Community Origins of Industrial Entrepreneurship in Colonial India

Bishnupriya Gupta[*], Dilip Mookherjee[†], Kaivan Munshi[‡], and Mario Sanclemente[§]

August 15, 2020

Abstract

We provide evidence of the role of community networks in emergence of Indian entrepreneurship in early stages of cotton and jute textile industries in the late 19^{th} and early 20^{th} century respectively, overcoming lack of market institutions and government support. From business registers, we construct a yearly panel dataset of entrepreneurs in these two industries. We find no evidence that entry was related to prior upstream trading experience or price shocks. Firm directors exhibited a high degree of clustering of entrepreneurs by community. Entry flows and investments were consistent with a model of network-based dynamics.

[*]Department of Economics, University of Warwick, Coventry, CV4 7AL, UK. Corresponding Author Email: B.Gupta@warwick.ac.uk

[†]Department of Economics 270 Bay State Road Boston, MA 02215, USA

[‡]Department of Economics, Yale University, New Haven, CT 06511, USA

[§]QuantumX, Santiago, Chile, 7580206.

We thank the Economic Development and Institutions (EDI) network, the International Growth Center (IGC) and Centre for Competitive Advantage in the Global Economy (CAGE) for financial support. The paper has benefited from comments of Sam Asher, James Fenske, Martin Fiszbein, Walker Hanlon, Debraj Ray and others attending presentations at the 2019 Royal Economic Society conference, 2018 World Economic History Congress, Delhi School of Economics, Paris School of Economics India-China conference, CEPR conference at NYU Abu Dhabi, EDI network meetings at the University of Namur, and seminars in Cambridge, Oxford, Hitotsubashi, Bonn, King's College London, LSE and Warwick.

1 Introduction

Differences in the timing and the determinants of early industrialization across countries have constituted an important area of research in economic history. Most of this research has focused on developed countries today. Property rights and well functioning goods and financial markets created conditions for the first industrial revolution in Britain 1760-1840 (North and Weingast 1989, Bogart and Richardson 2008, Mokyr 2009). The state played a more important role in follower countries that industrialized during the late 19^{th} century, such as Germany and the United States.[1] In Russia, tariffs were raised across the board in 1890 to increase revenue as well as to protect industry (Markevich and Nafziger 2017); the government subsidized railway construction and established policies to attract foreign capital (Kahan 1967). In Japan after the Meiji restoration of 1868, the state coordinated interactions between the financial sector and industry, facilitated imports and diffusion of machinery and technological know-how (Perkins and Tang 2017).

Countries that are less developed today were slower to industrialize, in the absence of well-functioning markets, supportive institutions and interventionist states. Despite this, a cotton textile industry developed and gained market share relative to imports in the late 19th century in pre-independence India. In this paper, we explore the extent to which the vacuum created by the absence of supportive institutions and policies was filled by ethnic networks that exchanged intermediate inputs, shared know-how, connections and capital amongst themselves, and overcame contractual moral hazards via informal community enforcement mechanisms. The setting is colonial India, where market institutions were weak and the colonial state had a limited interest in industrial development, suggesting a possible role of community networks to act as a substitute. In the Indian context, social networks are defined by occupation based caste or religious groups. These groups engaged in the same occupation and married within the group. Intra-group relationships manifested high levels of trust, mutual help and assisted economic transactions and sharing of information.

The role of community networks in solving contract enforcement in trade has been eloquently described in the work of Greif (2006). Social networks have been instrumental in economic activity in Africa (Fafchamps 2003) and in Chinese trade (Rauch 2003). Even in contemporary India, community networks continue to play an important role in credit and insurance (Banerjee and Munshi 2004; Munshi

[1] Effective sectoral tariffs of 30% were widely used in both these countries to protect infant industries (Irwin 2007, Webb 1980). Investment banks in Germany supported the development of large and technologically complex industries (Burhop 2006).

2011), hiring and referrals in labour markets (Beaman and Magruder 2012) and migration flows (Munshi 2003). However, the role of community networks in early industrialization of less developed countries has not received comparable emphasis in the literature. This paper provides evidence indicating the importance of community networks in early industrialization in colonial India.

Our paper supplements the literature on Indian business history with detailed empirical evidence. Business historians have accumulated case study based evidence on entry into industry from trade starting from the second half of the 19^{th} century (Goswami 1985; Rudner 1994, and Timberg 1978; Tripathi and Mehta 1990; Tripathi 2004), focusing on particular entrepreneurs and their communities. A limitation of these rich historical accounts is a lack of quantification and systematic evidence. In this paper, we build a new data set of entrepreneurs and estimate the effect of social networks in industrial entrepreneurship. We examine the cotton textile industry in Bombay, the leading industry in the 19^{th} century where Indian entrepreneurship emerged. We draw upon business directories to gather names of directors of listed firms in upstream and downstream activities of the concerned industry, and code their respective community identities from their names.[2] This enables us to construct a yearly panel data set of active entrepreneurs and their investments by community, track their backgrounds prior to entering the downstream industry and examine patterns of community homophily in the composition of firms. We use this evidence to understand the role of community networks in the process of entry into the downstream industry, while controlling for the role of price and other industry-wide shocks and community characteristics such as prior experience, literacy, population size or outside options.[3]

We use two pieces of quantitative evidence to assess the role of communities. First, we examine the extent to which entrepreneurs clustered by community within firms. Given the high degree of interdependence among entrepreneurs within the same firms, this is a natural way to assess the extent to which problems of trust and cooperation among principal shareholders and executive officers were overcome by partnering with members of the same community. Second, we use data on entry flows and investments to test a model of entry dynamics based on productivity-enhancing help provided by incumbent entrepreneurs to new entrants from their own community. The model is appropriate for

[2] There were two segments of the industry: an upstream component which used light machinery to process raw cotton and bale it for export, and a downstream segment which invested in heavy machinery to produce cotton yarn and cotton cloth mainly for the domestic market. We follow convention and use the term cotton textiles to refer only to the downstream segment.

[3] The data pertains to the stock of active entrepreneurs in different years, which includes the effects of entry as well as exit. However, entry flows dominate exits, so changes in the stock of active agents primarily reflect entry forces. We also show that the results are robust to adjusting for exits.

early stages of industrialization with stable market growth and a given set of communities with stocks of potential entrepreneurs with stationary outside options. Such early stages are typically characterized by growing stocks of incumbents from each community. The model generates network-based dynamics of community-specific incumbent stocks, in which differences in initial presence of different communities exponentially amplify over time. The network effect can be identified by the presence of a non-linear divergence effect for incumbent stocks, while controlling for year dummies (which include the effect of price and other industry-wide shocks) and community dummies (which capture differences in levels of community-specific unobserved characteristics such as education, ability, wealth and outside options).

The dynamics of entry into downstream cotton in the Bombay region after the US Civil War turn out to be consistent with predictions of the network-based model. The evolution of active entrepreneurs from different communities during early stages exhibited the nonlinear amplification of early community presence predicted by the model. We use yearly data for stocks of active entrepreneurs at the community level from 1866 until 1890. For reasons explained within the paper, given the identification assumption described above we are able to infer the existence of a network effect, but not obtain an unbiased estimate of its magnitude. Under the stronger assumption of exogeneity of initial community presence at the end of the US Civil War, we estimate the effect of an additional active entrepreneur from a given community in 1866 to be 2.85 additional entrepreneurs from the same community in 1880, and 5.75 additional entrepreneurs in 1890.

A potential alternative explanation for these findings could reside in differences in time trends of unobserved community characteristics such as wealth or ability. For instance, the exceptionally high levels of education, expertise and connection with the Western world of the Parsi community have noted by many historians (Anstey 1952, Buchanan 1934, Desai 1968 or Tripathi 2004). If education or wealth of the Parsi community were growing faster than the other communities, it could also generate a growing divergence between their respective incumbent stocks over time. However, the historical evidence reviewed in Section 2 reveals no such pattern of divergent trends in literacy or population shares; the differences across communities remained stationary in the late 19th century. To further address this concern, we use evidence based on entry-level investments of different communities. Under the alternative hypothesis, investments of Parsi entrants (and more generally, of communities with higher initial presence) should grow faster. We do not find any evidence of this.

In fact, it turns out that the trends in entrant investments were exactly consistent with a pattern of misallocation predicted by the network model: *marginal entrants from communities with larger initial* *presence are less productive and thus invest less. Moreover, the magnitude of this divergence should accelerate over time.* Our theoretical analysis shows the same pattern can appear for average entrants as well, under suitable conditions on the distribution of (unobserved) ability. While both average and marginal (10th percentile) investments of new entrants exhibit weak evidence of lower time trends, the divergence accelerated over time, an effect that was statistically significant. Hence the evidence on investment trends provides additional justification of our underlying identification assumption of parallel linear trends in unobserved community characteristics.

Besides community networks, we also examine the role of (and control for) other factors often emphasized in the literature as important determinants of early industrial entrepreneurship, such as pre-industrial accumulation of wealth (Marx 1887, Banerjee and Newman 1993) or experience in related upstream trading sectors in certain African countries (Sutton and Kellow 2010, Sutton and Kpentey 2012). In the context of the Bombay cotton industry, we find no evidence of a significant association of entry with prior upstream presence of the community during the US Civil War, a period during which Indian cotton traders achieved high export volumes and profits. 66% of entrepreneurs active during 1860-70 had no prior upstream experience in baling and trading raw cotton, a proportion which rose to 79% and 91% in the subsequent two decades. Hence accumulation of wealth or experience in the upstream sector was not a pre-condition for entry into the downstream industry. Nor were yearly cotton price movements an important determinant of entry patterns.

In the absence of a randomized experiment, identification of network effects is always difficult, owing to the possibility of alternative explanations based on unobserved covariates that might be correlated with network co-membership or size. In historical settings, of course, randomized experiments are not possible and one has to rely on historical data. Inferences regarding network effects then depend on the relative plausibility of alternative explanations. In our setting, the facts we document concerning divergence of initial disparities in entrepreneurship across different communities would have to rely on unobserved growth in access to knowhow or wealth which is selective to specific communities in the same location, and which is associated with shrinking size of entrant investments. Such explanations seem hard to conceive of, and seem inconsistent with available historical evidence, though of course we cannot rule them out definitively.

The remainder of the paper discusses the role of community networks in subsequent evolution of the cotton textile industry after 1890, as well as the the jute industry after the First World War. In the latter case we provide similar quantitative evidence: firms exhibited high degrees of clustering by community for newly entering Indian entrepreneurs, and the dynamic network model successfully predicts the evolution of community presence. Section 2 describes the historical background of the emergence of industrial entrepreneurship in India during the 19^{th} century, and the role of various communities in Western India that played a role in the development of the cotton textile industry. The theoretical model of networks is presented in Section 3, followed by data and descriptive statistics in Section 4 and empirical results in Section 5 for the cotton industry. Section 6 provides corresponding evidence from the jute industry after the First World War, while Section 7 concludes.

2 Emergence of Indian Entrepreneurship in the 19^{th} Century

Historical Origins

In the 1850s, the first cotton textile firm was set up by an Indian entrepreneur in Bombay, and the first jute textile firm by a British entrepreneur in Calcutta. But industrial investment did not pick up for another decade. By 1900 cotton and jute textiles accounted for just under half of total investment in Rupee Companies, registering in India. See Table 1.[4] Cotton textiles became the largest industrial sector, employing 17% of the workforce in Bombay city in 1901. The two main cities of industrial development were Bombay in the west and Calcutta in the east. Industrial investment was segregated by British owned companies in Calcutta and surrounding regions, and Indian owned companies in Bombay city and surrounding regions. British and Indian social networks were typically segregated into different sectors (Gupta 2014). The development of cotton textiles in Bombay city was initiated by Indian entrepreneurs. These entrepreneurs came from Indian trading communities in and around the city of Bombay.

These communities acted as agents of the East India Company in the settlements of the East India Company, Surat and Bombay on the western coast. They were also the contacts between the Company and the local consumers and producers. The five principal trading communities, included the Parsis, a small group of Zoroastrians who had migrated from Persia from the 8^{th} century, Hindu and Jain Vanias, the Muslim merchant communities of Bohras and Khojas, Jews who had migrated from Baghdad in the 18^{th} century and the Bhatias, a small Hindu sub-caste. The Hindu and Jain Vanias had a high degree of economic and social interactions (including inter-marriage), and will thus be treated as a single community ('Vanias') in the rest of the paper. These communities specialized in different pre-industrial sectors: Parsis in the opium trade in China, Vanias in moneylending, and the Muslims in

[4] Most of the investment in the railways and tea was in sterling companies and is not captured here.

Table 1: Share of Modern Sectors in Paid up Capital raised in India (%)

	Banking & Insurance	Transport	Mining & Plantation	Textiles	Food Industries
1851-65	35	7	20	12	0.6
1881	14	5	23	37	2
1890	13	8	19	46	1.8
1900	13	9	14	47	1.6

Source: Rungta 1970 (Appendix 8, 17)

Note: Textiles include the upstream and downstream of cotton and jute (95% of the mills are classified as specializing in one product until 1865, thereafter 60%). Rungta (p.284-85) notes that many of the firms were classified as multiple textiles if they showed capacity to produce more than one variety: cotton, jute, hemp, silk and wool. In practice only six produced a variety. The rest produced cotton textiles or jute textiles.

shipping routes to East Africa, while all were active in the raw cotton trade and as trading agents for Europeans within India.

The hinterland of Bombay city was India's cotton producing region. Raw cotton from the region was sold in the regional market and exported to Europe in periods of high demand. The civil war in the USA was one such period. The American civil war years of 1861-65 witnessed a big rise in export of raw cotton from India to Britain, as India replaced the United States as a source of raw cotton for the British textile industry. This provided a big impetus for investment in cotton pressing and baling, preparing raw cotton for export. With the end of the civil war, exports declined and the export trade in raw cotton became less profitable. The following decades saw entry into the downstream industry of textile production. The legalization of limited liability joint stock companies, investment in joint stock companies in cotton textiles became an attractive option, and cotton mills were set up in the city of Bombay. A number of early entrepreneurs were from the Parsi community. But other Indian merchant communities of Bombay were also involved. European firms had a large presence in the upstream sector of cotton press and baling that sold raw cotton for export. In the downstream industry, their presence was overshadowed by the Indian merchant groups. Morris (1982, p.580) suggests that European capital did not exceed more than 10-20% of total investment in the industry in Bombay.

Role of the State

In 1858, India came under Crown rule. This ended a 100-year rule of the East India Company and formally integrated India into the free trade regime of the British Empire. Under Company rule, the share of revenue from trade taxes had been small. British exports to India incurred tariffs of 5% on cotton piece goods and 3.5% on yarn. In 1859, this increased to 10% on piece goods and 5% on yarn as revenue concerns overrode the free trade argument (Harnetty 1972, p.7-10) and was strongly resisted by Manchester Chamber of Commerce. Under Crown rule, the British state in India was receptive to political lobbying by British interest groups. By 1862, tariffs on cotton textiles had been lowered to the previous level (Harnetty 1972, p.26). When Manchester demanded compensating excise on Indian products on the ground that the 5% tariff offered protection to Indian producers contrary to the principles of free trade, the government conceded. Hence for the post-1865 period that we study, tariffs for revenue purposes were combined with a countervailing excise tax on import competing goods, effectively equivalent to free trade. The early development of the Indian cotton textile industry was therefore not aided by protective tariffs as in the case of early industrialization experiences of USA, Germany and Russia.

Community as a Source of Capital

The cost of setting up a cotton mill was significant. Morris (1982, p. 575) shows that it cost between Rs 500,000-1 million to set up a textile firm in Bombay in the early phase. Although there was a formal banking sector and the Indian trading communities had shares in banks, the latter did not provide long term capital to industry. Bank lending was limited to short term working capital loans. The business historian Tripathi (2004) writes:

> "The tight hold of the mercantile interests on the sources of industrial finance could have been loosened, had there been an efficiently functioning stock market and an alternative source of credit, such as banks organized on modern lines. Such institutions were woefully underdeveloped at the turn of the century."

Before the establishment of a formal stock exchange in 1875, shares were sold by a handful of brokers in the metropolitan cities in open spaces. The value of individual shares was high and was typically bought by wealthy merchants and European residents in India. When the Bombay stock exchange was established in 1875, the number of brokers increased and the value of the average share came down, making purchase of stocks more widely available. High dividends of 10-15% offered by textile firms, made such investment comparable to the profit rate in trade and money lending (Rungta 1970, p.158).

However, community connections were important in selling shares. The capital for the first cotton mill in Bombay, set up by Parsi entrepreneur C.N. Davar, was raised from fifty of Bombay's leading merchants and the majority of investors were Parsis (Morris 1983, p. 574). A few years later, when Ranchhodlal, who did not belong to a trading caste, tried to set up a textile firm in Ahmedabad, local merchants were not willing to invest in his firm (Mehta 1991, pp 182-3). Similarly, when Tata, a Parsi, offered shares of the first cotton mills outside Bombay city to a prominent Marwari trader, the response was negative (Tripathi 2004, p. 121). Even for a late entrant in 1897 such as Lalbhai from the Jain community, friends and family were the main source of capital (Tripathi and Mehta 1990, p. 90). The capital market thus remained highly segmented despite the presence of a stock exchange. To quote Roy and Swamy (2016, p.146):

> "The typical industrial firm in 1900 was a company with shareholding of family and friends as well as the public and managed by another firm, which was a partnership or a company closely held by the same family or on rarer occasions, a trust. The company had a legal entity as a public body, but it was managed like personal property."

Community as a Substitute for Contract Enforcement and State Intervention

Industrial entrepreneurs and traders faced problems of contract enforcement, as legal institutions in India in the middle of the 19^{th} century were weak. There was no civil code for contract enforcement. During Company rule, problems related to contract enforcement appeared in different contexts: from the procurement of textiles from weavers to cultivation of opium and indigo, the East India Company adopted ad hoc solutions using customary systems. The Indian Contract Act was passed in 1872, but as with other legal interventions, it did not create an institutional setting similar to that in Britain. Roy and Swamy (2016) argue that Indian traders, creditors, and manufacturers largely continued to rely on pre-existing community norms and institutions that were outside the structure of formal British-Indian law. Community networks as described by Greif (2006) were the primary means of contract enforcement in trade and industry for the caste and religion based communities. Information and trust through the community network became a substitute for well-functioning public institutions.

Differences in Wealth, Literacy and Outside Connections Between Communities

A challenge to identify the role of community networks in facilitating entry into industry is the problem of distinguishing trends resulting from a dynamic network effect from trends in underlying community characteristics. If some communities had privileged access to capital or enjoyed certain benefits due to their connections with the British in India, and these were growing at rates positively correlated with initial entry patterns, it would confound the effect of the dynamics generated by the social network itself. We now review the available historical evidence on differences in trading experience, wealth, education, or outside options across the different communities during the 19th century. While there may have been differences in *levels* of some of these characteristics, we argue there is no evidence of any significant differences in time trends.

Prior to 1850, all five communities were active in trading, though in different commodities (opium, shipping and moneylending) and regions (spanning China, the Middle East, Africa and Europe). In 17*th* and 18*th* century, there was no difference between the Parsis and the Vanias, or Bohra and Khoja Muslims in their business connections in the region; all of them acted as agents for Europeans in the Indian market (Das Gupta 2001). Table 2 shows the importance of the various communities in Surat in the mid 18^{th} century.

Table 2: Non-European Trading Capital 1746 in Surat: Shares of communities (%)

	European protection	Independent	Total
Jew	1.2	0	1.2
Parsi	9.9	0	9.9
Hindu*	18.8	30.5	49.3
Muslim	2	31.7	33.7
Total	31.9	68.1	100

Source: Guha (1984). * refers primarily to Bhatias and Vanias.

The external environment or British policy did not favour any particular community. Tariffs and excise on textiles were levied uniformly. All trading groups collaborated with the British in trade as suppliers of raw materials and distributors of British goods. There is no evidence suggesting that any particular community had better access to credit. The Bank of Bombay financed cotton traders in the early stages and later on provided short term capital to the cotton mills (Bagchi 1987, p.110). Members of all communities were involved in the management of the bank. Parsis, Bhatias, Baghdadi Jews, Hindu and Muslim merchants were on the board of directors between 1876 and 1920 (Bagchi 1987, Appendix). Several cotton textile firms belonging to different social groups enjoyed a high credit limit with the bank discounting bills or borrowing against goods (Bagchi 1987, p. 337-8). They all relied on credit offered by British machinery exporting firms.

However, the communities differed in education and outside options. The merchant castes enjoyed higher literacy than the average for the population. Within trading communities, the Parsis were exceptional in their educational attainment. In 1881, over 70% of the Parsis were literate and 50% of the literates had secondary education comparable to the literacy of the Europeans in India. 5% of the literate Parsis and 4% of the literate European Christians had college education. Other communities were far less educated. The Vania trading groups had primary education, but a very small proportion went to college. Table 3 shows the differences in literacy rate among the different communities. However, *the key point to note is that literacy rates did not materially change over time for any community.* Nor do we see differential growth in population in Bombay city (Table 4) where most of the cotton textile industry was located until the 1890s.

Table 3: Literacy by Community in Bombay Presidency (%)

	1881	1891	1901
Hindu[1]	10.9	11.8	11.2
Parsi	72.9	76.3	75.2
Jain	51.5	53.4	48.9
Jewish	52.0	54.2	-
Muslim[2]	6.8	9.2	7.3
Christian	38.9	44.4	36.5
Total Bombay Presidency	11.1	12.6	11.67

Source: Censuses of India

[1] Hindus include Bhatias and Vanias. The 1911 Census shows male literacy for Bhatias was 56% and Vanias 60%.

[2] Within Muslims, the 1911 Census shows male literacy rate of 41% among the trading communities of Muslim Bohras, Khojas and Memons.

Table 4: : Population Share in Bombay City by Community (%)

	1864	1872	1881	1891	1901
Hindu[1]	64.1	63.4	65	-	65
Bhatia	2.7	1.5	1.2	-	1.2
Parsi	6	6.8	6.3	5.8	6
Jain	1	2.3[2]	2.2	-	1.8
Jewish	-	0.4	0.4	-	0.7
Muslim	17.8	21.4	20.4	18.9	20.1[3]
European Christian	1	1.1	1.4	-	1.5

Source: 1872- 1891: The Gazetteer Bombay City and Island 1909.
[1] Hindu includes Bhatias.
[2] Includes Buddhists (only 2% of the Population in 1881). 1901 data comes from the Census.
[3] Includes Bohra 1.5% and Khoja 1.4%.

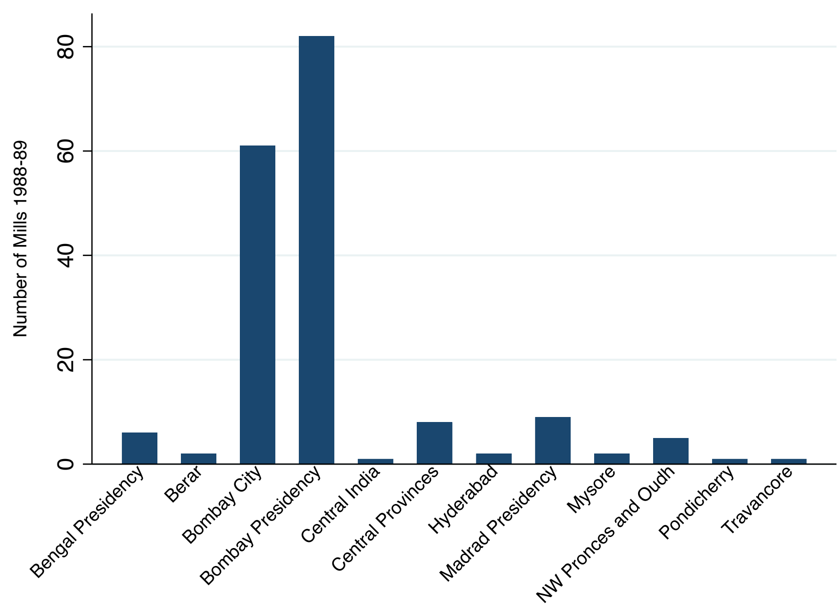
With regard to outside options, many of these trading communities experienced adverse price shocks in the respective commodities they traditionally specialized in during the first half of the 19th century, thereby motivating their transition into industrial entrepreneurship. This applies to the Parsis, Bhatia, Baghdadi Jews and Muslim Khojas. For instance, the Opium Wars in China during the middle of the 19^{th} century reduced the involvement in opium for the Parsis. The competition from European shipping from the late 1860s with the opening of the Suez Canal, made shipping unprofitable. And as the boom in cotton trade during the American civil war ended, the profits from the upstream industry exporting cotton bales evaporated, motivating entry from these communities into the downstream industry.

However the experience of the Vania traders in Ahmedabad was markedly different. Unlike the other four communities, their principal outside option was money lending until the late 1870s, where returns remained high until new legal changes affected moneylending. The peasant riots against moneylenders in 1875 in the Deccan region of Bombay Presidency, targeted to destroy debt records held by money lenders (Kranton and Swamy 1999). Following the riots, the government became increasingly concerned about agrarian distress from indebtedness and began to take the legal route to restrict moneylending. The Deccan Agricultural Relief Act of 1879 which regulated interest rates on loans to agriculture was the first of a series of legal interventions. Consequently, industry became more attractive as a contender for investment for the moneylenders only after the 1880. Hence compared to the other four communities, the Vanias of Ahmedabad were latecomers to the (downstream) cotton textile industry. The size of their population was substantially larger than the Parsis, Bhatias and the Muslim trading groups in

Bombay. After 1890, this group came to dominate the industry, entering the downstream industry in the Ahmedabad region.

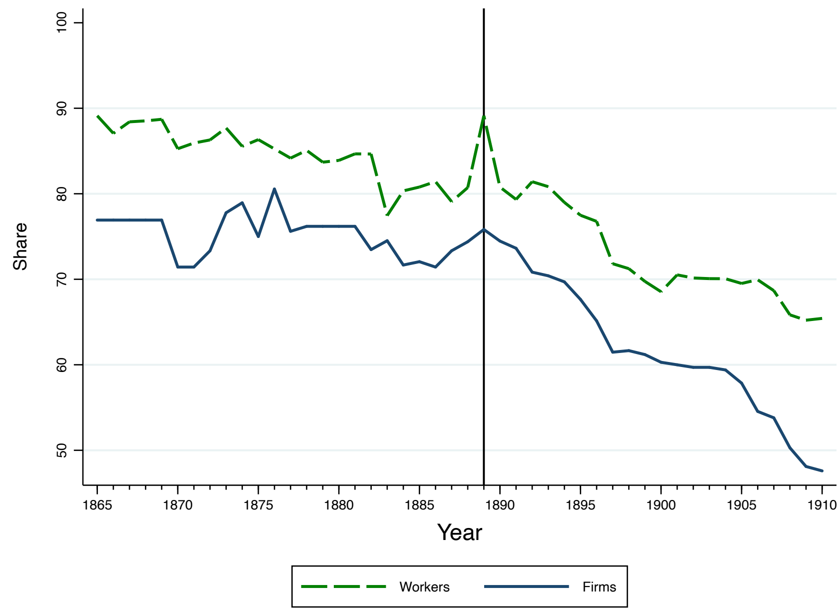
These trends are reflected in Figures 1 and 2. Bombay Presidency is divided into two sub-regions: Greater Bombay consisting of Bombay city and surrounding areas, and Greater Ahmedabad consisting of Ahmedabad and the surrounding areas. Figure 1 shows the dominance of the Greater Bombay region in the late 1880s. After 1890, the Ahmedabad region emerged as a competitor and the share of Greater Bombay within Bombay Presidency declined (as seen in Figure 2).

Figure 1: Location of Textile Mills in India 1888



Source: Statistical Tables Relating to Indian Cotton: Indian Spinning and Weaving Mills. Times of India, Bombay 1889

Figure 2: Cotton Textile Industry Share of Bombay City in Bombay Presidency (1865-1910)



Source: Report of Bombay Millowners' Association, 1925

Hence the assumption of similar levels and trends in outside options across the different communities is reasonable only for the early stage of the industry from 1865-1890 in the Greater Bombay area; accordingly our empirical analysis focuses on this period and location. In Section 5.2 we discuss the subsequent evolution of the industry after 1890.

3 Network-based Model of Entry Dynamics

The model extends Munshi (2011) and is a special case of Dai et al (2018) with a single destination sector.[5] There are successive cohorts of potential (infinitely lived) entrepreneurs (or agents, in short) $t = 1, 2, ...$ who have a once-and-for-all opportunity to enter the downstream industry at date t. Their outside option is to continue pursuing their current occupation, e.g., trading or the upstream stage of the same industry. The agents belong to different communities $c = 1, ..., C$. Individuals within a

[5]The latter paper tests a more detailed version of the same model in the context of growth of privately owned firms in China over the period 1990-2009, using comprehensive firm registration data covering all industries. The richness of the more contemporary Chinese data allows us to test implications of community networks on entry, size, and concentration across sectors and locations, using a different econometric strategy relating firm dynamics to a measure of quality of the community networks. The relative sparseness of the Indian historical data rules out the use of a similar estimation strategy, so we rely instead on inferences based on variation in initial presence of different communities with subsequent entry flows.

community vary in their entrepreneurial ability ω, drawn from a log-uniform distribution whose mean varies across communities. Specifically, log ω is distributed uniformly with mean log ω_c on support $[\log \omega_c - \frac{1}{2}, \log \omega_c + \frac{1}{2}]$. The profit of an agent with individual ability ω in their outside option is $\Pi_c \omega^\sigma$ where $\sigma \in (0, 1)$.

Communities can differ in levels of underlying characteristics such as access to capital (borrowing cost r_c), ability (ω_c) and outside options (Π_c), which are assumed to be fixed over time. The historical evidence described in previous sections suggests this is a reasonable assumption for the early stage of the cotton textile industry in the Greater Bombay region.

Entry decisions are irreversible: once a cohort t agent enters the downstream industry he stays there at all dates $t' > t$. It will be in the interest of every entrant to stay every period thereafter, so this assumption is not restrictive. Entrepreneurs are myopic and selfish: each cohort t agent decides whether to enter based on a comparison of his own profit upon entering the industry at date t, with his outside option. Extending the model to accommodate foresight makes it more complicated, while reinforcing further the network effects. Hence myopia is a useful simplifying assumption. Similarly, extension to incorporate some altruism towards fellow community members will also reinforce the network effects at the cost of complicating the analysis considerably.

Network effects in the downstream industry are represented by dependence of the total factor productivity (TFP) of any cohort t entrepreneur from community c on the presence of incumbents from the same community from all preceding cohorts $t-1$, denoted by n^c_{t-1}. A community c entrepreneur of ability ω will have a production function at date t:

$$y^c_t = A^c_t \omega^{1-\alpha} K^\alpha \tag{1}$$

where K denotes capital size chosen by the entrepreneur, $\alpha \in (0, 1)$ represents diminishing returns to capital, and the 'community-based TFP' (CTFP) is represented by

$$A^c_t = A_0 \exp\{\theta n^c_{t-1}\} \tag{2}$$

The underlying assumption is that incumbents from the same community share knowhow and help one another, to overcome the absence of government support and market institutions. This is supported by social sanctions involving suspension of social interactions by the community with deviants. Fresh entrants can benefit from these forms of intra-community cooperation; they do not contribute in the period that they enter but do contribute thereafter. Exchanges of different community members complement each other, giving rise to the exponential form (2) involving the size of the network. For instance, each entrepreneur relies on a given level of informal help $h > 0$ provided by other members to raise productivity or market access: $A_t^c = A_0(1+\gamma h)^{n_{t-1}^c}$, which reduces to expression (2) if we define $\theta \equiv \log(1+\gamma h)$. θ represents the strength of the network effect, manifested by a productivity spillover among network members. This formulation of spillovers differs from standard specifications of agglomeration spillovers in the economic geography literature, as the domain of the spillovers at a given location is restricted to entrepreneurs from the same community. In contrast, agglomeration effects are spillovers are location-specific but do not vary by community origins of the entrepreneurs involved.

It could be argued that active collaboration is actually limited to social groups (based on family ties, physical proximity or other sources of social association) which are smaller than the entire community. In that case the definition of network size for any given agent should be the number of incumbents from the same social group. If there are k social groups within the community, $m_{t-1}^c \equiv \frac{n_{t-1}^c}{k}$ denotes the (average) number of incumbents from each group, and θ' denotes the strength of ties within the group, formulation (2) would be replaced by

$$A_t^c = A_0 \exp\{\theta' m_{t-1}^c\} \tag{3}$$

for a representative agent of community c, which reduces to (2) if we define $\theta = \frac{\theta'}{k}$.

The product price q_t at date t is the same across all communities. Each entrepreneur and community takes this price as given; we therefore abstract from collusive behavior. Conditional on entering, an entrepreneur of ability ω will enter with capital size $K_t^c(\omega)$ which maximizes profit $\pi_t \equiv q_t A_t^c \omega^{1-\alpha} K - r_c K$. Hence realized profits and capital size in the downstream industry, conditional on entry at t by a potential cohort-t entrepreneur with individual ability ω and belonging to a community with an incumbent network size of n_{t-1}^c, are given by

$$\log \pi_t^c(\omega) = \log \omega + \frac{1}{1-\alpha}[\theta n_{t-1}^c + \log q_t] - \frac{\alpha}{1-\alpha}\log r_c + \log \mu + \frac{1}{1-\alpha}\log A_0 \tag{4}$$

$$\log K_t^c(\omega) = \log \omega + \frac{1}{1-\alpha}[\theta n_{t-1}^c + \log q_t] - \frac{1}{1-\alpha}\log r_c + \frac{1}{1-\alpha}[\log A_0 + \log \alpha] \tag{5}$$

where μ denotes $[\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}]$.

Entry is endogenous and incorporates a pattern of selection (analogous to the standard Roy model) based on entrepreneurial ability and network size as follows. A cohort t agent will decide to enter if $\log \pi_{t-1}^c(\omega) > \log \Pi_c + \sigma \log \omega$, i.e. if his individual ability ω exceeds the threshold $\underline{\omega}_t^c$ given by

$$\log \underline{\omega}_t^c = \frac{1}{1-\sigma}[\log \Gamma_c - \frac{1}{1-\alpha}\{\log q_t + \theta n_{t-1}^c\}] \tag{6}$$

where $\log \Gamma_c \equiv \Pi_c + \frac{\alpha}{1-\alpha} \log r_c - \log \mu - \frac{1}{1-\alpha} \log A_0$.

Owing to yearly inflows of fresh cohorts, network size n^c_{t-1} grows over time, while other community characteristics are fixed: hence the ability threshold (6) declines over time (assuming the product price is stationary). The implied dynamics of entry flows will then depend on whether the ability threshold is in the interior of the support of the ability distribution, which will be the case in early stages of the development of the industry. During the early stage, the threshold will be falling in t, implying that the entry flows will be rising. Once this early stage is over, there is no further scope for entry flows to rise: the threshold drops below the lower endpoint of the support, and *all* entrepreneurs in subsequent cohorts will enter.

We thus define the *early stage* of the entry process to be when the entry threshold is interior, i.e. the network size is still small:

$$n^c_{t-1} < \frac{1}{\theta}[(1-\alpha)\log \Gamma_c - \log q_t] \tag{7}$$

while the *mature stage* is characterized by the reversal of this inequality. The actual date of transition will of course depend on the realization of the price shocks, and is thus random. Moreover, the process can flip back and forth between the two stages. These complications will not arise if the product price does not fluctuate over time, which we abstract from to simplify the exposition.

The resulting entry flows are as follows. During the early stage, the proportion of each cohort that enters is:

$$e^c_t = \frac{1}{(1-\alpha)(1-\sigma)}[\theta n^c_{t-1} + \log q_t] + \delta_c < 1 \tag{8}$$

while during the late stage:

$$e^c_t = 1 \tag{9}$$

where $\delta_c \equiv \frac{1}{1-\sigma}\Gamma_c + \log \omega_c + \frac{1}{2}$ is a 'composite' community characteristic, which depends on mean ability, outside options and capital access of community c members.

The evolution of network size during the early stage is thus given by

$$n^c_t = n^c_{t-1} + e^c_t = M n^c_{t-1} + P \log q_t + \delta_c \tag{10}$$

where $M \equiv 1 + \frac{\theta}{(1-\alpha)(1-\sigma)}$ and $P \equiv \frac{1}{(1-\alpha)(1-\sigma)}$. M is the *network multiplier parameter* which exceeds 1 in the presence of network effects, and equals 1 otherwise. This parameter represents the incremental effect of higher network size in any cohort on the network size in the succeeding cohort.

In the mature stage, the dynamic of network size is

$$n_t^c = n_{t-1}^c + 1 = n_T^c + t - T \tag{11}$$

if T denotes the last date of transition into the mature stage.

We summarize the predicted network dynamic in the following Proposition.

Proposition 1 *During the early stage (assuming (7) holds at dates $1, \ldots, t$), the evolution of network size of community c given initial cohort size n_0^c is given by*

$$n_t^c = M^t[n_0^c + \frac{\delta_c}{M-1}] - \frac{\delta_c}{M-1} + P \sum_{k=0}^{t-1} M^k \log q_{t-k} \tag{12}$$

in the presence of network effects ($M > 1$), and

$$n_t^c = n_0^c + \delta_c.t + P \sum_{k=0}^{t-1} \log q_{t-k} \tag{13}$$

when they are absent ($M = 1$).

In the mature stage, network size follows (11) instead.

Network effects result in greater sensitivity of network size in later cohorts to initial network size during the early stage of the development of the industry: $\frac{\partial n_t^c}{\partial n_0^c} = M^t$ which is rising in t when $M > 1$. A network multiplier operates across successive cohorts: one more member raises the number of members at the next cohort by M, the one after that by M^2 and so on. Initial differences in network size across communities will be progressively magnified across time, a pattern of divergence that grows exponentially over time. In the absence of any network effect, this divergence can grow, if at all (i.e if the community characteristic δ_c is positively correlated with initial network size n_0^c), at a linear rate. In particular, the (community-specific) time trend for network size will be linear in the absence of network effects, and exponential in the presence of network effects.

During the mature stage, this difference disappears, as time trends become linear even in the presence of network effects.

Proposition 1 thus generates a strategy of testing for network effects, using the early stage of the development of the industry, under the assumption that relevant community characteristics (ability, access to capital or outside options) summarized by δ_c are fixed over time. Taking a quadratic approximation for $M^t = 1 + t\zeta + t^2\frac{\zeta^2}{2}$, (12) generates the following regression specification for dynamics of network size

$$n_t^c = [1 + \zeta t + \frac{\zeta^2}{2}t^2]n_0^c + \frac{1}{M-1}[\zeta t + \frac{\zeta^2}{2}t^2]\delta_c + P\sum_{k=0}^{t-1}(1 + \zeta k + \frac{\zeta^2}{2}k^2)q_{t-k} \tag{14}$$

in the presence of network effects ($\zeta > 0$). When network effects are absent, the regression is given by (13) instead, where community-specific quadratic time trends do not appear. Hence the significance of community-specific quadratic time trends, interacted either with initial network size n_0^c or community characteristic δ_c, indicates the presence of network effects. Since initial network size and δ_c are perfectly collinear, we cannot identify their effects separately. But this is unnecessary to show evidence of a network effect, i.e,, that ζ is positive. The significance of quadratic time trends interacted with either n_0^c or δ_c is sufficient to infer the existence of a network effect. Intuitively, they represent a pattern of divergence between the size of different communities among incumbents that is growing faster than can be represented by differential linear time trends.

3.1 Misallocation, Adverse Selection and Entering Capital Size

While community-based cooperation facilitates entry and thereby hastens the pace of industrialization, its scope is restricted to members of specific communities and excludes others. This would be reflected in growing divergence in the presence of different communities among incumbents during the early stage of development. This is in marked contrast to market or state-sponsored industrialization patterns whose scope is not restricted to particular social communities. As we shall now explain, the growing imbalance in community composition of entrepreneurs will be accompanied by a form of misallocation, where incumbents from favored communities with larger networks will tend to be of lower productivity. This reflects a pattern of adverse selection generated by the network-based mechanism, during the early stage of development. This pattern helps distinguish the network based model from competing explanations of non-linear amplification of differences in initial community presence, based on a positive correlation between levels and trends in community-specific ability or wealth (which would predict instead rising productivity of entrants from communities with higher initial presence).

Observe that the log of the productivity P_t^{mc} of a marginal entrant from cohort t of community c equals

$$\log P_t^{mc} = \log A_t^c + (1-\alpha)\log \underline{\omega}_t^c = \log A_0 - \frac{\theta\sigma}{1-\sigma}n_{t-1}^c - \frac{1}{1-\sigma}\log q_t + \frac{1-\alpha}{1-\sigma}\log \Gamma_c \tag{15}$$

during the early stage. This implies that marginal entrants (among any given cohort) from communities with a higher network size have lower productivity. Hence average productivity would rise if it were possible to replace the marginal entrant from a community with a larger network, by a marginal non-entrant from a different community with a smaller network (i.e,, who did not actually enter but was almost indifferent between entering and not). This represents a misallocation of talent in the economy.

The misallocation appears even when comparing the average productivity of **all** entrants (rather than marginal entrants) across communities of varying network size if the parameter σ (elasticity of the outside option with respect to ability) exceeds $\frac{1}{2}$, since the log of the productivity P_t^{ac} of an entrant from cohort t of community c equals

$$\begin{aligned}\log P_t^{ac} &= \log A_t^c + \frac{1-\alpha}{2}[\log \omega_c + \frac{1}{2} + \log \underline{\omega}_t^c] \\ &= \log A_0 + \frac{1-2\sigma}{1-\sigma}\theta n_{t-1}^c + \frac{1-\alpha}{2}[\log \omega_c + \frac{1}{2} + \frac{1}{2(1-\sigma)}\log \Gamma_c] - \frac{1}{2(1-\sigma)}\log q_t\end{aligned}$$

However, in the opposite case where $\sigma < \frac{1}{2}$, entrants into larger network communities are of higher average productivity. Hence adverse selection for average entrants is a stronger form of misallocation which may or may not arise.

How can we test for such forms of adverse selection? Note the close connection between entering capital size and productivity P of an entrant:

$$K(P; r_c) = [\frac{\alpha P}{r_c}]^{\frac{1}{1-\alpha}} \tag{16}$$

If access to capital varies across communities, this makes it difficult to test for misallocation directly by comparing entering capital sizes in different communities. Instead we can test for adverse selection by comparing time trends in entrant capital size, since according to the model these reflect changes in network size while capital access is unchanging over time within any community. To elaborate on this, observe that among the entrants within a given community, those entering with a smaller capital size are of lower productivity. Hence we can identify the marginal entrant in a community-cohort pair, by the entrepreneur with the smallest entering capital size. The model predicts the smallest entering capital size from a given community c at any given date t, is decreasing in the network size n_{t-1}^c, and therefore in the determinants of this size (which has been provided in Proposition 1). In particular, it predicts that *the smallest entering capital size in a given community will be decreasing in the entry date* t*, that this declining time trend will be steeper for communities with higher initial presence* n_0^c*, and finally that this tendency will accelerate over time (owing to the accelerating growth of network size).* We can examine this for the average entrants as well, a stronger form of misallocation.

4 Cotton Industry: Data and Descriptive Statistics

Our data come from business directories, which list all registered firms in the industry in the region and the names of directors and partners. The firm level data used in the the paper comes from the Bombay Almanac, Times of India Calendar and Directory and Thacker's Indian Directory; see the Appendix for further details. This data has been digitised by us and to the best of our knowledge is being used for the first time. All subsequent figures and tables are based on this data.

Each entrepreneur is either a director or a partner in a listed firm. We construct a data set at the level of the entrepreneur using names of partners and directors. The data include all entrepreneurs in registered firms in Bombay Presidency (Greater Bombay and Greater Ahmedabad). We observe them at the point of entry, and track their previous and subsequent trajectories between 1860 and 1910. The data includes some characteristics of the firm that the entrepreneur is listed in: the year of establishment, its location and in many cases paid-up capital. From the names of the entrepreneurs, we can trace the community they belong to. Following the discussion in the preceding section, we classify communities principally on the basis of religion and caste. Within Hindus, we distinguish between the Bhatias and Vanias. For reasons explained earlier, the term 'Vania' will include Hindu and Jain Vania merchants. We construct a panel data set of entrepreneurs by community. The communities are Parsi, Bhatia, Vania, Baghdadi Jews and Khoja Muslim. The reference community is European, mainly British. As discussed in the introduction, our main analysis will focus on the early phase of development of the industry. As Figure 1 shows, until 1890 most firms were situated in Greater Bombay, so it is unlikely that we are confounding network effects with agglomeration effects.

Figure 3: Downstream sector (1860-1890)

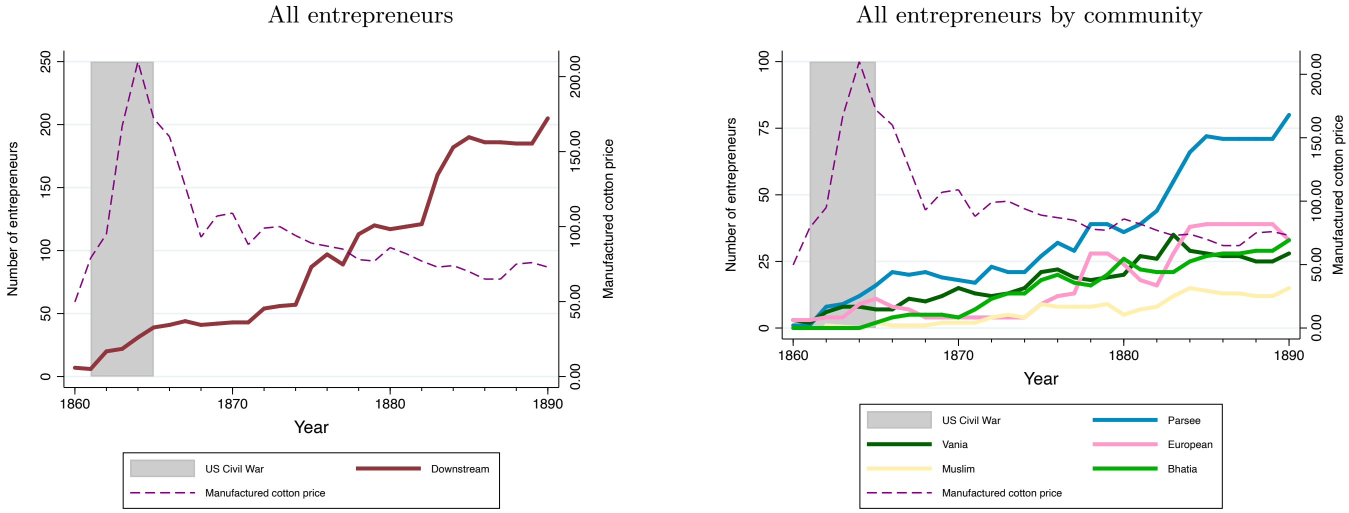


Figure 3 shows the evolution of the stock of total number of entrepreneurs in a given year between 1860 and 1890, and its breakdown between different communities. As we are measuring stocks in a given year, we take into account exit as well as entry. The graph shows a steady growth in the number of entrepreneurs entering in downstream industry between 1860-90. There is an upward spurt during 1875-80, followed by periodic spurts during the 1880s. We also observe the dominance of the Parsis at the initial stage and a pattern of non-linear divergence in the presence of different communities. Observe also that rising entry is not accompanied by rising prices of manufactured cotton.

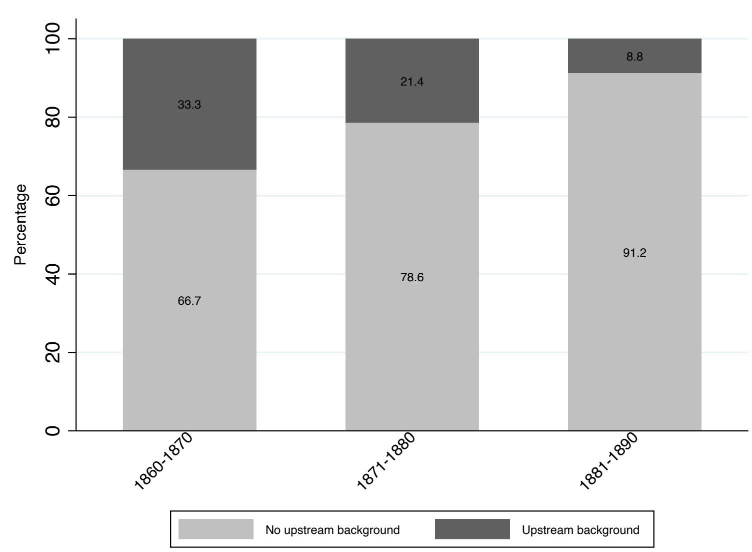
Our data includes paid up capital of firms. The role of prior wealth accumulation in the emergence of industrial entrepreneurship has been stressed by Marxian theories of 'primitive accumulation' (see Marx 1887 (Chapter 26) and Harvey 2005) as well as more recent theories of occupational choice in the presence of credit constraints and lumpiness of industrial capital (Banerjee and Newman 1993). The capital required to set up a textile mill was substantial. Table 5 shows the capital needed to set up a textile firm. It shows that capital per firm in downstream industry was twice as large as the capital in upstream industry during 1860-90. The upstream industry comprised of mainly ginning and baling of raw cotton, and was technologically far less sophisticated than the downstream mills. For these reasons, we will focus on entry into the downstream industry, and treat the upstream industry as a related pre-industrial trading activity. The substantial capital required to set up a firm in the downstream may explain why growth in entry was gradual and took almost a decade after the end of the Civil War to accelerate further. The opening of the stock market in 1875 may have facilitated the transition to a faster growth phase.

Table 5: Capital per firm in Rupees

	Upstream		Downstream	
Period	*mean*	*median*	*mean*	*median*
1860-1890	578,990	400,000	1,206,293	1,000,000
1891-1910	345,283	225,000	1,106,615	943,500

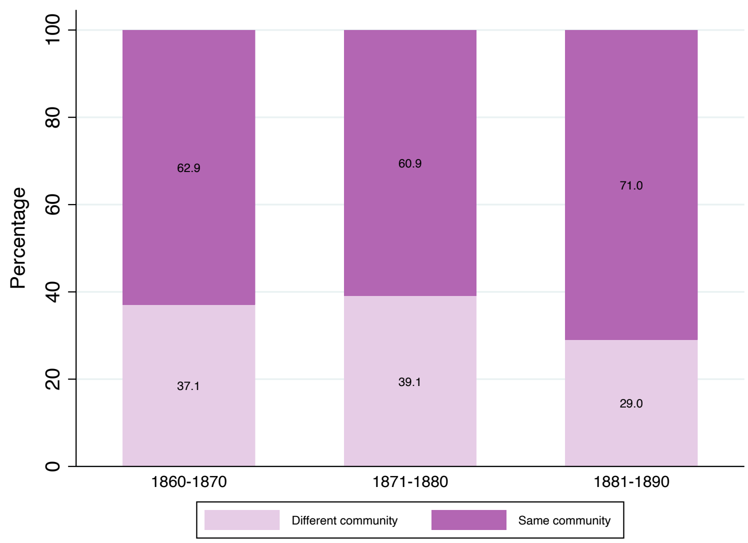
The role of prior experience in pre-industrial trade in entrepreneurship has been highlighted in the recent 'enterprise map' studies from some contemporary African countries, which suggest pre-existing upstream experience is an important determinant of entry into downstream industry (Sutton and Kellow 2010; Sutton and Kpentey 2012). To check the relevance of this in the current context, we calculate the proportion of downstream entrants who had earlier participated in the upstream industry. Figure 4 shows that less than 35% of the downstream entrants had prior upstream experience. This proportion declined further over the subsequent periods. By 1890 less than 10% of the entrants had prior upstream experience.

Figure 4: Upstream experience in downstream entry



Next, we examine clustering by community within firms. Figure 5 shows over 60% of all entrepreneurs within any firm belonged to the same community. Moreover, this proportion rose to 71% during the decade 1880-90 and suggests continued importance of community links throughout successive decades, despite the growth of the stock market.

Figure 5: Community concentration in firms at entry

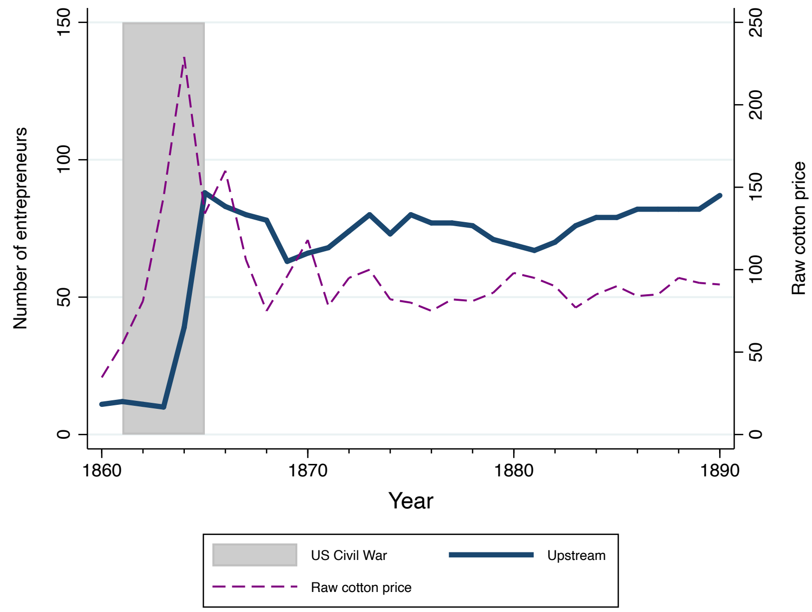


We construct a second panel of paid up capital per entrepreneur at the community level, using firms at entry point only. The variable output per entrepreneur is constructed by using firm capital at the point of entry and dividing this by the total number of entrepreneurs listed as directors or partners in the firm. This is used in the next section to test for misallocation. Unfortunately reporting of capital size during post-entry years was not systematic, so we could not construct a reliable data set on post-entry growth rates of firm size.

5 Empirical Results for Cotton Entry

5.1 The Early Phase 1866-1890

Figure 6: Upstream entrepreneurs (1860-1890)



The Civil War witnessed a boom in the upstream industry, resulting from a spike in the world price of raw cotton. As Figure 6 shows, there was a sharp upward increase in the number of entrepreneurs and firms in the upstream industry during the Civil War period. Once the War ended, the price of raw cotton exports fell sharply, and the number of entrepreneurs in the upstream sector plateaued. In contrast, the number of entrepreneurs in the dowstream sector accelerated after 1875 (as seen in Figure 3). The downstream industry followed a different trajectory from the upstream. A likely reason is that the downstream produced cotton yarn or cloth mainly for the domestic market rather than the international market and the upstream produced mainly for export until 1865.

The regression analysis focuses on the early period in the development of the downstream industry, following the US Civil War years. We test the model developed in Section 3 using data for the downstream industry over 1866-1890. The initial presence of each community (date 0 in the model) is represented by number of entrepreneurs already present in the downstream industry in 1865. We stop in 1890 for the reasons explained in previous sections: (a) entry from communities in Bombay stabilized as the 'early stage' came to an end, and (b) the model would fail to capture the experience of the Vanias whose outside options in moneylending contracted sharply in the 1880s, inducing them to enter in large numbers into the downstream cotton industry from the 1890 onwards in a new location. The regression also tests the role of prior upstream experience and opportunity to accumulate profit during the Civil War years, measured by 1865 upstream community presence.

Table 6 shows the presence of different communities in the upstream and downstream industry during the Civil War years. There were 38 entrepreneurs downstream in 1865, dominated by Parsis, followed by European and Vania communities, and also featuring a small number of Muslims, Bhatias and other Indian communities.

Table 6: Stock of entrepreneurs at the end of the US Civil War (1865)

Community	Upstream	Downstream
Bhatia	9	2
European	41	11
Vania	9	7
Muslim	2	2
Parsi	26	16

Column 1 of Table 7 shows results for the following regression specification based on (14), run for the period 1866-1890:

$$n_t^c = \phi_c * t^2 + \delta_c * t + \gamma_c + \nu_t + \epsilon_t^c \tag{17}$$

where $\phi_c, \delta_c, \gamma_c$ denote community c dummies, n_t^c denotes number of incumbent entrepreneurs from community c in year t, and $t = 1$ corresponds to year 1866. ν_t is a year dummy representing (current and lagged) price or other industry level shocks, and ϵ_t^c is an error term. This is a particular version of equation (14) predicted by the network model, in which the initial network size is subsumed within the community dummies interacted with linear and quadratic time trends. In the absence of a network effect, the community specific quadratic time trend dummies ϕ_c should all equal zero. Since the estimated coefficients are not of intrinsic interest, we report only the F-statistic for this test. Column 1 shows the null hypothesis of a zero network effect is rejected with a p-value less than .01.

To confirm the pattern of divergence predicted by the network model, and obtain an estimate of the magnitude of the network effect, Column 2 replaces the community dummies by their levels of initial presence n_0^c in 1866, and also includes upstream presence n_0^{uc} in that year:

$$n_t^c = \chi n_0^c * t^2 + \delta n_0^c * t + \gamma n_0^c + \eta n_0^{cu} + \nu_t + \epsilon_t^c \tag{18}$$

Table 7: Regression: Stock of entrepreneurs at community level (1866-1890)

VARIABLES	(1)	(2)	(3)	(4)	(5)
n_0^c downstream $* t^2$		0.008***	0.008***	0.008***	0.007***
		(0.002)	(0.002)	(0.001)	(0.001)
n_0^c downstream $* t$		-0.082	-0.082**		
		(0.058)	(0.039)		
n_0^c downstream		1.673***			
		(0.252)			
n_0^c upstream		-0.322***			
		(0.054)			
L1 Log cotton price ratio					5.947
					(4.223)
L2 Log cotton price ratio					6.609*
					(3.647)
Year dummies	Y	Y	Y	Y	N
Community dummies	Y	N	Y	Y	Y
Community $* t$	Y	N	N	Y	Y
Community $* t^2$	N	N	N	N	N
F test Community $* t^2$	21.44				
Prob >F	0.00				
Observations	125	125	125	125	125
R-squared	0.969	0.874	0.944	0.965	0.940

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Intercept not reported

This specification requires the stronger assumption that levels of initial presence n_0^c were uncorrelated with unobserved community characteristics. Since this assumption is strong, the estimates reported are purely illustrative. The estimate of χ is 0.008, significant at the 1% level, while that of δ is negative and statistically insignificant. The estimated coefficient η of initial upstream presence is negative and significant, reinforcing the evidence against hypotheses stressing the role of prior accumulation of skills or wealth in the upstream sector.

Columns 3,4 and 5 show that the estimated magnitude and significance of χ in Column 2 is ro-

Table 8: Regression: Stock of entrepreneurs at community level, assuming no exits (1866-1890)

VARIABLES	(1)	(2)	(3)	(4)	(5)
n_0^c downstream $* t^2$		0.009***	0.009***	0.009**	0.014***
		(0.004)	(0.002)	(0.004)	(0.001)
n_0^c downstream $* t$		-0.023	-0.023		
		(0.081)	(0.059)		
n_0^c downstream		1.926***			
		(0.352)			
n_0^c upstream		-0.145**			
		(0.064)			
L1 Log cotton price ratio					-1.581
					(5.570)
L2 Log cotton price ratio					9.833**
					(4.699)
Year dummies	Y	Y	Y	Y	N
Community dummies	Y	N	Y	Y	Y
Community $* t$	Y	N	N	Y	Y
Community $* t^2$	Y	N	N	N	N
F test Community $* t^2$	14.09				
Prob >F	0.00				
Observations	125	125	125	125	125
R-squared	0.992	0.936	0.978	0.990	0.973

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Intercept not reported

bust under alternative specifications of (18). Column 3 replaces the initial downstream and upstream presence by community dummies; Column 4 additionally replaces the interaction of linear time trend with n_0^c by δ_c. Finally, Column 5 replaces year dummies in Column 4 by lagged ratio of manufactured goods price and raw cotton price. The two period lagged price ratio is marginally significant, while the magnitude and significance of the estimated χ increases.

The common estimate of χ across Columns 2-4 corresponds to a value of the network effect ζ (in equation (14)) of approximately 0.13.[6] In other words, one extra entrant from a community in any given year was associated with 0.13 more entrants from the same community in the following year. Using the quadratic approximation for M^t, this cumulates to 2.85 additional entrants fifteen years later, and 5.75 more entrants twenty five years later.[7]

Table 7 used data for the actual stock of incumbents from each community at different dates, which reflects the joint effect of entry and exits. In addition we sometimes observe entrepreneurs listed in some years, drop out for intervening years and then get listed again subsequently. This could reflect errors of omission in the listing of data for some years, or intermittent entry and exit by certain entrepreneurs. However our theory focused entirely on entry decisions, assuming that entrepreneurs that once enter do not ever exit. We now check the extent to which the preceding results continue to hold if we focus in the data on the entry process exclusively. To do this we replace the actual community-year incumbent stock data series by cumulating past flows of entrants from each community. This would have been the stock of incumbents if entrants never exited. Table 8 shows the results of the same regression as Table 7, using this alternative series as the dependent variable. It is evident that the results are almost entirely the same, and thereby unaffected by the pattern of exits or measurement error resulting from omission of some entrepreneurs in the business registers for some years.

Next, we examine evidence for misallocation as predicted by the network model: that entering capital size declines faster among marginal entrants from communities with higher initial presence. To avoid effect of outliers, we use as dependent variable K_t^{mc}, the 10th percentile of the distribution of capital per entrepreneur for firms from community c (i.e., firms for which c constitutes the largest proportion of entrepreneurs) that enter in year t. Combining equations (15, 16) with (18), we obtain

$$\begin{aligned}\log K_t^{mc} &= -\frac{\theta\sigma}{1-\sigma}[\chi n_0^c * (t-1)^2 + \delta n_0^c * t + \gamma n_0^c + \eta n_0^{cu} + \nu_t] \\ &\quad -\frac{1}{1-\sigma}\log q_t + \frac{1-\alpha}{1-\sigma}\log \Gamma_c + \log A_0 \end{aligned} \qquad (19)$$

Hence the regression specification is (if initial network size n_0^c is exogenous):

$$\log K_t^{mc} = \mu_1 n_0^c * t^2 + \mu_2 n_0^c * t + \mu_3 \delta_c + \nu_t' + \epsilon_{ct}' \qquad (20)$$

where $\mu_1 \equiv -\frac{\chi\theta\sigma}{1-\sigma} < 0, \mu_2 \equiv \frac{(2\chi-\delta)\theta\sigma}{1-\sigma}$. If we drop the exogeneity assumption of initial network size, we

[6]This follows from $\frac{\zeta^2}{2} = .008$.

[7]Recall the quadratic approximation is $M^t = 1 + \zeta * t + \frac{\zeta^2}{2} * t^2$.

can replace n_0^c interactions with time trends with corresponding community dummy interactions:

$$\log K_t^{mc} = \phi_{1c} * t^2 + \phi_{2c} * t + \phi_{3c} + \nu_t' + \epsilon_{ct}' \tag{21}$$

where network effects imply the existence of community-specific quadratic and linear time trends.

Table 9 shows these regression results, restricting the sample to community-year pairs where there was positive entry of firms from the community in question. As in the entry regressions, Column 1 corresponds to specification (21), and shows that the community specific quadratic time trends are jointly significant. Column 2 corresponds to specification (20). Column 3 shows the estimated μ_1 when $n_0^c * t$ is replaced by $\delta_c * t$. It is negative in both columns 2 and 3, and significant at the 10% and 5% levels respectively. The estimated μ_2 is positive, but recall this is consistent with the network model (e.g., if $2\chi > \delta$ in (19)). The relative magnitudes of μ_1 and μ_2 indicate that after $t \geq 6$, the negative quadratic effect dominates, so communities with higher initial presence experienced lower trends for all but the first few years. Hence we find evidence consistent with the network model, and against the alternative hypothesis that communities with higher initial presence had faster growing wealth or ability. Column 5 checks directly for the alternative hypothesis using the interaction of initial presence with linear time trends alone, while dropping its interaction with t^2. The linear interaction is slightly negative and insignificant. Hence there is no evidence in favor of the alternative hypothesis.

Table 9: Regression: Log of capital per entrepreneur (10th percentile): 1866-1890

VARIABLES	(1)	(2)	(3)	(4)
n_0^c downstream $* t^2$		-0.002*	-0.004**	
		(0.001)	(0.001)	
n_0^c downstream $* t$		0.062*		-0.001
		(0.030)		(0.009)
Year dummies	Y	Y	Y	Y
Community dummies	Y	Y	Y	Y
Community $* t$	Y	N	Y	N
Community $* t^2$	Y	N	N	N
F test Community $* t^2$	5.02			
Prob >F	0.05			
Observations	31	31	31	31
R-squared	0.9583	0.7367	0.9289	0.6582

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Intercept not reported

Table 10 shows corresponding regressions for the log of capital per entrepreneur K_t^{ac} for the average (rather than marginal) entrants from community c in year t. The results are very similar to those of marginal entrants, thus providing evidence of the stronger form of inter-community misallocation.[8]

[8]When we replace the dependent variable in Tables 9, 10 by log of entering firm size, the coefficients of $n_0^c * t^2$ continues to be negative, though statistically indistinguishable from zero. Hence there is no evidence that entering capital size of firms exhibited a larger upward trend among communities with higher initial presence. This addresses the possible concern that the preceding results were driven by increasing entry flows of entrepreneurs per firm, which dominated higher time trends in entering capital size per firm among communities with higher initial presence.

Table 10: Regression: Log of capital per entrepreneur (Average): 1866-1890

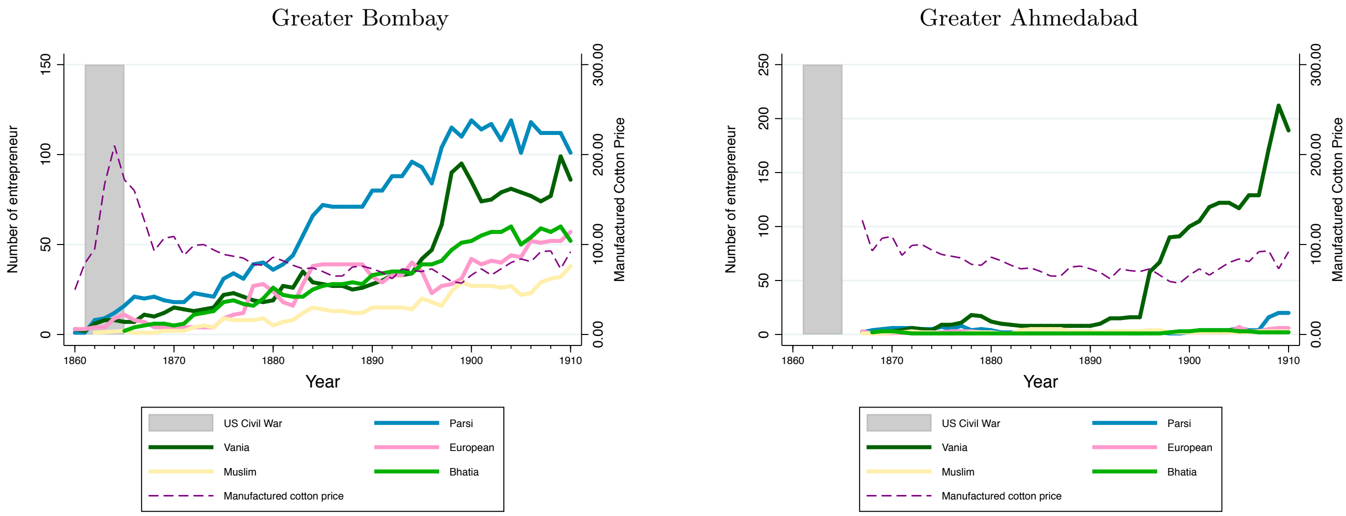
VARIABLES	(1)	(2)	(3)	(4)
n_0^c downstream $* t^2$		-0.002*	-0.003***	
		(0.001)	(0.001)	
n_0^c downstream $* t$		0.063**		-0.001
		(0.028)		(0.009)
Year dummies	Y	Y	Y	Y
Community dummies	Y	Y	Y	Y
Community $* t$	Y	N	Y	N
Community $* t^2$	Y	N	N	N
F test Community $* t^2$	5.02			
Prob >F	0.05			
Observations	31	31	31	31
R-squared	0.9583	0.7367	0.9289	0.6582

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Intercept not reported

5.2 Evidence from the Longer Time Span 1860-1910

Figure 7 extends the time plot of presence of different communities until 1910 in the two locations: Greater Bombay and Greater Ahmedabad. It shows the emergence of the Greater Ahmedabad region as a new location for the industry. There is a steep surge of entry between 1895-1900 is accounted for mainly by the Vania community, who became the largest group in the industry. The process of entry of the Vanias resembles a cascade, consistent with the hypothesis of strong community networks.

Figure 7: Entrepreneur stocks in the downstream cotton textile industry by location (1860-1910)



The entry of Vanias in the Greater Ahmedabad region represented the arrival of a new community. Their connections with the initial stock of entrepreneurs in Greater Bombay was unlikely and the network-based model on the basis of presence in 1865 would therefore be unable to capture this. Other assumptions of the network model would also cease to be valid after the passage of almost three decades, such as the scope for entry rates to continue to rise further. For example, a community may have entered sufficiently to exhaust any further scope for their entry rates to rise. Indeed, Figure 7 shows that the presence of different communities in the Greater Bombay region tended to plateau after 1900, while Vanias continued to enter in the Greater Ahmedabad region. It is therefore not surprising that the significance of the network effects that we found during the phase 1865-90 does not extend when we run the corresponding regressions for network size and entering capital per entrepreneur for the longer period 1865-1910.[9]

However, to confirm our hypothesis that the growth of the industry beyond 1890 still featured the importance of community networks, we check the extent of clustering by community within firms until 1910. In figure 8, we show this by location. Entry into the Greater Ahmedabad region after 1890 exhibits even larger clustering than the Greater Bombay region, with nearly 90% of entrepreneurs from the same community. Moreover, prior upstream experience was even less important for the late entrants than for the previous entrants in both locations (Figure 9).

[9]We do not show the regressions for the longer time span, in order to conserve space.

Figure 8: Community concentration in cotton textile firms at entry by location (1860-1910)

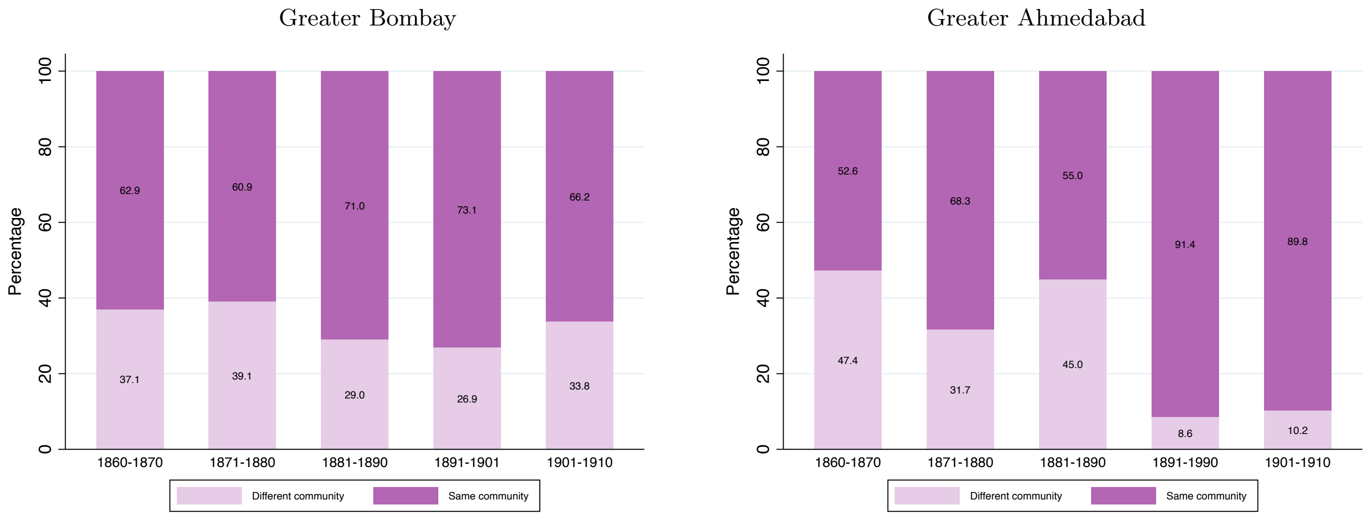
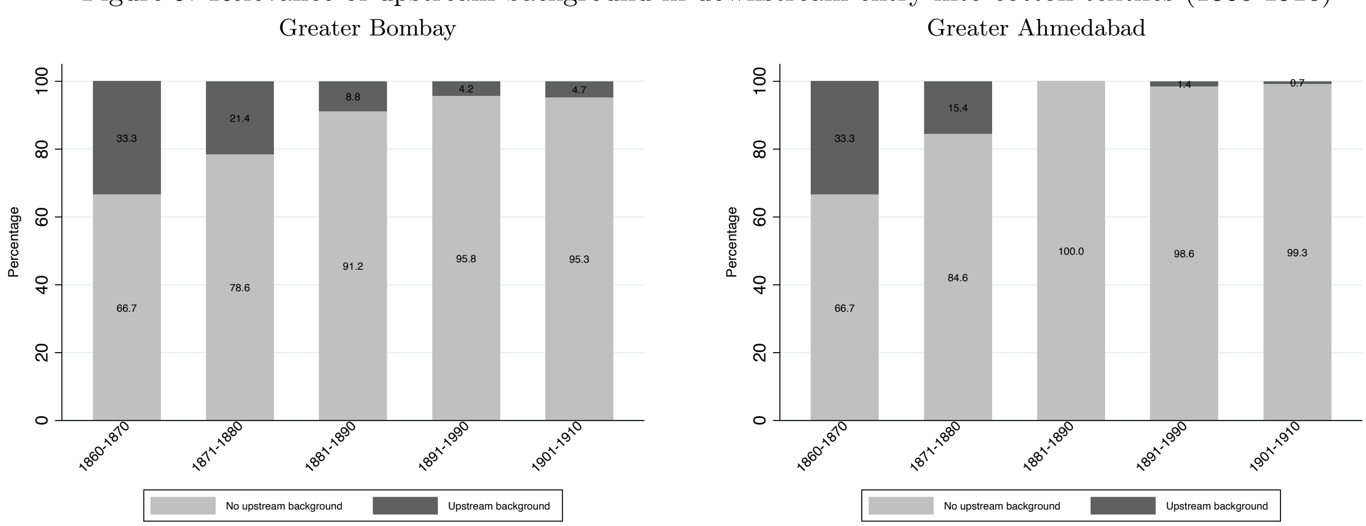


Figure 9: Relevance of upstream background in downstream entry into cotton textiles (1860-1910)



6 Entry Patterns in the Jute Industry

Unlike Bombay, the commercial life of Calcutta was dominated by British firms. While Bombay was the hub of Indian capital, Calcutta was the centre of British firms in tea, coal and jute. Many investors were British, resident in India. The Calcutta industry we focus on is jute, for a number of reasons. Until 1930 this was the most important instance of entry by Indian entrepreneurs after cotton. Jute is also closest to cotton textiles in terms of technology and capital requirements. It developed with

British entrepreneurship around the same time as cotton textiles in Bombay. But unlike cotton, jute was mainly sold in the export markets and the British firms had an advantage in the export trade. In 1866, there were four British firms in jute. From the mid 1870s, the industry expanded rapidly and by 1900, there were 32 British firms, but no Indian firms. The situation changed with the first World War, when Indian entrepreneurship appeared for the first time.

Indian presence in Calcutta's commercial sector had been low compared to Bombay. The Marwari traders, the dominant group among Indian communities in Eastern India, were involved in money lending, trade, brokerage with European companies and speculation (Tripathi 2004, p.166). The futures market in opium, gold specie and later raw jute and hessian was started by the Marwaris and became the focus of speculation. Several Marwaris first created their wealth in the opium trade. Among them, Birla, Hukumchand and Chamaria were the key players (Timberg 1978, p160-61). While the export trade in raw jute and jute textiles had been controlled by the Europeans before the war, from 1914 Marwaris traders Birla and Hukumchand became involved in this trade (Bagchi 1994, p.179).

British Managing Agency Houses (which managed and controlled firms across various industries, including jute) had Marwaris as brokers. The close contact of the Marwari families with British firms opened up a channel of entry into the industry. The first World War was a demand shock for the jute industry, just as the American civil war had been in the case of cotton half a century earlier. As demand for jute goods increased, raw jute prices rose; share prices of jute firms rose 8-10 times (Goswami 1985). Many British shareholders sold shares to their Marwari contacts. The Marwaris also lent short term capital to British firms against a collateral of shares. Both interactions allowed the Marwaris to acquire shares in British firms and get elected on the boards of British Companies. In 1918, of the 114 directors in British owned jute firms, only 3 were Marwaris. By 1924, of the 46 British firms, 19 had Marwari directors on their boards (Goswami 1985). Two Indian firms were set up in 1918 by Birla and Hukumchand. While Hukumchand's firm was primarily self-financed, capital for Birla's firm was raised more widely (Timberg 1978, p171). As in cotton, the initial stock of entrepreneurs opened up the way for further entry, which continued until the beginning of the Great Depression. Unlike cotton textiles, the entry of Indian traders into the jute industry happened both through acquisition of shares in British firms and starting new firms.

Figure 10 shows entry into the downstream jute industry from 1914 to 1930. Until 1918, only a handful of Indian entrepreneurs from three communities were present: Marwaris, Baghdadi Jews and Bengalis. After 1918, there was a steep rise in Marwari entry.

Figure 10: Entrepreneur stocks in the Jute downstream sector (1914-1930)

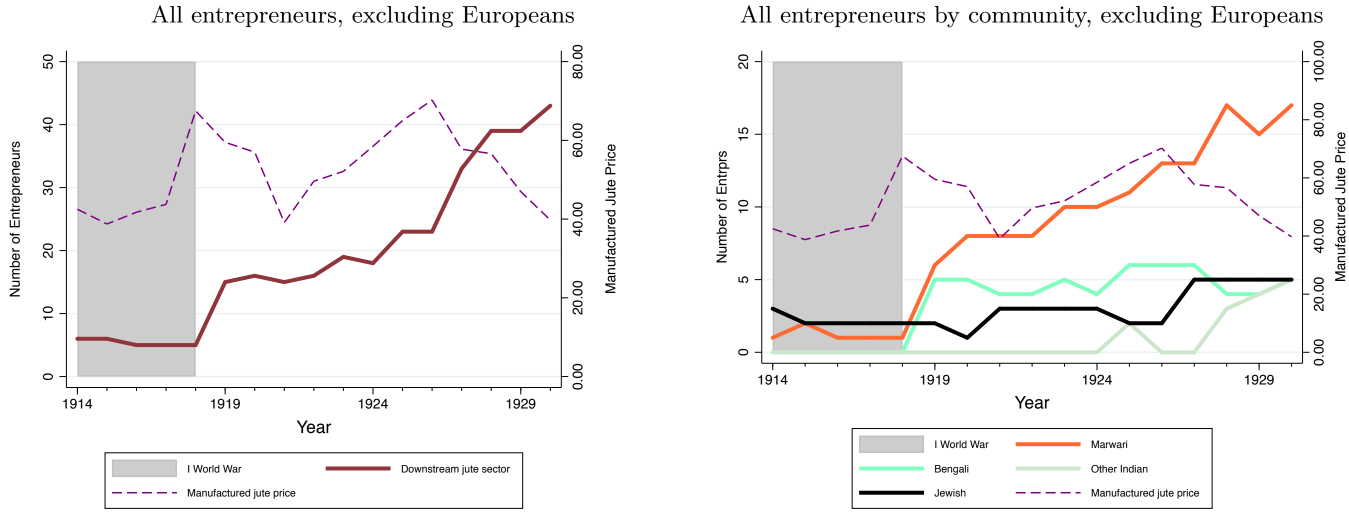


Figure 11: Relevance of upstream background in downstream entry in jute

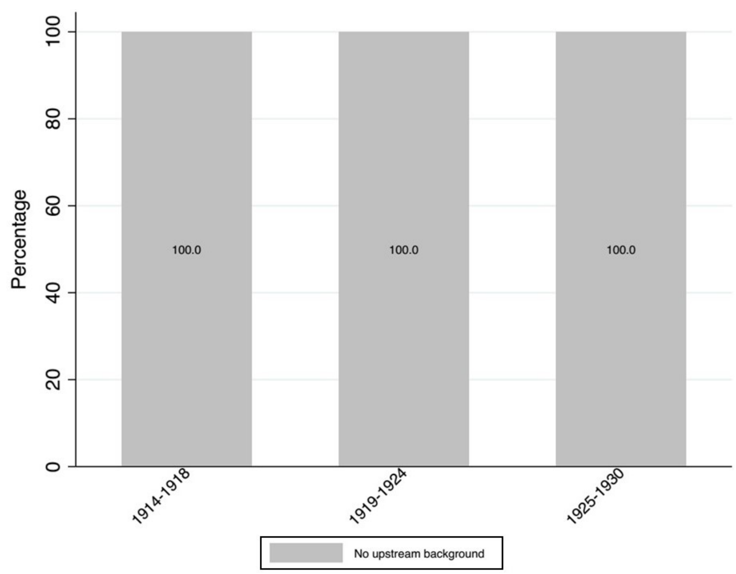


Figure 12: Community concentration in downstream jute firms at entry (1914-1930)

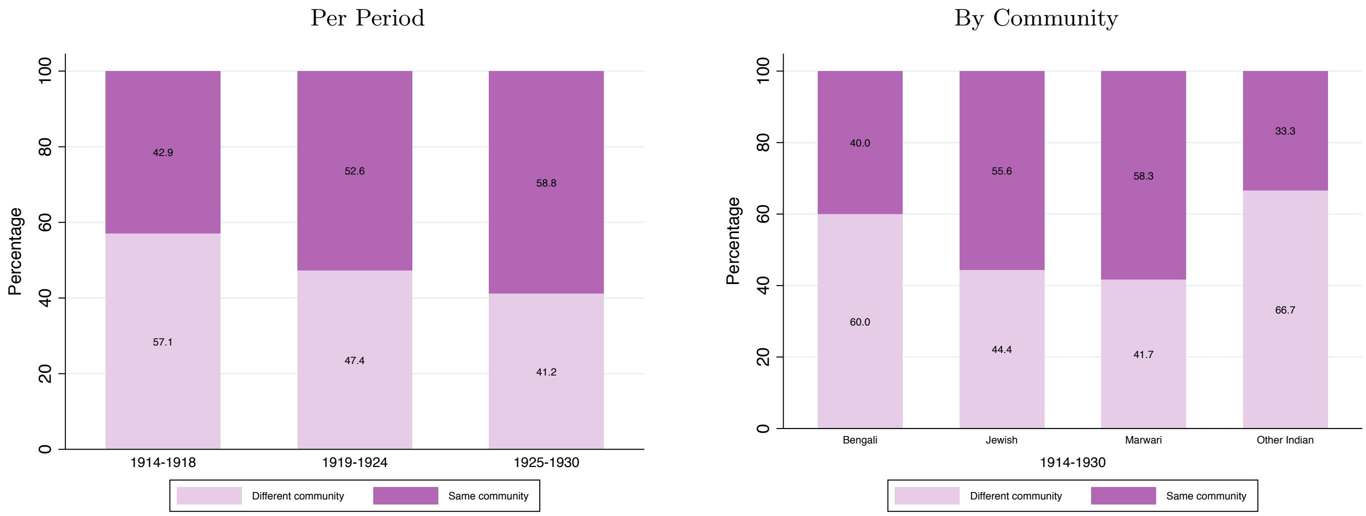


Figure 11 shows that Indian entrants did not have upstream experience. Analogous to the cotton industry, we define jute baling as the upstream sector. Figure 12 shows the extent of clustering by communities within firms. More than half the partners within the average firm belonged to the same community. The clustering was the highest (58%) among the Marwaris, the dominant Indian community. In this industry, the extent of clustering was lower than that observed in the early phases of the cotton textile industry. This is perhaps explained by the different path of entry compared to cotton textiles. Many Marwari entrepreneurs entered by acquiring shares in existing firms from British investors initially.

Finally, we test the model of network-based dynamics by using a regression specification analogous to that used for the cotton industry in Table 11. The results are very similar to those for the early stage of Indian entrepreneurship in downstream cotton in Bombay. The short time span for the jute data, however, do not allow us to estimate the corresponding regressions for capital size of entrants.

Table 11: Stock of entrepreneurs at community level in downstream jute (1919-1930)

VARIABLES	(1)	(2)	(3)	(4)
n_0^c downstream $* t^2$		0.005***	0.005***	0.005***
		(0.002)	(0.002)	(0.002)
n_0^c downstream $* t$		-0.077***	-0.077***	
		(0.023)	(0.023)	
n_0^c downstream		1.129***		
		(0.071)		
n_0^c upstream		0.614***		
		(0.081)		
Year dummies	Y	Y	Y	Y
Community dummies	Y	N	Y	Y
Community $* t$	Y	N	N	Y
Community $* t^2$	Y	N	N	N
F test Community $* t^2$	2.80			
Prob >F	0.04			
Observations	60	60	60	60
R-squared	0.995	0.988	0.993	0.995

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Intercept not reported

7 Concluding Comments

This paper has presented evidence from colonial India on the role played by community networks in early stages of development of downstream cotton and jute industries, when supporting market institutions and state policy were missing. We have put together data on individual entrepreneurs in the cotton and jute industry, thus filling a gap in the literature (e.g., Bagchi (1994, p. 185) states: "No detailed figures on the distributions of directorships among various Indian communities are available"). We use this data to show that community connections were important in early entry. Entrants tended to join firms dominated by entrepreneurs from the same community. Presence of communities in the early stages predict presence several decades later, consistent with the hypothesis that help provided by early incumbents facilitated entry of entrepreneurs from the same community, controlling for fixed differences in unobserved community characteristics, varying prices and other time varying shocks in the industry.

We also found evidence of misallocation of talent resulting from the reliance on community networks, thereby ruling out alternative explanations of observed entry dynamics based on faster growth in education or wealth of communities with higher initial presence. Entrants in communities with higher initial and subsequent presence tended to enter with lower capital sizes, indicating they were of lower ability compared to potential entrants from other communities. Owing to this it is difficult to make any inferences concerning the aggregate welfare impact of the network effects. On the one hand, the help provided by incumbents from the same community encouraged entry, thereby alleviating the underinvestment associated with lack of market institutions or infant industry protection from the state. This is likely to have constituted a first order positive welfare effect, which would have to be set against the welfare losses resulting from the misallocation due to lack of uniformity in networks across communities.

The results in this paper raise further interesting questions, pertaining to the role of community networks in later stages of evolution of these industries. After 1890, the cotton industry witnessed emergence of new locations and new communities. Our empirical methodology could not be applied to draw any clear inferences regarding the role of networks at these later stages. A related question concerns the causes of delayed entry by specific communities in specific regions. In the case of cotton, the later development of the Greater Ahmedabad region, may be explained by the disappearance of the primary pre-industrial occupation (moneylending) of the relevant community (Hindu and Jain Vanias) following the Deccan riots in the 1870s. In similar vein, delayed entry into industrialization in South India by the community of Chettiars (after 1930s) is possibly explained by the decline of their moneylending activities in Burma in the aftermath of the Great Depression. Conversely, the early lead of Greater Bombay may have owed to the earlier disappearance of profitable trading opportunities in opium and cotton enjoyed by Western Indian business communities. The lack of pre-industrial accumulation of wealth or skill in related trading sectors seems unlikely to explain delayed industrialization in other regions. On the contrary, these delays probably arose precisely owing to the continued profitability of these pre-industrial trading opportunities, which reduced the incentive of the concerned communities to enter industry. In other words, successful community networks can hinder progression into the next higher stage of development, an effect that did not manifest itself in the phenomena studied in this paper. Exploring this hypothesis remains an interesting task for future research.

References

[1] Allen, R.C. 2009. *The British industrial revolution in global perspective* (Vol. 1). Cambridge: Cambridge University Press.

[2] Bagchi, A.K., 1972. *Private investment in India, 1900-1939*, Delhi: Cambridge University Press & Orient Longman.

[3] Bagchi, A.K., 1987. *The Evolution of the State Bank of India: The era of the presidency banks, 1876-1920* (Vol. 2). Oxford University Press.

[4] Bagchi, A.K, 1994. European and Indian entrepreneurship in India 1900-30, in R.K. Ray, ed., *Entrepreneurship and Industry in India 1800-1947*, Delhi: Oxford University Press.

[5] Banerjee A. and Newman A., 1993. Occupational Choice and the Process of Development. *Journal of Political Economy* Vol. 101, No. 2. (Apr., 1993). pp. 274-298.

[6] Banerjee, A. and Munshi, K., 2004. How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur. *The Review of Economic Studies, 71*(1), pp.19-42.

[7] Beaman, L. and Magruder, J., 2012. Who gets the job referral? Evidence from a social networks experiment. *American Economic Review, 102*(7), pp.3574-93.

[8] Bogart, D. and Richardson, G., 2011. Property rights and parliament in industrializing Britain. *The Journal of Law and Economics, 54*(2), pp.241-274.

[9] Burhop, C., 2006. Did banks cause the German industrialization?. *Explorations in Economic History, 43*(1), 39-63.

[10] Dai R., D. Mookherjee, K. Munshi and X. Zhang, 2018. Community Networks and the Growth of Private Enterprise in China. Working Paper, Institute for Economic Development, Boston University.

[11] Dasgupta, U. ed., 2001. *The World of the Indian Ocean Merchant, 1500-1800: Collected Essays of Ashin Das Gupta*. Oxford University Press, USA.

[12] Davis, J. H., & Irwin, D. A., 2008. The antebellum US iron industry: domestic production and foreign competition. *Explorations in Economic History, 45*(3), 254-269.

[13] Desai, A.V., 1968. The origins of Parsi enterprise. *The Indian Economic & Social History Review, 5*(4), pp.307-317.

[14] Fafchamps, M., 2003. *Market institutions in sub-Saharan Africa: Theory and evidence.* MIT press.

[15] Goswami, O., 1985. Then came the Marwaris: Some aspects of the changes in the pattern of industrial control in Eastern India. *The Indian Economic & Social History Review, 22*(3), pp.225-249.

[16] Greif, A., 2006. *Institutions and the path to the modern economy: Lessons from medieval trade.* Cambridge University Press.

[17] Guha, A., 1984. More about the Parsi Seths: Their Roots, Entrepreneurship and Comprador Role, 1650-1918. *Economic and Political Weekly*, pp.117-132.

[18] Gupta, B., 2014. Discrimination or social networks? Industrial investment in colonial India. *The Journal of Economic History, 74*(1), pp.141-168.

[19] Harnetty, P., 1972. *Imperialism and free trade: Lancashire and India in the mid-nineteenth century.* Manchester University Press.

[20] Harvey D., 2005, *The New Imperialism.* Oxford University Press.

[21] Irwin, D. A. 2007. Tariff incidence in America's gilded age. *The Journal of Economic History, 67*(3), pp. 582-607.

[22] Kahan, A. (1967). Government policies and the industrialization of Russia. *The Journal of Economic History, 27*(4), pp. 460-477.

[23] Kamenar, M. P. 1998. Parsis in Maritime Trade on the Western coast of India from 17th to nineteenth century in M. B. Mody (ed) *The Parsis in Western India: 1818-1920*, Allied Publishers: Bombay: pp. 129-145.

[24] Kranton, R.E. and Swamy, A.V., 1999. The hazards of piecemeal reform: British civil courts and the credit market in colonial India. *Journal of development Economics, 58*(1), pp.1-24.

[25] Kumar, D. and Desai, M., 1983. *The Cambridge economic history of India* (Vol. II) Cambridge.

[26] Markevich, A and Nafziger, S, 2017, The state and market in Russian industrialization, 1870-2010, in O'rourke, K. H and Williamson J. G, *Spread of Modern Industry in the Poor Periphery since 1871*, Oxford University Press: pp. 33-62.

[27] Mehta, M., 1991. *Indian Merchants and Entrepreneurs in Historical Perspective: With Special Reference to Shroffs of Gujarat,* 17^{th} *to* 19^{th} *Centuries.* Academic Foundation.

[28] Marx K., 1887, *Capital* (Vol. 1). Progress Publishers, Moscow.

[29] Mehta, S.D., 1953. *The Indian Cotton Textile Industry: An Economic Analysis*, GK Ved for the Textile Association, India.

[30] Mokyr, J., 2009. *.The enlightened economy. An economic history of Britain, 1700,* p.1850.

[31] Morris, M.D., 1965. *The emergence of an industrial labor force in India: A study of the Bombay cotton mills*, 1854-1947. Univ of California Press.

[32] Morris, M.D. 1983. "Growth of Large-scale Industry to 1947." In Dharma Kumar, ed., *The Cambridge Economic History of India, vol. 2*, Cambridge: Cambridge University Press: 553-676.

[33] Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics, 118*(2), pp.549-599.

[34] Munshi, K., 2011. Strength in numbers: Networks as a solution to occupational traps. *The Review of Economic Studies, 78*(3), pp. 1069-1101.

[35] Oŕourke, K. H. 2000. Tariffs and growth in the late 19^{th} century. *The Economic Journal, 110*(463), pp. 456-483.

[36] Perkins, D and Tang, JP, 2017, The East Asian Pioneers: Japan, Korea and Taiwan in O'rourke, K. H and Williamson J. G, *Spread of Modern Industry in the Poor Periphery since 1871*, Oxford University Press: pp. 169-196.

[37] Rauch, J.E. and Trindade, V., 2002. Ethnic Chinese networks in international trade. *Review of Economics and Statistics, 84*(1), pp.116-130.

[38] Ray, R.K., 1979.*Industrialization in India: growth and conflict in the private corporate sector, 1914-47.* Delhi: Oxford University Press.

[39] Ray, R.K., 1994. *Entrepreneurship and India in India, 1800-1947*. Delhi: Oxford University Press

[40] Roy, T. and Swamy, A.V., 2016. *Law and the economy in colonial India*. University of Chicago Press.

[41] Rudner, D.W., 1994. *Caste and capitalism in colonial India: the Nattukottai Chettiars* (p. 93). Berkeley: University of California Press.

[42] Rungta, S., 1970. *The Rise of Business Corporations in India 1851-1900* (No. 8). CUP Archive.

[43] Rutnagur, S.M. ed., 1927. *Bombay industries: the cotton mill. A review of the progress of the textile industry in Bombay from 1850 to 1926 and the present constitution, management and financial position of the spinning and weaving factories*. Indian textile journal.

[44] Shah, A. M., 2010. *The Structure of Indian Society: Then and Now. Routledge.*

[45] Sutton John and Nebil Kellow, 2010. *An Enterprise Map of Ethiopia*. International Growth Centre, London.

[46] Sutton John and Bennet Kpentey, 2012. *An Enterprise Map of Ghana*. International Growth Centre, London.

[47] Timberg, T.A., 1978. *The Marwaris, from traders to industrialists*. Stosius Incorporated/Advent Books Division.

[48] Tripathi, D. and Mehta, M., 1990. *Business Houses in Western India: A Study in Entrepreneurial Response, 1950-1956*, Jaya books.

[49] Tripathi, D., 2004. *The Oxford history of Indian business*. Oxford University Press.

[50] Webb, S. B. 1980 Tariffs, cartels, technology, and growth in the German steel industry, 1879 to 1914. *The Journal of Economic History, 40*(2), pp. 309-330.

8 Appendix

8.1 Data sources

We have several data sources which we collected and digitalized. In this paper, we construct panels for cotton and jute industry in India, between 1850 and 1930 at the entrepreneur level.

Our data provides information on preindustrial activity from before the first waves of industralization, documenting the presence of cotton merchants and jute balers, and up to 1930 with the impact of the Great Depression. In cotton sector, industralization started right after the American civil war, while in jute sector Indian entry took place after 1914.

8.1.1 Cotton industry registers

"*The Bombay Almanac, Directory, and Register*", (1806 - 1868)

We construct the List of Merchants, from the records of firms and partners in firms organised by occupation for years: 1840, 1850, 1855, 1860, 1865. In this dataset is also available the list of Joint Stock Companies from 1860 onward.

" *Bombay Calendar and Almanac*", (1853 - 1861)

From this dataset we only use years 1860 and 1861, where we obtain: (i) the list of firms, location, and capital; (ii) the list of entrepreneurs and individuals by occupation.

"*Times of India Calendar and Directory (Bombay)*", (1862-1887)

We use this dataset for the period (1862 - 1884), where we collect: (i) the list of firms, location, and capital; (ii) the list of entrepreneurs.

"*Thacker's Indian Directory (Bombay)*", (1885 - 1960)

This is our main data source, which we use from year 1885 onward. From here we construct: (i) the list of firms, location, and capital; (ii) the list of entrepreneurs.

8.1.2 Jute industry registers

"*New Calcutta Directory*" (1856 - 1863)

We construct yearly cross sections between 1860 and 1869: (i) the list of firms, location, and capital; (ii) the list of entrepreneurs.

"*Thackers Bengal Directory*" (1863 - 1884)

From this dataset we construct yearly cohorts from 1870 to 1875, in addition to 1880. The information found in the directory is: (i) the list of firms, location, and capital; (ii) the list of entrepreneurs.

"*Thacker's Indian Directory (Calcutta)*", (1885 - 1960)

This dataset is used to obtain: (i) the list of jute balers; (ii) the list of firms, location, and capital; (iii) the list of entrepreneurs and individuals by occupation.

From 1885 to 1905 we collected data every five years, and form 1905 we construct our yearly cohorts up to 1930.

8.1.3 Cotton and jute prices

"*Index Number of Indian Prices*" India Office Records (1861-1931)

The most relevant pieces of information extracted from this source are the following:

Exported articles: (i) Cotton, Raw Broach (Bombay) per candy of 784 lb; (ii) Cotton, Manufactured Yarn 20s (Bombay) per lb and T. cloth (Bombay) per lb (only between 1874-1931); (iii) Jute, Raw Picked & Ordinary (Calcutta). Per bale of 400 lb.; (iv) Jute, Manufactured: Gunny bags (Calcutta). Per bale of 100 lb

Imported articles: (i) Cotton, Manufactured Grey shirtings (Calcutta) and Grey yarn Banner Mill (Calcutta)

8.1.4 Entrepreneur and firm panels construction

We first pool yearly cross-section cohorts with all firms in upstream and downstream sector. From each firm we obtain the names of partners, directors and managing agents. Each of them were associated to a community origin, which defined the firm's community by simple majority. Raw and manufactured cotton (and jute) prices from the *Index Number of Indian Prices* were yearly assigned to entrepreneurs and firms panels.

Entrepreneurs panel collapses yearly information at individual level, as a single entrepreneur could be part of more than one firm in the same or different sector during the same year. Therefore, in our panel we are able to identify whether an entrepreneur is incumbent in one sector and later entrant in the other. Information on capital requirement is summarised at individual level as well.

Firms panel gathers all relevant information at entrepreneurs level. We assign a community to each entrepreneur using his name. We calculate the proportion of entrepreneurs belonging to the same community in a firm and the firm's community identity is based on the community of the majority of entrepreneurs. For each firm we assign a code or ID, which enables to track the same firm despite of firm's names changes. The location of firms is obtained from registered office address, and mills or presses location.

Cross section cohorts for cotton industry panels: (1860-61) *Bombay Calendar and Almanac*; (1862-1884) *Times of India Calendar and Directory (Bombay)*; (1885-1910) *Thacker's Indian Directory (Bombay)*;

Cross section cohorts for cotton jute panels: (1860-1869) *New Calcutta Directory*; (1870-1875) *Thackers Bengal Directory*; (1885-1905) for every five years *Thacker's Indian Directory (Calcutta)*; (1905-1930) *Thacker's Indian Directory (Calcutta)*