

FAMILIAR STRANGERS: OVERSEAS CHINESE, LINEAGE CONNECTION AND FOREIGN INVESTMENTS IN CHINA

Fanghao Chen

Peking University

Ruichi Xiong

University of Toronto

Xiaobo Zhang

Peking University & IFPRI

September 25, 2021

Southwestern University of Finance and Economics

- Governments of developing countries strive to attract FDI.
 - often in the form of introducing new foreign firms.
 - in the hope that FDI generates positive spillovers to the domestic sector (Harrison and Rodríguez-Clare, 2010).
- Immature market environment often deters foreign investments.
- "Why doesn't capital flow from rich to poor countries?" (Lucas, 1990)
 - information frictions
 - institution weaknesses
 - corruption
 - infrastructure insufficiency

- The case of China challenges the conventional wisdom:
 - remained closed for 30 years before the reform in 1979.
 - rose to a hub for multinationals and foreign firms within very short time.
 - IFDI volume only second to the US from 1979 to 1999 (Huang, 2003).
- Anecdotes attribute this achievement to the "Huaren", i.e. overseas Chinese (Vogel, 1990; Tang, 2006):
 - socially connected to ancestral hometowns through lineage.
 - pioneered in an immature market.
 - used informal networks to overcome entry barriers.
 - planted the seed investment for later industrial growth of China.
- How important are the roles of Huaren in explaining the entry and survival of foreign firms? Is there a causal effect of lineage connection? What are the long-term implications?



Figure 1. Examples of Huaren Foreign Firms

- Build a comprehensive dataset on the universe of foreign firms (including wholly-owned, joint-venture, foreign-invested) controlled by Huaren in China from 1980 to 2014, using firm registration data.
 - leveraging the spellings of Chinese surnames.
 - account for more than 60% of all foreign firms; $\geq 90\%$ before 1995.
- Exploit variations from both China's gradual openness and surname distributions across prefectures, and adopt a DID design with differential treatment.
- Find that stronger lineage connection, following the local opening shocks, facilitates the entry of Huaren foreign firms through reducing information frictions and strengthening contract enforcement.
- Show that lineage-driven FDI led to foreign investments in the long run.

Historical Background

- China has a long history of emigration.
 - 1300s-1840s: South East Asia
 - 1840s-1949: North and South America, Australia
 - 1949-1979: HMT
 - 1979-present: US, Canada, Europe
- One of the largest migrant network in the world.
 - 5.8 million in stock (Global Migrant Origin Database, 2007)
 - high income level and social status in host countries
 - 3/4 of \$ 369 billion worth wealth was controlled by Huaren (The Economist, 2020)
- Though as emigrants, Huaren (of earlier generations) remained socially connected to their ancestral hometowns (Kuhn, 2008).
 - During 1949 and 1979, they kept in touch with lineage members through letters and remittance.



Figure 2. Letters, Remittance, Ancestral Temple, and Genealogy

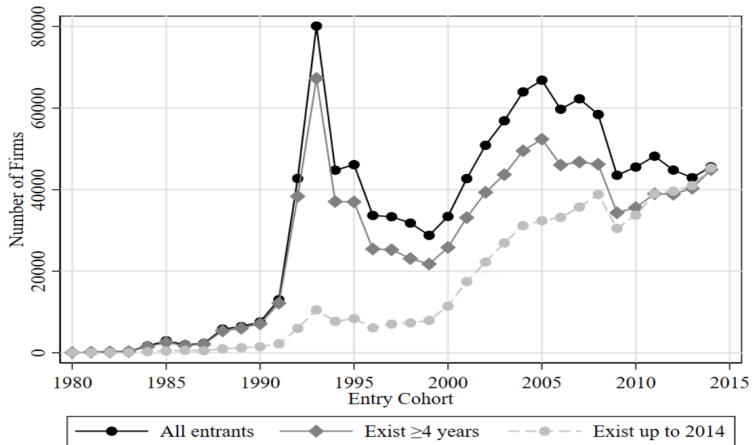


Figure 3. Entry and Survival of Foreign Firms by Cohort

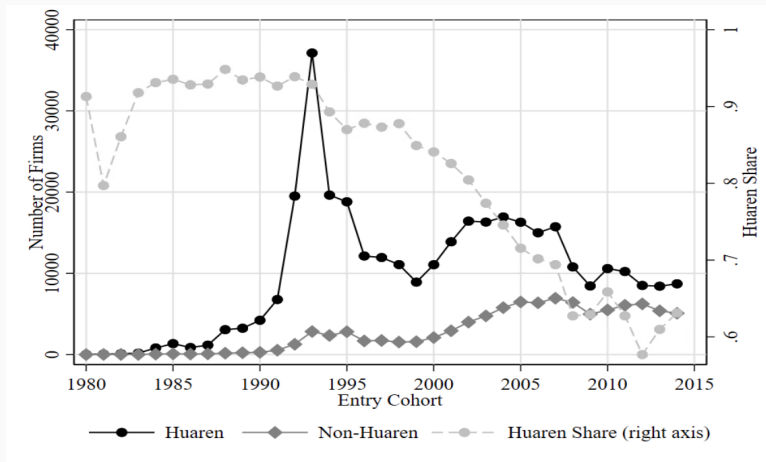


Figure 4. Breakdown of Foreign Firms: Huaren versus Non-Huaren

Related Literature and Contribution

- **Social affinities on trade and investment:** Rauch and Trindade (2002), Javorcik et al. (2011), Ma (2018), Burchardi et al. (2019)
 - a new type of social tie: lineage connection (between surname and locality)
 - a dynamic perspective: separate two margins——entry & survival
- **International migration and empirical studies on names:** Sequeira et al. (2019), Tabellini (2020), Abramitzky et al. (2020)
 - effects of emigration on home countries: facilitate foreign investments
 - use surnames to proxy one's lineage/ancestry
- **FDIs and their spillovers in China:** Cheng and Kwan (2000), Gao (2003), Tong (2005), Amit and Javorcik (2008), Du et al. (2008), Huang et al. (2016), Lu et al. (2017)
 - a causal examining of the role played by lineage network on Huaren foreign firms across Chinese prefectures.
- **Studies on Chinese economy using firm registration data:** Dai et al. (2019), Allen et al. (2019), Bai et al. (2020). We focus on the subset of foreign firms.

- **Firm registration data**

- the universe of firms ever registered in China up till 2014
- entry & exit dates
- industry code (4-digit), region code (6-digit), ownership code (4-digit)
- up-to-date registered capital (2014). **strongly correlated with assets, employment, and sales** (Bai et al. 2020)
- list of shareholders & registered personnel.

- **Population census 2005**

- 0.2% representative sample of population
- region code (6-digit)
- *individual surnames*

- **Other sources**

- CSMAR
- Input-Output Table 2002

Identifying Huaren Foreign Firms & Surnames

- Use ownership code to select the subgroup of foreign firms.
 - any change in the nature of ownership will automatically create a new legal entity (Chen et al. 2019)
 - immediate shareholding structure do not imply ultimate control structure, e.g., the existence of holding shells (Bai et al. 2020)
 - richer features: joint-venture (JV), headquarter or branch, HMT
- Use surnames to identify Huaren.
 - extract surnames of registered personnel in foreign firms. [▶ See details](#)
 - Huaren = Chinese surname + non-mainland ID (passport, HMT residency)
- Use the surname of the Huaren legal representative to proxy the firm's lineage
 - surnames of shareholders in foreign-registered firms not observable.
 - legal rep. unique in every firm. [▶ See details](#)
 - legal rep. \approx highest executive position. [▶ See details](#)

Surname-Based Lineage Connection

- The lineage connection between surname s and prefecture p is:

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

- where E_{sp} denotes the population of surname s in prefecture p .
- a measure of lineage connection not necessarily driven by surname size

► See details

- novelty of this measure:
 - practical: lack of statistics for emigration in China
 - methodological: pre-determined variable (Clark, 2015; Bai and Kung, 2020)
- assume surname distribution remains relatively stable from 1981 to 2005.
 - Hukou registration not formally relaxed until 2000s (Tombe and Zhu, 2019)
 - exclude ethnic minority provinces and prefectures that experience influx of internal immigrants: Beijing, Shanghai, Guangzhou, Shenzhen

► See details

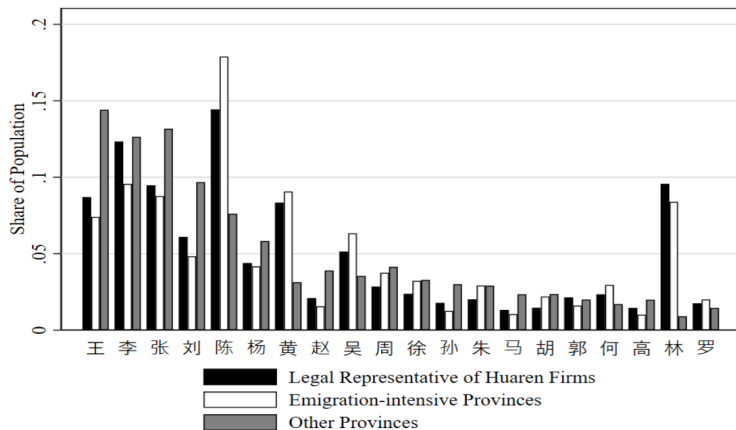


Figure 5. Surname Distributions in Different Populations

Comparison with Other Data and Cross Validation

- Foreign Firm Entry vs. Official FDI Inflows. [▶ See details](#)
- Firm registration data vs. Annual Survey of Industrial Firms. [▶ See details](#)
- Huaren foreign firms vs. HMT firms. [▶ See details](#)
- Huaren foreign firms in Fujian Province during the 1980s. [▶ See details](#)
- Lineage connection measures based on 2005 Census vs. based on firm registration data. [▶ See details](#)

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

- surname s , prefecture p , cohort c (1981-1996, the "Reform Era")
- Y_{spc} : outcome variable, e.g. number of entrants, number of survived firms
- m_{sp} : lineage connection between surname s and prefecture p
- $S_{s,p,c-1}$: number of incumbent firms
- η_{sp} , ϕ_{pc} , θ_{sc} , controlling for:
 - geographic characteristics, entrepreneurial culture
 - place-based policies (and agglomerations)
 - surname-specific advantages
- $Open_{pc}$: =1 if prefecture p has conducted opening reform in year c , =0 otherwise

China's Staggered Opening Reform: $Open_{pc}$

Year	Open Policy	Open Regions
1980	Special Economic Zone	Shenzhen, Zhuhai, Shantou, Xiamen Dalian, Qinhuangdao, Tianjin, Yantai,
1984	Open Coastal City	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

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

- surname s , prefecture p , cohort c (1981-1996, the "Reform Era")
- $\tau = c - \tau^p$ is the normalized time window relative to opening reform
- τ^p is the year of openness reform firstly launched in prefecture p
- $Open_{pc}^{\tau} = 1$ if prefecture p in year c is τ year relative to reform, =0 otherwise
- Common trend assumption holds if β^{τ} insignificant for $\tau < 0$

Baseline Results

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

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

Testing Common Trend Assumption

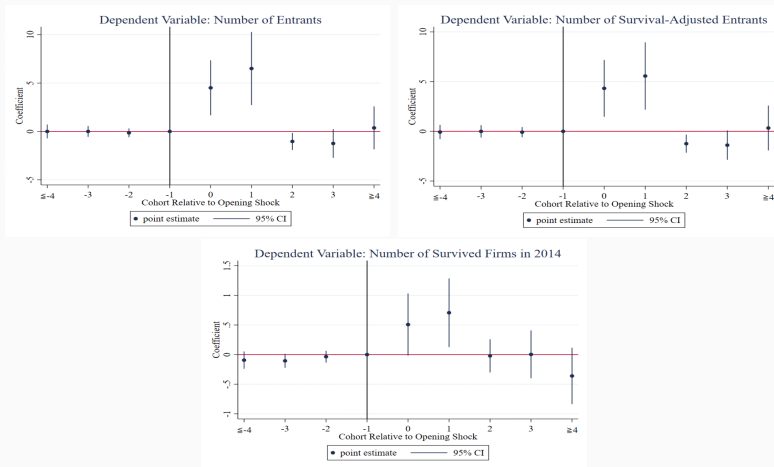


Figure 6. Estimates of β^τ in Equation (3)

Testing Selective Entry Hypothesis

	Survival Ratio in 2014 Conditional on Entry	Average Registered Capital in 2014 Conditional on Entry
	(1)	(2)
Open × Lineage Connection	-0.333 (0.825)	-7040.914 (7984.062)
Adj. R^2	0.364	0.054
N	20,633	20,633
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.

Robustness Checks I: Alternative Dependent Variables

	At Least One Entrant	At Least One Survived Entrant in 4 Years	At Least One Survived Entrant in 2014	Arcsinh (Total Registered Capital in 2014)
	(1)	(2)	(3)	(4)
Open x 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.

Robustness Checks II: Alternative Regression Setups

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

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

Robustness Checks III: Subsample Regressions

	Open × Lineage Connection			
	Excluding Emigration- intensive Provinces	Excluding Emigration- intensive Surnames	Within Emigration- intensive Provinces	Excluding FDI-intensive Prefectures
	(1)	(2)	(3)	(4)
Number of Entrants	0.642*** (0.233)	1.733*** (0.332)	6.001*** (2.476)	1.534*** (0.440)
Survival-Adjusted Number of 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 Entrants is the number of entrants that survive for more than 4 years (included). Emigration-intensive provinces include Guangdong, Fujian, Zhejiang. Emigration-intensive surnames refer to the 20 most populous surnames among all Huaren legal representatives from 1981 to 2014. FDI-intensive prefectures are those whose cumulative number of hosted foreign firms rank top 20 among all prefectures during 1981 to 2014.

Permutation Tests: Reshuffling Lineage Connection Measures Within Prefecture

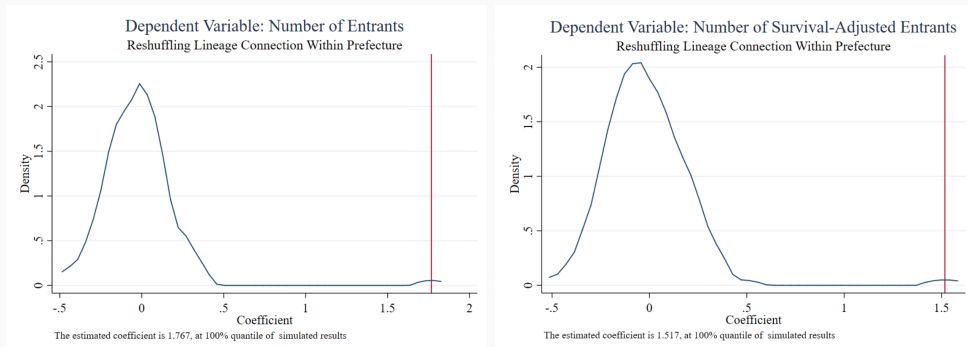


Figure 7. Permutation Tests

Permutation Tests: Reshuffling Lineage Connection Measures Across Prefecture

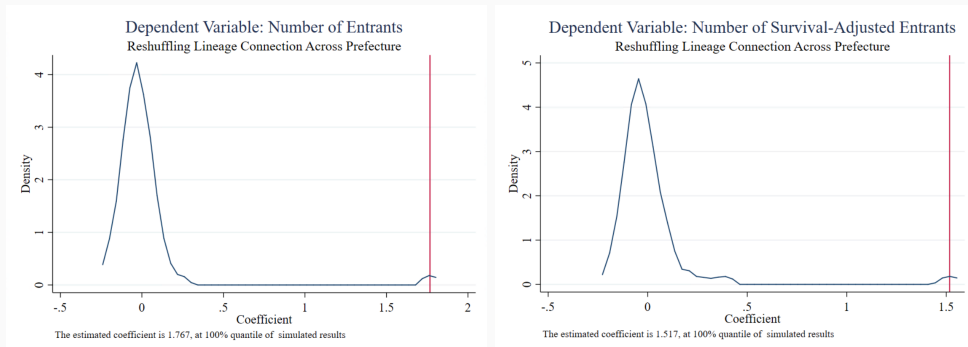


Figure 8. Permutation Tests

Mechanism: Reducing Information Frictions

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

Mechanism: Strengthening Contract Enforcement

	Number of Entrants	Survival-adjusted Number of Entrants
	(1)	(2)
Panel A: Branch Firm		
Open x Lineage Connection	8.795*** (3.363)	7.633** (2.997)
Open x Lineage Connection x Branch Firm	-15.823*** (6.255)	-13.749*** (5.547)
Adj. R^2	0.452	0.444
N	2,690,048	2,690,048
Panel B: Labor Intensive		
Open x Lineage Connection	-1.864* (1.043)	-1.650* (0.928)
Open x Lineage Connection x Labor Intensive	4.990** (2.466)	4.433** (2.226)
Adj. R^2	0.665	0.645
N	2,690,048	2,690,048
Panel C: Supplier Intensive		
Open x Lineage Connection	-2.331* (1.156)	-2.030** (1.010)
Open x Lineage Connection x 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

Lineage-Driven Foreign Entrants as IV for Huaren FDI

$$FDI_p^l = \sum_s \sum_{c=1981}^{1996} \hat{\beta} \times Opening_{pc} \times m_{sp} \quad (4)$$

$$p = \alpha + \theta \times FDI_p^l + \lambda \times X_p + \epsilon_p \quad (5)$$

$$\pi_p = \alpha + \gamma \times \hat{FDI}_p + \lambda \times X_p + \epsilon_p \quad (6)$$

- following Sequeira et al. (2019)
 - isolate lineage-driven FDI that is orthogonal to geographic characteristics, entrepreneurial culture, place-based policies, surname-specific advantages, etc.
- $\hat{\beta}$ estimated from Equation (2)

$$FDI_p^l = \sum_s \sum_{c=1981}^{1996} \hat{\beta} \times Opening_{pc} \times m_{sp} \quad (7)$$

$$FDI_p = \alpha + \theta \times FDI_p^l + \lambda \times X_p + \epsilon_p \quad (8)$$

$$\pi_p = \alpha + \gamma \times \hat{FDI}_p + \lambda \times X_p + \epsilon_p \quad (9)$$

- FDI_p^l : lineage-driven Huaren firm stocks in 1996
- FDI_p : prefecture-level Huaren firm stocks in 1996
- π_p : measures of long-term foreign investments
- X_p is a vector of control variables: log GDP per capita in 1996, years since openness reform, distance to sea, surname fragmentation, SEZ dummy, Coastal Open City dummy, province FEs.

Lineage-Driven FDI Spillovers on Foreign Investments in the Long Run

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

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

Conclusions & Follow-ups

- For the first time, document quantitatively the prevalence of Huaren foreign firms in China.
- Lineage connection lowered entry barriers by reducing information frictions and strengthening contract enforcement.
- Lineage-driven FDI led to long-term foreign investments.
- Implications: Developing countries have comparative advantage to make use of informal institutions.
- More follow-up questions: transition from informal to formal institution during development process.

Thank you

Fanghao Chen

Ph.D. Candidate, National School of Development, Peking University

fhchen2017@nsd.pku.edu.cn

<https://fanghaochen.github.io/homepage/>

Ruichi Xiong

Ph.D. Candidate, Rotman School of Management, University of Toronto

ruichi.xiong@rotman.utoronto.ca

Xiaobo Zhang

Chair Professor, Guanghua School of Management, Peking University

Senior Fellow, International Food Policy Research Institute

x.zhang@nsd.pku.edu.cn & x.zhang@cgiar.org

Backup Slides

Registered Capital vs. 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	15065	124964
Year Fixed Effects	Y	Y
Industry Fixed Effects	Y	N
Firm Fixed Effects	N	Y

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

Algorithm to Extract Surname

- Names of registered personnel in foreign firms:
 - English entries: e.g. "Fanghao Chen" (7.51%)
 - Chinese entries: e.g. " 陈方豪" (91%)
 - mixed entries: e.g. " 陈方豪 (Fanghao Chen)" (1.49%)
- Construct a English spelling-Chinese spelling mapping based on mixed entries.
 - e.g. $1 \text{ Zhang} = 0.8 \text{ 张} + 0.2 \text{ 章}$ (an illustrative case)
- Rule out non-Chinese surnames using Chinese spelling.
 - e.g. Japanese, Korean, transliterated names.
- Manually process highly-unstructured entries.

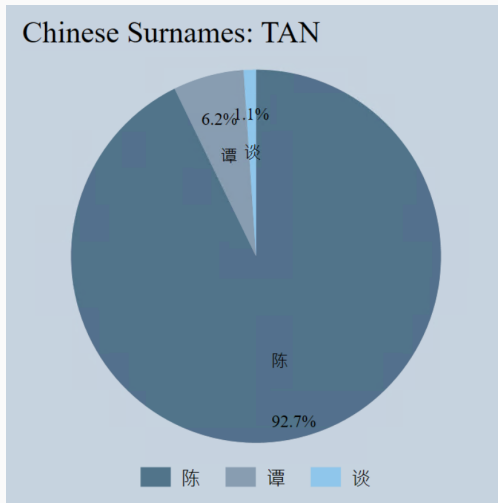
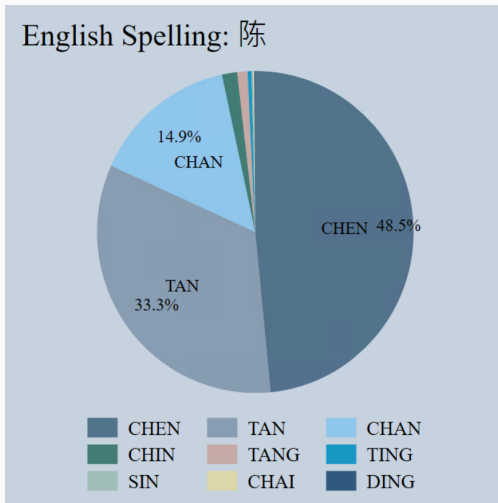


Figure 9. An Example of Spelling-Character Mapping between English and Chinese

Structure of Registered Personnel in 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%

Notes: The samples are registered personnel in foreign firms ever existed from 1985 to 2014.

Legal Representative vs. Top Executive Position within Foreign Firms

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

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

Table A1: Correlation of historical surname share with that in the 2005 census

	Surname share in total population				
	(1) All sur- names	(2) Top-100 surnames	(3) Non-top-100 surnames	(4) All sur- names	(5) All sur- names
Surname share in CBDB	1.134*** (0.194)	1.202*** (0.241)	0.934*** (0.136)	1.176*** (0.223)	1.222*** (0.260)
Surname rank				0.040 (0.029)	0.191 (0.154)
Squared term of surname rank					-0.029 (0.024)
Observations	493	100	393	493	493
R-squared	0.774	0.742	0.347	0.778	0.781

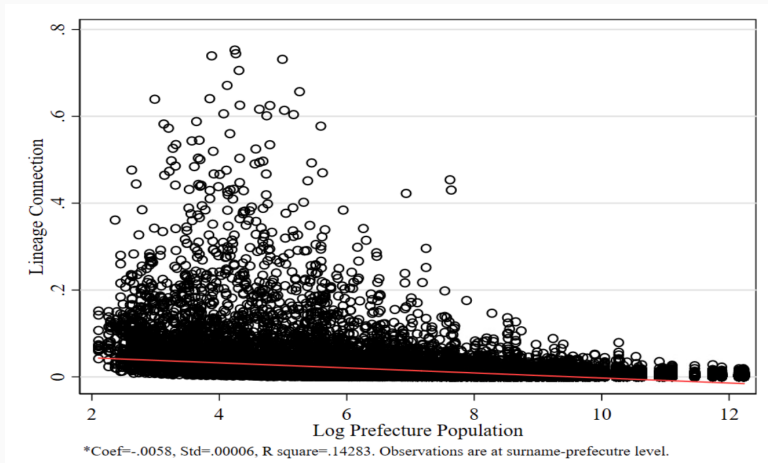
Note: * significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses.

Table A2: Correlation of historical surname share with that in the 2005 census

	Surname share in total population of each prefecture			
	(1)	(2)	(3)	(4)
		Small sam- ple size	Medium sample size	Large sam- ple size
Surname share in CBDB	0.375*** (0.031)	0.255*** (0.038)	0.442*** (0.042)	0.637*** (0.041)
Observations	52,751	17,748	17,748	17,255
R-squared	0.284	0.191	0.339	0.480

Note: * significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses.

Lineage Connection Measure vs. Prefecture Population



Foreign Firm Entry vs. Official FDI Inflows

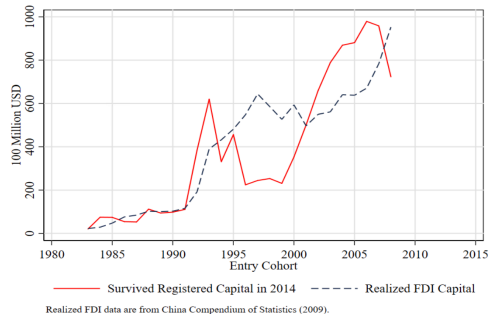
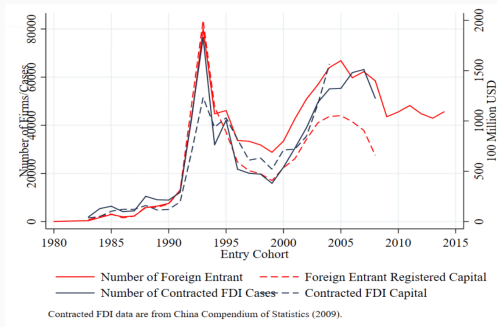


Figure 10. Foreign Firm Entry vs. Official FDI Inflows

Firm Registration Data vs. Annual Survey of Industrial Firms

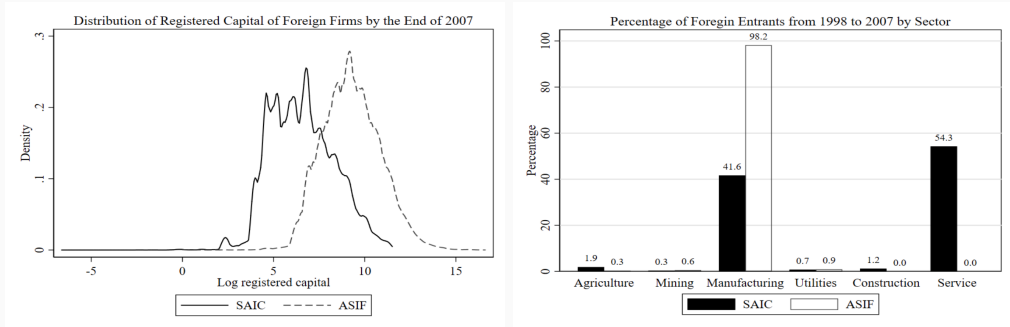


Figure 11. Comparing two popular data sources

Firm Registration Data vs. Annual Survey of Industrial Firms

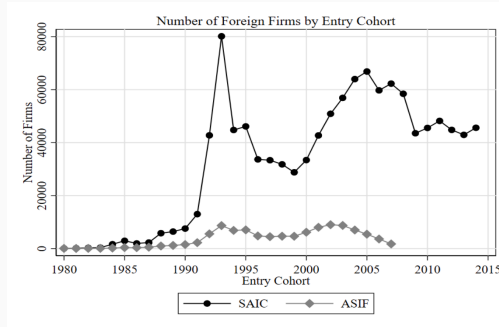


Figure 12. Comparing two popular data sources

Huaren Foreign Firms vs. HMT Firms

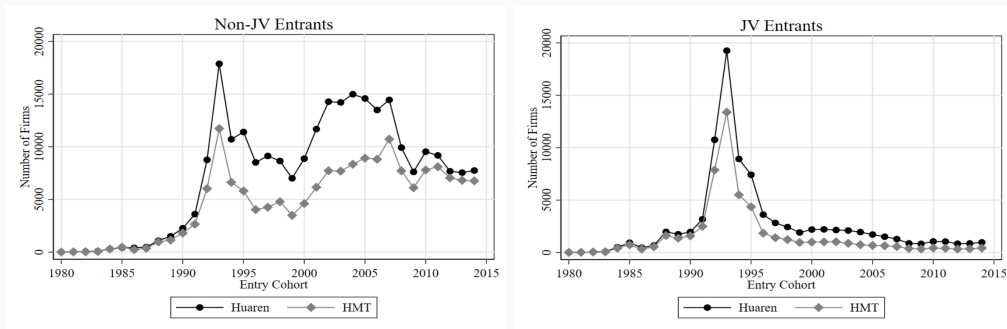


Figure 13. Huaren Foreign Firms vs. HMT Firms

Huaren Foreign Firms in Fujian Province during the 1980s

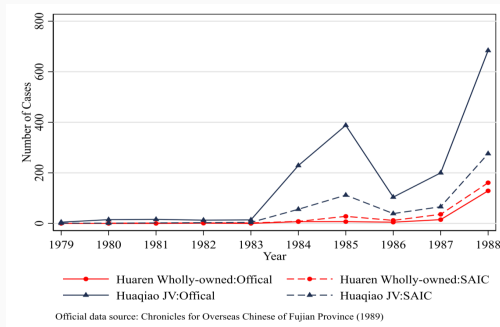


Figure 14. Huaren Foreign Firms: SAIV vs. Local Chronicle

Lineage Connection Measure: 2005 Census vs. Firm Registration data

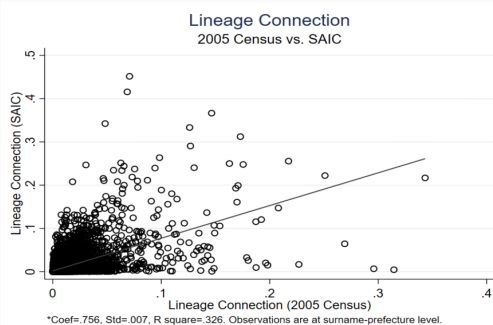
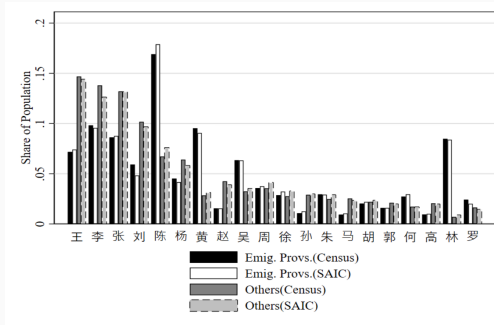


Figure 15. Lineage Connection: Census vs. SAIC