora entrepreneurs. To empirically test this auxiliary prediction of information channel, we examine the heterogeneous effects of lineage connection on the entry of diaspora foreign firms across province-years with differential predetermined telephone exchange capacity. Back in the 1980s and 1990s, the telephone exchange capacity determined whether diaspora entrepreneurs could communicate easily with local people they were connected to. To formally test, 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 9 show our results. The coefficients of the new interaction terms remain positive and significant for the number of diaspora entrants and survival-adjusted number of diaspora entrants. These findings are in line with the expectation that communication infrastructure are complementary with lineage connection since overseas Chinese can use the communication infrastructure to access the more information supplied by diaspora entrepreneurs' local contacts following the local opening reform.

Second, 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 frag-

**Table 9:** Sharing Information

	Number of Diaspora Entrants	Survival-adjusted Number of Diaspora Entrants
	(1)	(2)
<b>Panel A: Communication Infrastructure</b>		
Open $\times$ Lineage Connection	-2.507 (1.908)	-1.814 (1.846)
Open $\times$ Lineage Connection $\times$ Telephone Exchange Capacity	1.074*** (0.409)	0.868** (0.386)
Adj. $R^2$	0.454	0.437
N	666,397	666,397
<b>Panel B: Social Fragmentation</b>		
Open $\times$ Lineage Connection	-9.356*** (2.936)	-7.201*** (2.899)
Open $\times$ Lineage Connection $\times$ Surname Fragmentation	11.963*** (3.336)	9.365*** (2.899)
Adj. $R^2$	0.409	0.395
N	1,186,664	1,186,664
<b>Panel C: Pioneer Firm</b>		
Open $\times$ Lineage Connection	7.424** (3.063)	6.390** (2.712)
Open $\times$ Lineage Connection $\times$ Pioneer Firm	-13.080** (5.650)	-11.263** (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 the 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.

mented. Then for less connected diaspora entrepreneurs, searching for local information would be harder in such places. Therefore lineage connection would have higher effects on diaspora entrants 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, in Panel B of Table 9 we discover stronger effects of lineage connection in prefectures with higher social fragmentation as measured by one minus the Herfindahl-Hirschman Index (HHI) of surname concentration. This result corroborates the information channel for explaining the large causal effects of lineage connection we identified on diaspora foreign firm entry.

Last, for entrepreneurs, local information would be more abundant and informative if there had already been entrants in the same industry at that location. In another word, it would be harder for a pioneering firm, defined as the the first-ever foreign entrant in a 2-digit industry in a prefecture, to access useful market information through their lineage connection. To test this prediction, we further decompose 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 pioneering firm dummy. Panel C of Table 9 demonstrates that diaspora firms of stronger lineage connection tend to enter into existing industries as imitators rather than new industries as pioneers following the local opening reform. This suggests that information on proven business opportunities could pass through lineage group. Since we focus on a period when most industries in China just began to take root, i.e. the Reform Era, 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). It should be noticed that, the effect of lineage connection on pioneering firms is actually negative. That's to say higher lineage connection even redistribute diaspora entrepreneurs from being pioneering to imitating. However, this does not mean lineage connection does not help motivate the entry of pioneers. Most of the

pioneering foreign firms are diaspora firms as seen in Figure 2. And as pioneers induce imitators, these early pioneering diaspora entrants could have long-term benefits on both non-diaspora foreign and domestic firm entry and investments, which will be discussed in later section.

### 5.3 Enforcing Contract

Weak contract enforcement has also been seen as key barrier to deter foreign investments from flowing into developing countries. And in the context of China, informal institution such as lineage groups, as argued by Greif (1993) and Greif and Tabellini (2017), could be useful to strengthen contract enforcement under circumstances with weak or no enforcement institutions. We conjecture that the effect of lineage connection would be higher when demand for contract enforcement is higher. In the following analysis, we present how we test this auxiliary prediction from three aspects and show the results which all support the contracting channel in Table 10.

When a foreign entrepreneur consider entering China, he or she has to decide whether to set up the firm as a headquarter firm or a branch firm with headquarter registered outside China. The former choice requires stronger commitment than the latter since the relative autonomy of a headquarter firms is associated with higher administrative and monitoring costs, which demands better contracting environment. Therefore we conjecture that lineage connection would have larger effect on the entry of diaspora foreign firms as headquarter firms. Panel A of Table 10 finds that diaspora 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 and suggests that lineage connection indeed helps 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 a foreign entrepreneur

**Table 10:** Enforcing Contract

	Number of Diaspora Entrants	Survival-adjusted Number of Diaspora Entrants
	(1)	(2)
<b>Panel A: Branch Firm</b>		
Open $\times$ Lineage Connection	8.795*** (3.363)	7.633** (2.997)
Open $\times$ Lineage Connection $\times$ Branch Firm	-15.823*** (6.255)	-13.749*** (5.547)
Adj. $R^2$	0.452	0.444
N	2,690,048	2,690,048
<b>Panel B: Labor Intensive</b>		
Open $\times$ Lineage Connection	-1.864* (1.043)	-1.650* (0.928)
Open $\times$ Lineage Connection $\times$ Labor Intensive	4.990** (2.466)	4.433** (2.226)
Adj. $R^2$	0.665	0.645
N	2,690,048	2,690,048
<b>Panel C: Supplier Intensive</b>		
Open $\times$ Lineage Connection	-2.331* (1.156)	-2.030** (1.010)
Open $\times$ Lineage Connection $\times$ Supplier Intensive	6.196** (2.736)	5.397** (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 the 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 \(2007\)](#).



needs to hire a large group of local employees or sign contracts with many local suppliers, the demand for better contract enforcement would be very high. We hypothesize that lineage connection would be more valuable for diaspora entrants in industries with either higher labor intensity or intermediate good intensity. To formally test the hypothesis, we further breakdown entrants into labor-intensive industries and non-labor-intensive industries (for the subsample of manufacturing industries only), supplier-intensive industries and non-supplier-intensive industries. Results in Panel B and Panel C indicate that diaspora foreign firms with stronger lineage connection tend to enter into labor-intensive and supplier-intensive industries. These results suggest that lineage connection helps mediate access to labor and intermediate inputs through strengthened contract enforcement.

## 6 Seeding Effects in the Long Run

We have estimated the causal effects of lineage connection on the entry of diaspora foreign firms during the take-off stage of Chinese economy. However, the short-run surge of diaspora firm entry during the Reform Era may or may not have meaningful long-term impacts on Chinese economy. This section, taking one step further, asks whether those industrial “seeds” planted by the early DDI had long-lasting spillover effects on non-diaspora foreign firm entry and domestic private firm entry and if yes how large the seeding effects are. To achieve this goal, we proceed in two steps: first show the seeding effects with event studies and then estimate the seeding multipliers on non-diaspora foreign firm entry and domestic private firm entry separately with an innovative IV strategy.

As FDI literature have 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).

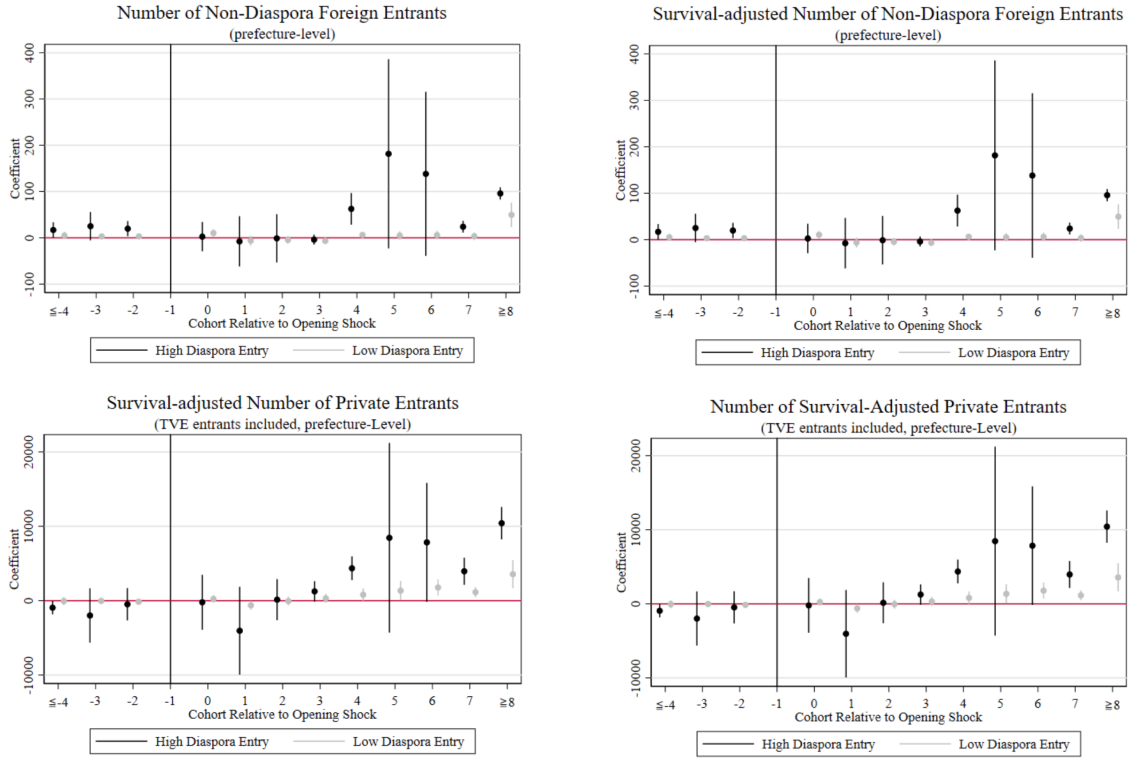
Compared with past literature on identifying FDI spillovers in China, our paper makes several significant contributions. First, we extend the research periods to the early 1980s when the spillover effects of early FDI in China (we call the “industrial seeds”) tend to be the strongest. In contrast, most existed studies only focus on late 1990s to mid 2000s due to data limitations. Second, our paper draws upon the most comprehensive dataset on FDI covering all sectors while past literature mainly use the above-scale manufacturers<sup>24</sup>. Last but not the least, our paper establishes the causal link between early DDI and future foreign investments in the long run. This fills the gap between anecdotal narratives and formal empirical examination on the importance of overseas Chinese in China’s rapid rise as a hub for multinationals and foreign firms during the process of China’s Reform and Opening-Up. Besides, this also proposes a new type of FDI spillovers from a perspective other than the traditional horizontal/vertical FDI spillovers across sectors.

## 6.1 Spillover Effects of DDI over Time

We try to identify the spillovers from diaspora direct investments on non-diaspora direct investments and examine the spillovers over time to test our seeding hypothesis: effects only take place gradually after years. For this purpose, we still leverage the opening shock during the reform era to conduct event studies. Our outcomes now are, however, non-diaspora direct investments including non-diaspora foreign firm entry and domestic private firm entry at prefecture level. That’s to say our event studies are now at prefecture-cohort level instead of prefecture-surname-cohort level. It is not surprising that there would be large effects of opening shock on prefecture-level firm entry. We further separate prefectures into two groups by their predetermined stock of diaspora firms before their opening: prefectures with higher-than-median diaspora firms versus prefectures with lower-than-median diaspora firms and conduct the event studies for both the two groups.

---

<sup>24</sup>See Appendix A.5 for a detailed discussion.



**Figure 6:** Event Studies for Prefectures with High v.s. Low DDI before Opening

The results are shown in Figure 6. We immediately see two patterns from the results: (1) the effects of the opening shock on non-diaspora foreign firm entry and domestic private firm entry, in contrast with the diaspora firm entry, only show up after 4-5 years; (2) these effects only exist in prefectures with higher-than-median stock of diaspora firms. Taken these two findings together with the previous result that the effects of opening shock on diaspora firm entry only happened immediately after the opening shock but not over time, we conclude the early entry of diaspora firms facilitated by lineage connection have seeding spillover effects on both non-diaspora foreign firm entry and domestic private firm entry. And these entrants, according to ?, account for most of China's GDP growth from 1978 to 2005.

## 6.2 Isolate Lineage-driven DDI

We have identified the seeding effects of DDI on other types of investments in the long run. In another word, DDI in early Refomr Era had multiplier effect if we think DDI as a policy lever. The implication of our findings therefore favors policies aim at attracting DDI. Then it is very important to know how large the multipliers are for evaluating whether the costs of attracting DDI outweigh the benefit DDI will bring to the country.

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 arose from the interaction of surname distribution and the gradual opening of Chinese prefectures during the Reform Era as instruments for the presence of DDI in 1996. 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 DDI inflows to the prefecture. So the exclusion restriction for using lineage-driven DDI as instrument for early DDI would be very likely to hold.

Our IV strategy has three stages. In the “zero stage”, we estimate our baseline regression in Equation 2 with number of diaspora entrants as outcome and get the point estimate of the coefficient  $\hat{\beta}$ . With the estimated coefficient, we are able to get the predicted value  $\hat{\beta} \times Opening_{pc} \times m_{sp}$ —the lineage-driven DDI at surname-prefecture-cohort level. We then aggregate them over all surnames and entry cohorts to get a predicted value at prefecture level:

$$DDI_p^l = \sum_s \sum_{c=1981}^{1996} \hat{\beta} \times Opening_{pc} \times m_{sp} \quad (5)$$

which represents the cumulative lineage-driven DDI for each prefecture in year 1996. We also have observed stock of diaspora entrants for each prefecture in year 1996. Then the first stage regression where we use lineage-driven DDI as the IV for observed DDI in 1996

can be estimated as following:

$$DDI_p = \alpha + \theta \times DDI_p^l + \lambda \times X_p + \epsilon_p \quad (6)$$

where  $DDI_p$  is the observed number of diaspora entrants in prefecture  $p$  in 1996.  $X_p$  denotes a set of controls at prefecture level. And with predicted prefecture-level number of diaspora entrants from this first-stage regression, we can perform the second stage of estimation as following:

$$\pi_p = \alpha + \gamma \times D\hat{D}I_p + \lambda \times X_p + \epsilon_p \quad (7)$$

where  $\pi_p$  is our long-term outcome of interest.

### 6.3 Estimate Long Run Seeding Multipliers

In this subsection, we estimate long-term effects of the lineage-driven DDI, which is the dominant form of inward FDI during the period when China's market economy was largely immature, on the future non-diaspora investments.

To identify the long run effects of lineage-driven DDI, we implement the methodology described in Section 6.2. First, 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 diaspora foreign firms survived in 1996 to construct lineage-driven DDI at the prefecture level  $D\hat{D}I_p$  and use it as dependent variable in the first stage (Equation 6). After conducting the first stage with lineage-driven predicted DDI from our baseline regression, we perform the second-stage regressions for a series of long-run measures of non-diaspora investments including non-diaspora foreign firm stocks, domestic private firm stocks and these survived firms' registered capital respectively 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 in 1978, distance to sea, social fragmentation index, SEZ dummy, *Open Coastal City* dummy and province fixed effects.

**Table 11:** Seeding Effects of DDI on Non-Diaspora Investments

	Non-Diaspora Foreign Firm Stocks in 2014	Domestic Private Firm Stocks in 2014	ArcSinh Registered Capital of Non-Diaspora Foreign Firms in 2014	ArcSinh Registered Capital of All Domestic Private Firms in 2014
	(1)	(2)	(3)	(4)
Panel A: Foreign Investment in 2014				
Diaspora Firm Stocks in 1996	0.696*** (0.273)	285.064*** (109.929)	0.005*** (0.002)	0.004*** (0.001)
N	231	231	231	231
Panel B: First Stage Dependent Variable: Diaspora Firm Stocks in 1996				
Lineage Driven Diaspora 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 Coastal 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.

The results are presented in Table 11. We find that prefectures with more historical lineage-driven DDI today has higher non-diaspora foreign investments and domestic private investments in terms of both survived number of firms and registered capital in 2014. We call these findings the “seeding” effect of early DDI on future direct investments. The second stage in Panel A of Table 11 also shows the magnitude of seeding multipliers. We see that in an average prefecture, one more diaspora firm in 1996 lead to 0.696 more non-diaspora foreign firm in 2014. This is quite large especially if we think averagely the non-diaspora foreign firms are larger in size. The estimate on number of domestic private firms in 2014 is even astonishing. We find that one more diaspora firm stock in 1996 seed for 285 more domestic private firm at prefecture level. Though domestic private firms are generally smaller than foreign firms, this still shows that early diaspora entrants have per-

sistent effects on the future business dynamism. We also take advantage of the fact that we can observe the up-to-date registered capital in 2014 and estimate the effects of diaspora firm stocks in 1996 on registered capital of non-diaspora foreign firms and domestic private firms in 2014. As expected, accounting for size differences smooths the differences of the estimates for non-diaspora foreign firms and domestic private firms, one more diaspora firm in 1996 bring 0.4-0.5 percent increase in registered capital for both of them. To be noticed that this is not a small multiplier. Recalling the findings in Figure 6, average differences in diaspora firm stock in 1996 at prefecture level are more than 100 and then prefectures with high level of diaspora firm stocks would get 40-50 percent higher registered capital.

These high estimates support the persistent influence of overseas Chinese on economic development in China not only during the early stage of China's Reform and Opening-Up, but also in the long run. The large multipliers also rationalize the national policies in China targeting diaspora investors. The significant positive spillover effect on non-diaspora foreign investments in 2014 is especially surprising as we tend to think of diaspora foreign investments as speculators who sought short-term rents and being inferior in quality to non-diaspora counterparts during China's transition to market economy. One plausible explanation is that the business information acquired by diaspora foreign firms spillover to the non-diaspora foreign firms in the broad international market, and the regions that had received more DDI develop better contracting environment through frequent interactions with diaspora foreign firms.

## 7 Conclusions

The lineage connections between diaspora foreign firms and their ancestral hometowns are believed to be one of the important drivers for China's remarkable industrial growth over the last forty years. Our paper, for the first time in the literature, empirically docu-

ments the prevalence of diaspora foreign firms during the take-off period of Chinese economy and causally identifies the effect of lineage connection on the entry of diaspora firms, leveraging a unique dataset containing the universe of diaspora firms and the surnames of their legal representatives. We show that the effects are primarily driven by reducing information frictions and strengthening contracting enforcement in an immature market environment, thus lowering the entry barriers for diaspora firms with stronger lineage connection.

This paper contributes to an improved understanding of the FDI policy for the developing countries. While opening-up has numerous potential benefits for developing countries, it is often hard for them to kick start the virtuous cycle. Plagued with information frictions and poor contracting environment, 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 contracting 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's share in international trade volumes and FDI inflows rocket up from literally zero to the world-leading level, even before the accession to WTO.

We also find that these early lineage-driven FDI exerts positive large spillovers on non-diaspora investments in the long run, which suggests that the local business environment keeps improving as a result of frequent contacts with diaspora foreign firms. Although the sign and magnitude of FDI spillovers are generally inconclusive in different contexts, our empirical evidences suggest, from a new perspective than the traditional FDI spillover literature, that the multiplying effects brought by introducing FDI as industrial seeds on non-



diaspora foreign direct investments and domestic private business dynamism are large enough in magnitude to outweigh its costs during the take-off stage of China. Using policies levers targeting diaspora communities to attract FDI are by no means unique experiences of China. The International Organization of Migration, in a 2005 survey, found that more than 90% of the countries had policies or programs targeting diaspora<sup>25</sup>. DDI also accounts for a significant proportion of the FDI in other developing countries such as 25% during 1998-2014 in Armenia (Riddle, Hrivnak and Nielsen, 2010) and 26% in India during 1991-2001 (Wei and Balasubramanyam, 2006). A great research avenue for future is to explore whether the social ties between diaspora communities and their ancestral hometowns play similar roles as what we found in the case of China and how large the potential seeding effects of DDI on non-diaspora direct investments in other developing countries.

---

<sup>25</sup>The report is retrieved from [https://publications.iom.int/system/files/pdf/wmr\\_2005\\_3.pdf](https://publications.iom.int/system/files/pdf/wmr_2005_3.pdf)

## 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 Research.
- 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, Julian, 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.
- Kerr, William, and Ramana Nanda.** 2009. “Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship.” *Journal of Financial Economics*, 94: 124–149.
- Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebe, 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.
- Lever-tracy, Constance, David Ip, and Noel Tracy.** 1996. *The Chinese diaspora and Mainland China: An emerging economic synergy*. Macmillan Press LTD.
- 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.
- Riddle, Liesl, George A Hrivnak, and Tjai M Nielsen.** 2010. “Transnational diaspora entrepreneurship in emerging markets: Bridging institutional divides.” *Journal of International Management*, 16(4): 398–411.

- Saxenian, AnnaLee.** 2007. *The new argonauts: Regional advantage in a global economy*. Harvard University Press.
- 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*. Harvard University Press.
- Wang, Jin.** 2013. “The economic impact of Special Economic Zones: Evidence from China.” *Journal of Development Economics*, 101: 133–147.
- Wei, Shang-Jin, and Xiaobo Zhang.** 2011. “Sex ratios, entrepreneurship, and economic growth in the People’s Republic of China.” National Bureau of Economic Research.
- Wei, Yingqi, and Vudayagiri N Balasubramanyam.** 2006. “Diaspora and development.” *World Economy*, 29(11): 1599–1609.
- 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.1 Algorithm to Identify Overseas Chinese and Their Surnames

This section introduces the algorithm we use to identify overseas Chinese 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 A.1 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 A.1.

Below we provide detailed introductions to the algorithm we execute to identify overseas Chinese 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 match 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, the name will be tagged as 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): (1) the match with Japanese surname dictionary is successful; (2) the ID type of the personnel is foreign passport; (3) 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): (1) the match with Korean surname dictionary is successful; (2) the ID type of the personnel is foreign passport (3) 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 leftmost character of the first string as the surname for the ethnic Chinese, if the length of first string of name entry is shorter than 4. On the other hand, We extract leftmost two characters of the first string as the surname for the ethnic Chinese , if the length of first string of name entry is exactly 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 leftmost string or rightmost 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 leftmost and rightmost 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 leftmost character as surname by default (in our database, surnames are more likely to be identified at the leftmost 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 assign each English-spelt surnames into Chinese characters in a probabilistic way with its relative presence in the mixed entries we analyzed below as bootstrapped weight.

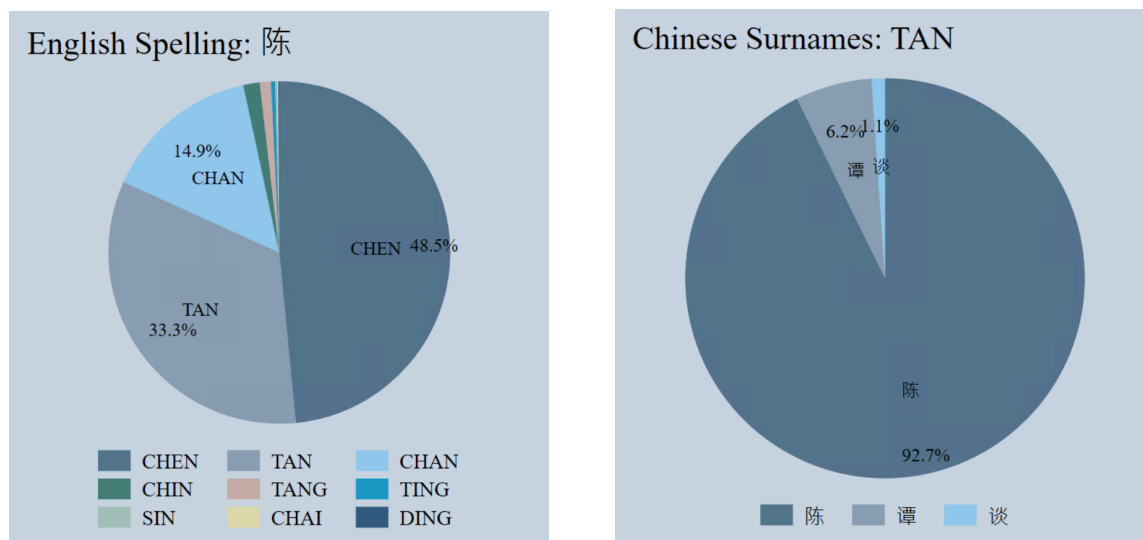
**Names Written in Both Chinese and English.** The group of names written in both Chinese and English serve as a “bridge” between English spellings and Chinese characters. We construct an English-Chinese surname mapping based on the group of names written in both Chinese and English, which enables us to do probabilistic assignment for each English-spelt Chinese surnames into Chinese characters.

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 to do 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 as ethnic Chinese. And we further deem the identified Chinese character as ethnic Chinese’s surname. We also identify the matched English string

as the legitimate spelling for the Chinese character. If both matches are successful, we keep the lefttest 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 randomly assign English-spelt surnames into Chinese characters with empirical weights we observed in the mapping.

**Determining Overseas Chinese.** The previous steps identify whether a person is ethnic Chinese or not. We further separate mainlander and overseas Chinese according to the ID type associated with each personnel. An overseas Chinese is an ethnic Chinese that holds a non-mainland Chinese ID—including passport of a foreign country, travel permit for Hong Kong, Macau, Taiwan (HMT) residents.



**Figure A.1:** An Example of Spelling-Character Mapping between English and Chinese

## A.2 Legal Representative and Executive Position

**Table A.1: 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

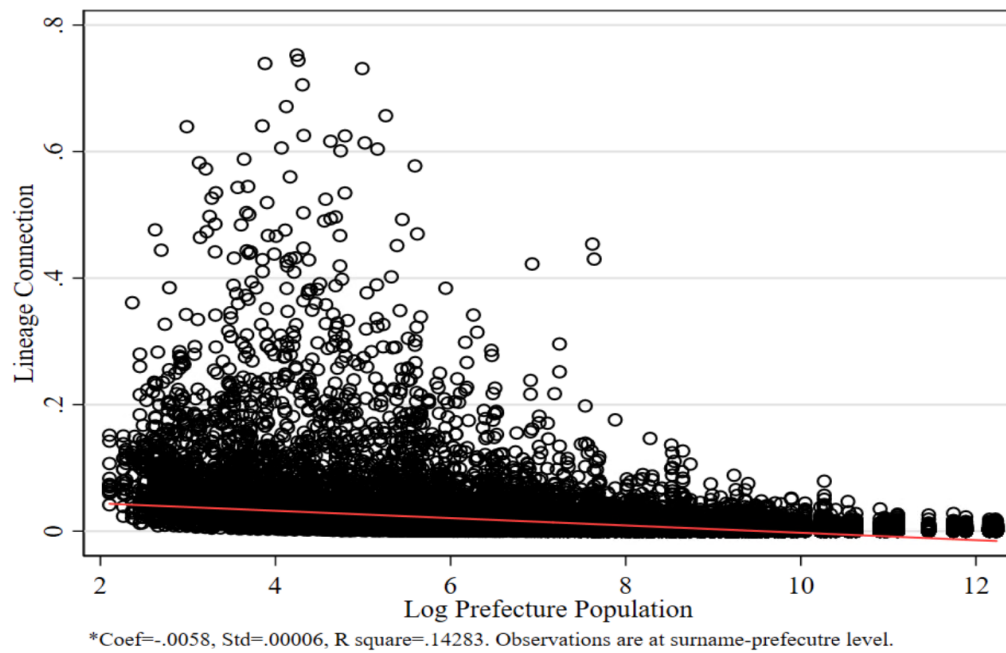
**Note:** The sample includes all registered personnel working in foreign firms ever existed from 1981 to 2014. Conditional on being a legal representative, a person has 93.35%(=1334367/1429323) chance of holding a top executive position within a foreign firm. Conversely, the chance is reduced to 74.41% (=1334367/1793275) for one's being a legal representative given that he or she holds a top executive position.

**Table A.2: Personnel Structure 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 sample includes all registered personnel working in foreign firms ever existed from 1981 to 2014.

### A.3 Lineage Connection vs. Population Size



**Figure A.2:** Lineage Connection Measure vs. Population Size

## A.4 Registered Capital vs. Other Economic Outcomes

**Table A.3:** Registered Capital v.s. Other Economic Outcomes

	Log Registered Capital	
	(1)	(2)
Log Employment	0.025*** (0.003)	0.018*** (0.004)
Log Assets	0.976*** (0.003)	0.416*** (0.004)
Log Sales	-0.124*** (0.003)	-0.005 (0.003)
$R^2$	0.710	0.946
N	150065	124964
Year Fixed Effects	Y	Y
Industry Fixed Effects	Y	N
Firm Fixed Effects	N	Y

**Note:** Sample include 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.



## A.5 Comparison with Other Data Sources on FDI

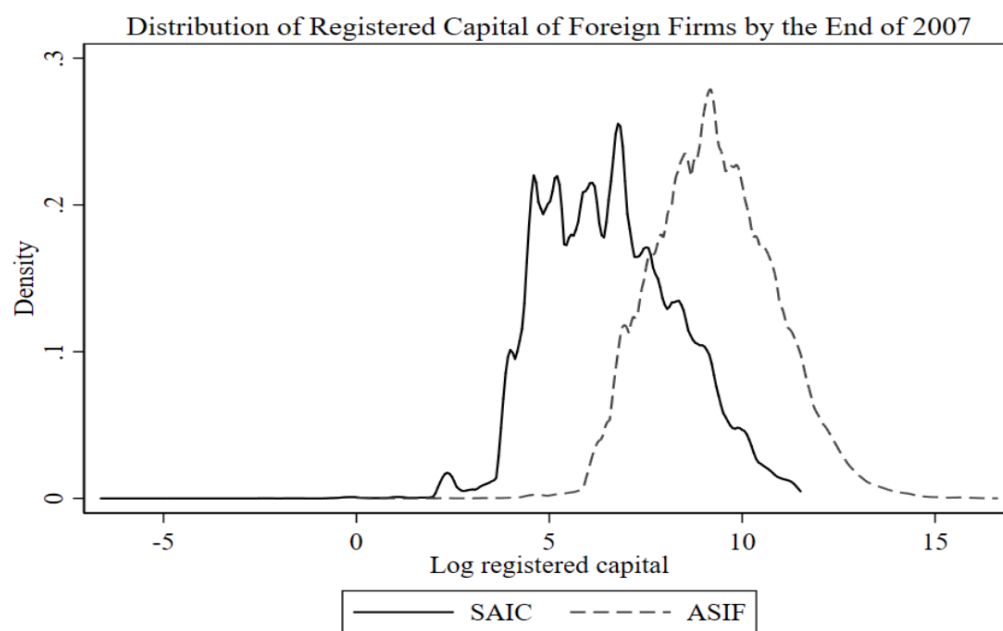
In this section, we illustrate the advantage of our comprehensive foreign firm dataset and cross-validate our data with other data sources.

**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 to 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 A.3 plots the distribution of registered capital of foreign firms by the end of 2007, constructed from SAIC and ASIF respectively. It is clear that SAIC has representation for smaller firms while ASIF does not. Second, Figure A.4 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. And 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 A.5.

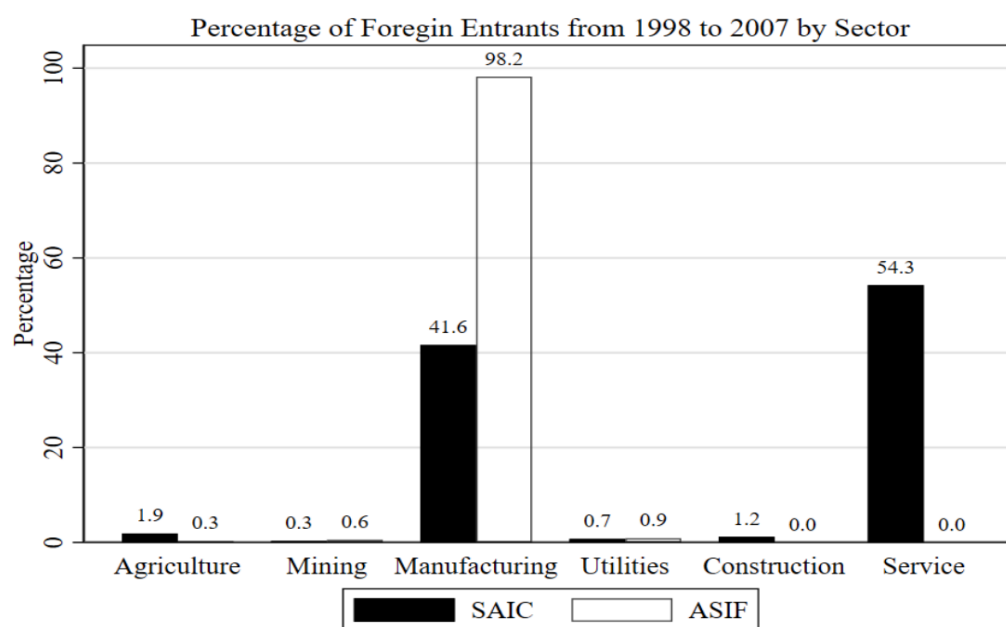
**Diaspora firms vs. HMT firms..** Due to data limitation, prior literature often rely on HMT firms to proxy diaspora foreign firms (Lin, Liu and Zhang, 2009; Huang, Jin and Qian, 2013). There are at least two biases for such approach. First, a lot of Europe-based and US-based multinationals invested 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 A.6, our estimated number of diaspora foreign entrants 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 Hong Kong.

**Foreign firms vs. FDI.** The notion of foreign firm is related to, but not perfectly comparable to foreign direct investment (FDI). To be precise, FDI is a monetary concept.

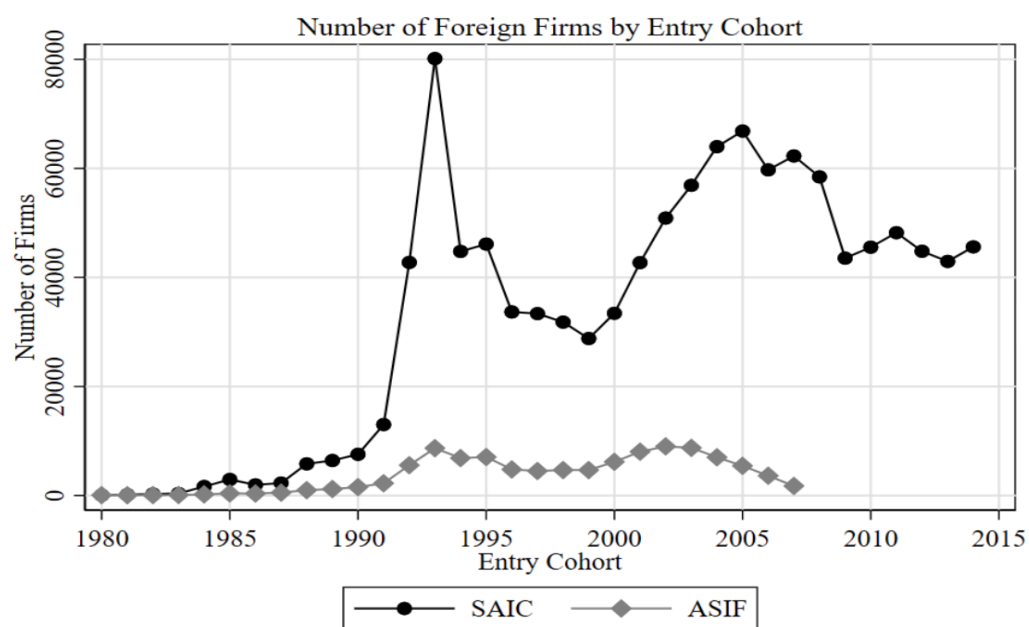
Besides, FDI includes both the initial and follow-up investment from a foreign-owned entity. We do, however, cross-validate our estimates of foreign firms both in counts and volumes with official FDI data supplied by the Ministry of Commerce of China. Figure A.7 suggests that the contracted FDI, measured in number of cases or USD, is highly correlated with our measure of foreign firm entry. And the realized FDI measured in USD is highly correlated with our measure of survived registered capital of foreign firms by 2014.



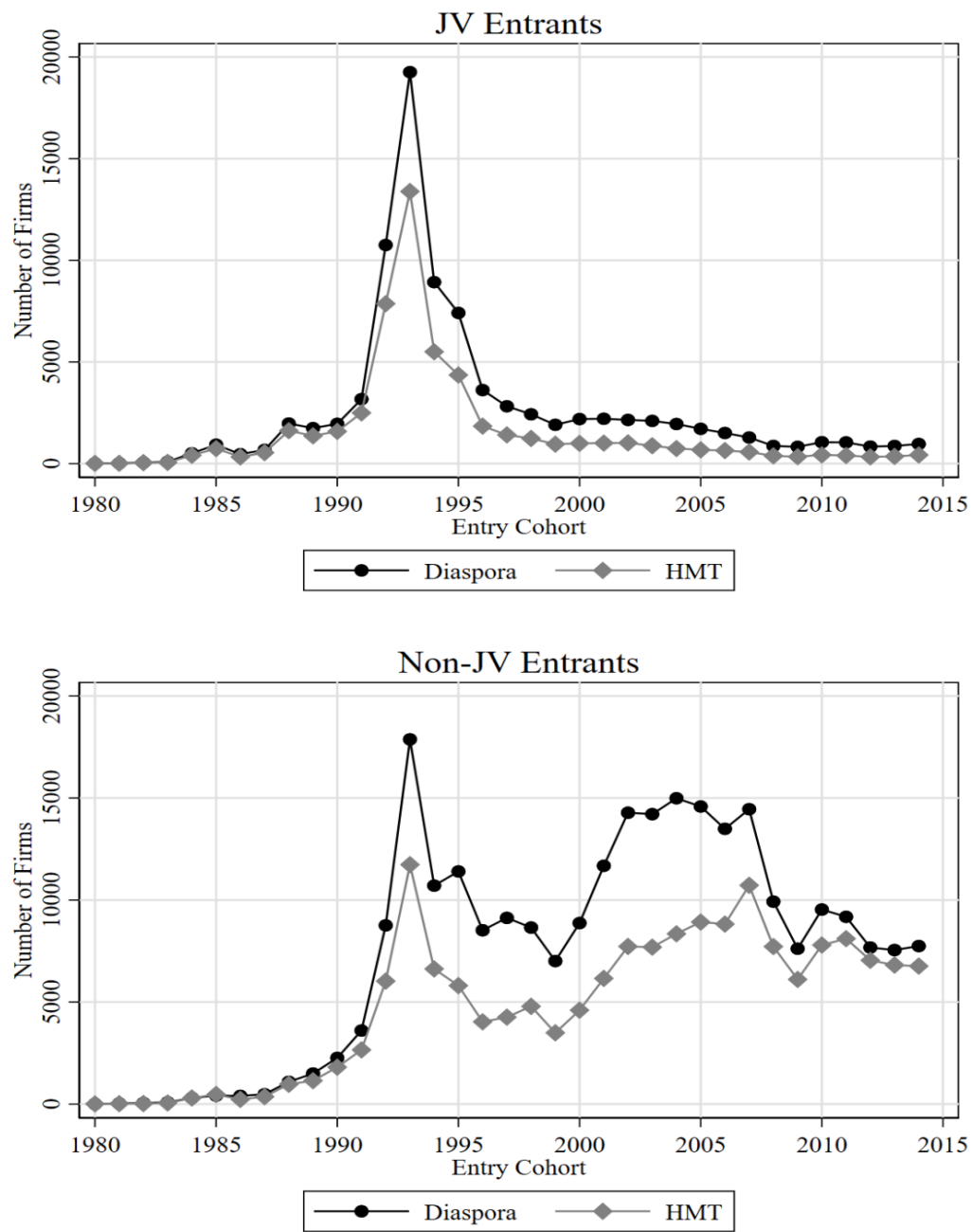
**Figure A.3:** SAIC v.s. ASIF: Firm Size



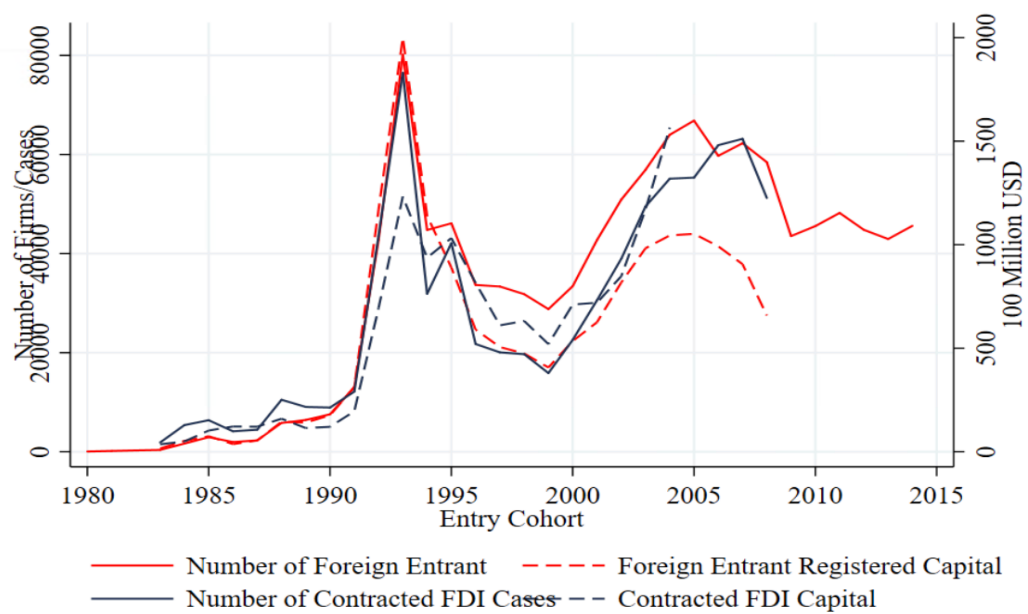
**Figure A.4:** SAIC v.s. ASIF: Sector



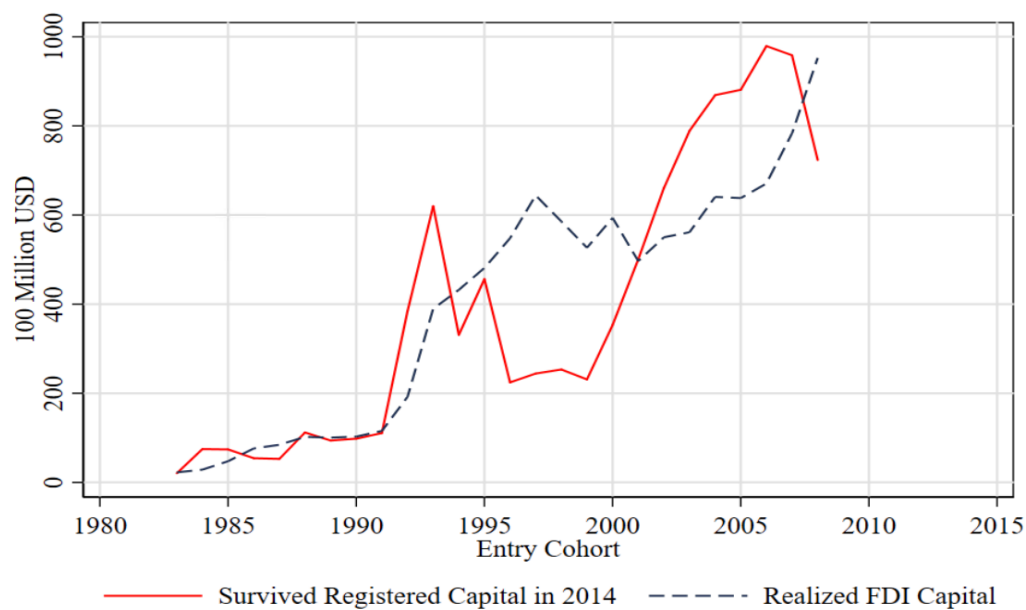
**Figure A.5:** SAIC v.s. ASIF: Counts of Entrant



**Figure A.6:** Diaspora Firms v.s. HMT firms



Contracted FDI data are from China Compendium of Statistics (2009).

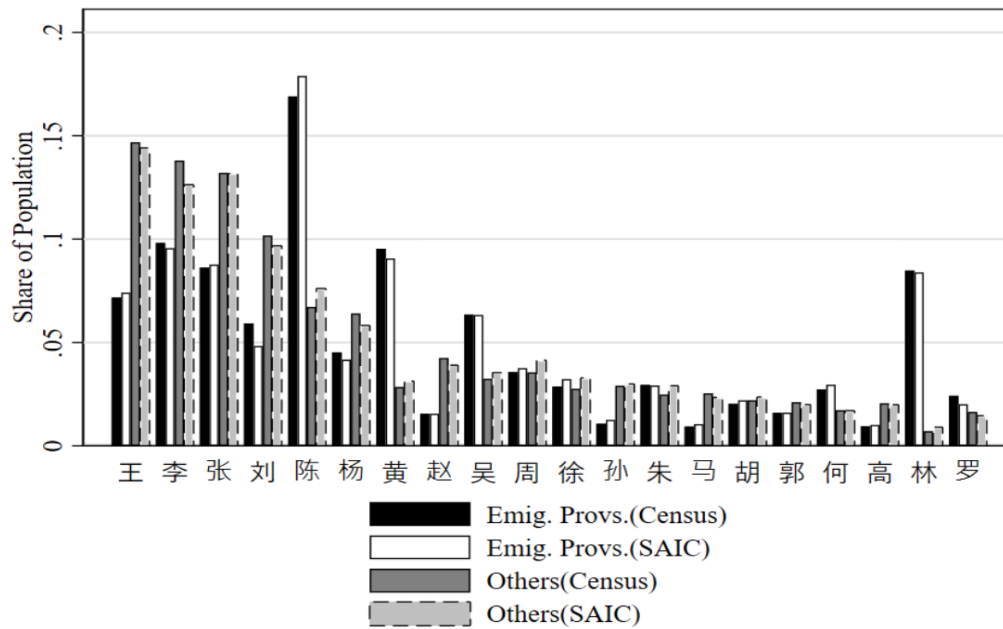


Realized FDI data are from China Compendium of Statistics (2009).

**Figure A.7:** Foreign firms v.s. official statistics of FDI inflows

## A.6 Lineage Connection Measure: Census v.s. SAIC Personnel

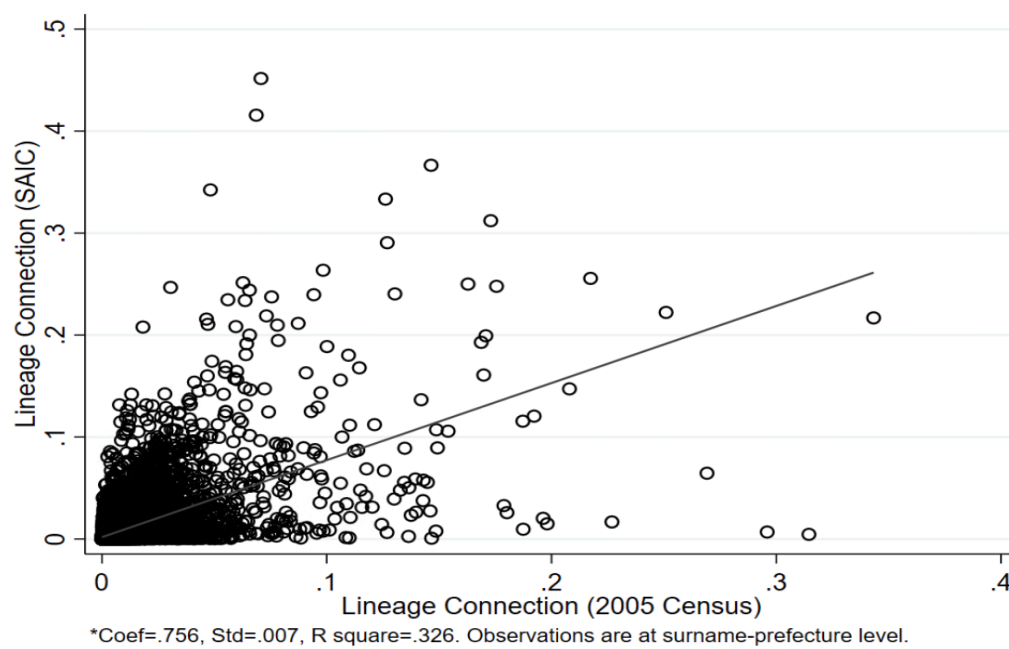
We construct an alternative lineage connection measure from the list of registered personnel that worked in domestic firms established before 1992<sup>26</sup> using the SAIC database. The distribution for top 20 surnames constructed from two data sources are very similar. And overall there is significant correlation between the two measures as shown in Figure A.9.



**Figure A.8:** Surname Distributions: Census v.s. SAIC

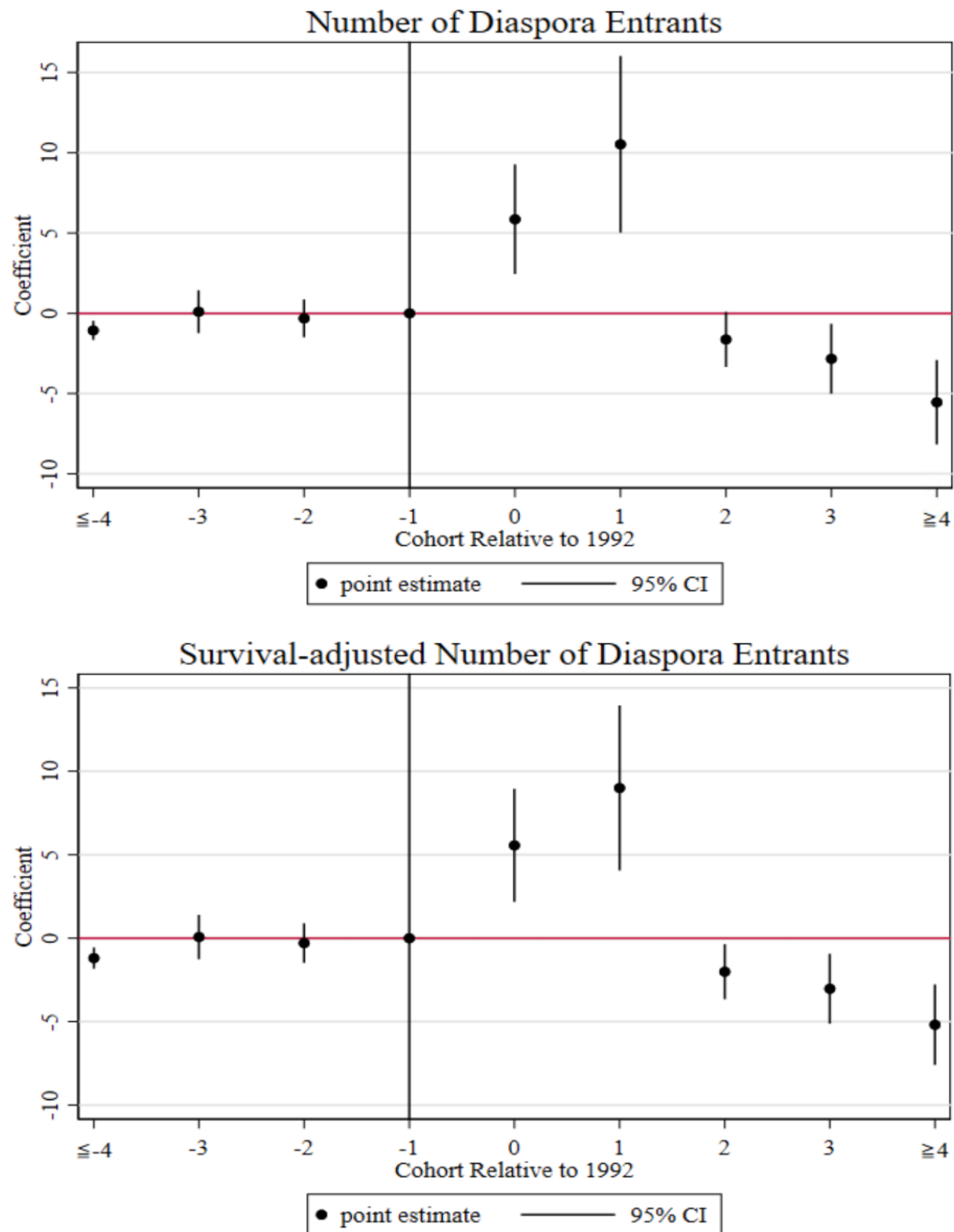
<sup>26</sup>Xu and Zhang (2009) documents that there was almost no internal migration before the early 1990s in China. Private firms are rather scarce then. Most employees and managers worked in township-village enterprises (TVEs) and have lived in the same community for generations.





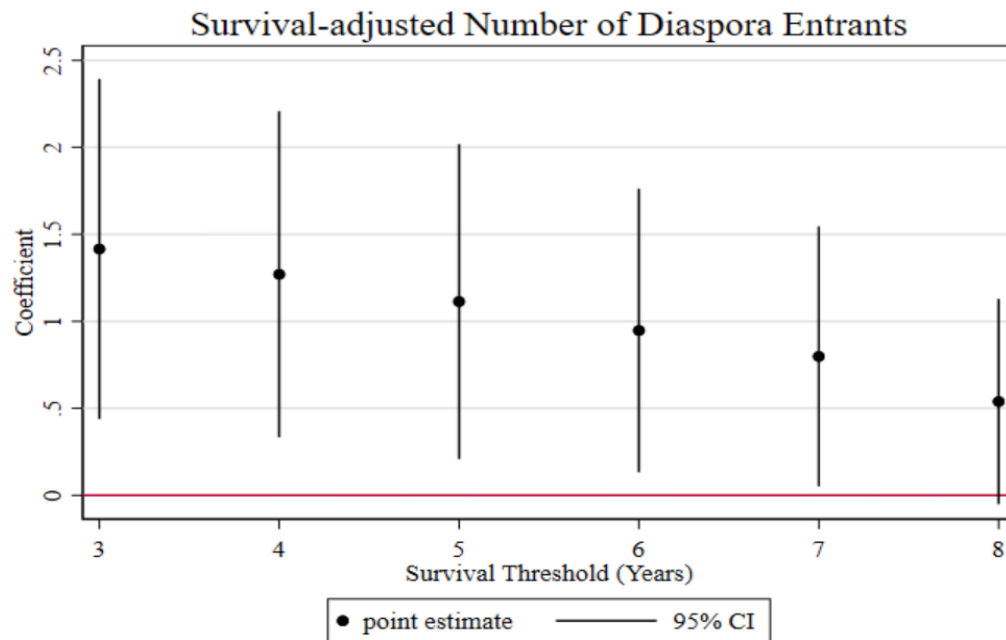
**Figure A.9:** Lineage Connection Measures: 2005 Census v.s. SAIC

## A.7 Validity of the Staggered Opening Shocks



**Figure A.10:** Event Study with 1992 Southern Tour by Xiaoping Deng as the Static Shock

## A.8 Using Different Thresholds for Adjusting Entry



**Figure A.11:** Adjust Entry by Survival with Different Thresholds