Community Networks and the Growth of Private Enterprise in China*

Ruochen Dai† Dilip Mookherjee‡ Kaivan Munshi§ Xiaobo Zhang¶

March 16, 2018

Abstract

In this paper, we identify and quantify the role played by community networks, organized around the birth county, in the growth of private enterprise in China. The model that we develop generates dynamic predictions for firm entry, sectoral and spatial concentration, and firm size across birth counties with different levels of social connectedness (measured by population density) when networks are active. We validate each of these predictions with unique administrative data that cover the universe of registered firms and provide information on entrepreneurs' birth counties. The rich set of results that we obtain, taken together, allow us to rule out alternative non-network explanations. Having validated the model, we estimate its structural parameters and conduct counter-factual simulations, which indicate that entry and (initial) capital stock over the 1995-1999 period, a critical period in Chinese industrial development, would have been 25% lower in the absence of community networks. Implications for industrial policy in economies where networks are active are consequently derived.

Keywords. Migration. Misallocation. Institutions. Growth and Development. Macroeconomic Implications. Community Networks. Entrepreneurship
JEL. J12. J16. D31. I3.

*We are grateful to numerous seminar participants for their constructive comments. Research support from the Economic Development Initiative (EDI)and Cambridge-INET is gratefully acknowledged. We are responsible for any errors that may remain.

†Peking University
‡Boston University
§University of Cambridge
¶Peking University

1 Introduction

China has witnessed the same degree of industrialization in three decades as Europe did in two centuries (Summers, 2007). This economic transformation began in the early 1980's with the establishment of township-village enterprises (TVE's) and accelerated with the entry of private firms in the 1990's. Starting with almost no private firms in 1990, there were 15 million registered private firms in 2014, accounting for over 90% of all registered firms and 60% of aggregate industrial production. The surge in the number of private firms has had major macroeconomic consequences. China is the largest exporter in the world today and, depending on how the accounting is done, the world's largest or second-largest economy (Wu, 2016).

While China's growth has been impressive by any yardstick, what is perhaps most striking about this growth is that it occurred without the preconditions that are generally believed to be necessary for economic development; i.e. without effective legal systems or well functioning financial institutions (Allen, Qian, and Qian, 2005). The government compensated for some of these institutional limitations by providing infrastructure and credit (Long and Zhang, 2011; Wu, 2016). However, there were clearly other powerful forces at work that allowed millions of entrepreneurs, most of whom were born in rural areas, to establish and grow their businesses. Allen, Qian, and Qian (2005) hint at the nature of these forces by noting that alternative channels based on reputation and relationships must have been present to support such a high level of growth in the absence of market institutions. And case studies of production clusters; e.g. (Fleisher, Hu, McGuire, and Zhang, 2010), indicate that long-established relationships among relatives and neighbors (from the rural origin) substituted for legal enforcement of contractual relationships among firms. Our analysis formalizes these ideas, shifting attention from the government to the role played by informal community networks in the establishment and evolution of private enterprise in China.

China is divided into approximately 2,000 counties and 250 cities (which are further divided into urban districts). There is an emerging literature that describes the role played by hometown or birth county networks in facilitating China's unprecedented rural-urban labor migration; e.g. Honig (1996), Ma and Xiang (1998), Zhao (2003). If the sending county is the domain around which migrant labor networks are organized, then we expect that this will also be the domain around which business networks supporting county-born entrepreneurs are organized. Our strategy to identify and quantify the contribution of these birth county networks to the growth of private enterprise in China exploits exogenous variation in social connectedness in the population across counties. We expect entrepreneurs born in counties with greater social connectedness to have access to better performing networks, which will have verifiable consequences for entrepreneurial choices and firm outcomes.

The model that we develop to identify a role for birth county networks derives the dynamic relationship between social connectedness in the birth county and the entry of firms, sectoral concentration, and firm size when networks are active. There are many counties of origin, distinguished by their social connectedness, in the model. At every origin, an equal-sized cohort of agents is born in each period. Each agent makes a once-and-for-all occupational choice, selecting between a traditional, non-entrepreneurial, sector and two business sectors. The profit in the traditional sector is an increasing function of the agent's ability. The profit in the business sectors, which are identical in their 'fundamentals,' is determined by the agent's ability, the capital that is invested in the firm, and 'community' TFP (CTFP), which reflects learning spillovers or gains from intra-network cooperation.[1] CTFP is a dynamically evolving community level variable, which depends on the size of the incumbent network; i.e. the stock of entrepreneurs from the community who previously selected into that sector, and the quality of the network, which is increasing in social connectedness.

[1]Although the interest rate is the same for all agents and constant over time in the model, larger and higher quality networks could also lower borrowing costs for their members. The resulting dynamics turn out to be very similar to those generated by productivity spillovers and, thus, the alternative network channels cannot be separately identified.

The network dynamics in each community commence with a small exogenous number of entrepreneurs in each business sector, and a slight advantage in one sector. A fixed fraction of agents receive an opportunity to become an entrepreneur in each subsequent cohort. Each such agent receives a single referral to enter one of the business sectors, with the fraction of agents receiving referrals from a given sector equal to the share of incumbent entrepreneurs from their community already in that sector. There are thus two sources of dynamic increasing returns in our model. One source is the interaction between network quality and size; an additional entrant in any period induces additional entry in the next period, with a compounding effect in subsequent periods that is amplified in high quality networks. This dynamic multiplier, associated with post-entry network benefits, is augmented by the second source of increasing returns operating through the referrals process, which increasingly channels firms into the initially (slightly) favored sector. Greater sectoral concentration generates greater aggregate entry by channeling firms into a single sector where they can take better advantage of the increasing returns generated by network size. The two sources of dynamic increasing returns thus complement each other, with the flow of entrants and sectoral concentration evolving together and increasing more steeply over time when networks are of higher quality; i.e. when they are drawn from populations with greater social connectedness.

Among the agents receiving a referral in a given birth cohort, only those with ability above a threshold level will select into business. This threshold declines over time (across successive cohorts) as CTFP and, hence, profits in the business sector grow, with a steeper decline in more socially connected populations because their networks are growing especially fast. Higher CTFP thus has two conflicting effects on the initial size of the marginal entrant's firm: the direct effect is to increase firm size by raising firm level TFP, but the negative selection on ability works in the opposite direction to lower the firm's TFP. In our model, the latter effect dominates; the marginal entering firm will be smaller in more socially connected populations, with this (negative) relationship growing stronger over time as their networks strengthen. Under specific conditions on the model parameters, this result is shown to hold for average firm size as well. This contrasts with the model's predictions for the post-entry growth in firm size. This growth is driven by changes in CTFP and is independent of initial size. Thus, firms owned by entrepreneurs from more socially connected origin counties will start smaller but subsequently grow faster.

We measure social connectedness by the population density in the entrepreneur's birth county. Social interactions within counties will be increasing in spatial proximity, which is increasing in population density. More frequent social interactions; i.e. greater social connectedness, allow for more effective social sanctions and, hence, higher levels of trust and economic cooperation. While this reasoning makes sense for county-born entrepreneurs, it may not for city-born entrepreneurs for two reasons. First, population densities are over four times larger in cities than in counties on average. If there is an upper bound to an individual's social interactions, then spatial proximity may not be a constraint to these interactions in the city. Second, multiple social groups co-exist in the city. If the number of groups is increasing in population density and social interactions cross group lines, then the within-group social interactions that are necessary for economic cooperation could even be decreasing in population density. Providing support for the distinction between counties and cities, we show with data from the China Family Panel Survey (CFPS) that the frequency of local social interactions, as well as trust in local residents, is increasing with population density in counties but not in cities. The core analysis thus focusses on entrepreneurs born in counties, most of whom would have been first-generation entrepreneurs (although the city-born entrepreneurs will serve as a useful placebo group). Their firms account for two-thirds of all registered private firms in China, highlighting the substantial inter-generational occupational mobility that characterizes the Chinese development experience.

We test the predictions of the model with unique administrative data covering the universe of registered firms in China that we have obtained from the State Administration of Industry and Commerce (SAIC).

The following information is available for each firm: establishment date, 4-digit sector, location, ownership-type, registered capital (initial and subsequent changes), and a list of major shareholders and managers, with their citizenship ID. The county of birth can be extracted from the citizenship ID, and the firm's legal representative is designated as the "entrepreneur" in the analysis.[2] The analysis is restricted to private firms and covers the 1990-2009 period, starting with the first wave of private entry and ending with the financial crisis. As discussed, population density proxies for social connectedness in counties but not in cities. The core analysis thus focuses on county-born entrepreneurs, using population density from the 1982 population census, prior to large-scale labor migration in China, as a predetermined measure of social connectedness in their birth counties. As predicted by the model, we establish that the flow of entrants and sectoral (and spatial) concentration are (i) increasing in birth county population density at each point in time, (ii) increasing over time, and (iii) increasing more steeply in population density over time.[3] In addition, we establish that the marginal entrant's ability, measured by his education, and initial capital (both for the marginal and the average entrant) are decreasing over time within birth county-sectors, more steeply in higher population density birth counties.

Although the SAIC registration database is well suited to examine entry, sectoral and location choice, and initial firm size, it is less useful for analyses of firm growth. Given the administrative costs involved, firms do not adjust their registered capital from year to year. We thus turn to the SAIC inspection database, which includes annual asset information from 2004 onward, and the 2004 and 2008 rounds of the industrial census, which also includes asset information, to test the model's predictions for firm growth. We estimate a positive and significant relationship between birth county population density and the average annual growth of the entrepreneur's firm over the 2004-2008 period, with both datasets. The model generates a rich set of predictions with respect to birth county population density when community networks are active, and we are able to validate each of them.

As a final, most direct, test of the network model, we estimate the relationship between initial entry in the 1990-1994 period, by firms from a given birth county in a particular sector-destination, and subsequent entry by firms from that birth county. Providing support for the network multiplier effect, we find that one additional initial entrant results in seven additional entrants over the 2000-2004 period and nine additional entrants over the 2005-2009 period. The initial effects are, moreover, stronger for firms from higher population density counties. In contrast, conditional on the number of initial entrants from the birth county, the total number of initial entrants in the sector-destination has no predictive power for the number of subsequent entrants. This striking result indicates that the birth county is indeed the domain around which business networks are organized in China and that these networks operate independently within sector-destinations. Analyses of agglomeration effects that ignored birth county information would evidently paint an erroneous picture of spillover effects in China.

Our use of a predetermined population-based measure of social connectedness as the source of forcing variation in the empirical analysis avoids many of the pitfalls encountered by analyses based on endogenously determined network characteristics. Nevertheless, population density is not a statistical instrument for social connectedness. It is possible that population density is (accidentally) correlated with other direct determinants of entry, concentration, and firm size. We account for this possibility in different ways; by including different determinants of the outcomes of interest in the estimating equations, by implementing placebo tests, and by ruling out alternative explanations (by showing that they cannot account for all the results that are obtained). We systematically examine alternative explanations by analyzing a simple model that retains the

[2]The legal representative and the largest shareholder are born in the same county 90% of the time.

[3]Although we do not explicitly model location choice, the predictions for spatial concentration (across locations outside the birth county) are identical to those for sectoral concentration. The analysis of spatial concentration is based on the 60% of entrepreneurs who establish their firms outside their birth county.

essential features of our model, but without the network component; i.e. with community TFP replaced by an exogenously increasing productivity term at the destination. The alternative explanations are generated by introducing additional sources of heterogeneity at the origin or the destination, which are correlated with population density, in turn.

We show that additional heterogeneity at the origin; with higher population density counties having larger populations, higher education, or lower moving costs, can generate greater entry, but cannot explain the dynamic patterns of sectoral and spatial concentration, nor the observation that firms from high population density counties start small but, nevertheless, grow faster. Heterogeneity at the destination, with firms from higher population density birth counties having access to more favorable destinations, can explain all the results that are obtained, but there are a number of reasons why this assumption may not be valid. First, we find no evidence that these firms ended up at destinations (or in sectors) that received a larger number of entrants overall. Second, the initial capital and firm growth results are robust to the inclusion of destination fixed effects. Third, entry into particular sector-destinations is driven entirely by initial entry into those sector-destinations by firms from the same birth county, with a stronger initial effect for firms from higher population density counties. If all firms in a given sector-destination were benefiting from the same economic environment, then this result would not be obtained.

Having tested and validated the model with networks, the next step in the analysis is to quantify the impact of these networks on firm entry and growth (measured by capital stock). To do this we estimate the structural parameters of the model and then conduct a counter-factual experiment in which the networks are shut down. Although the model is extremely parsimonious, it does a good job of matching entry and initial capital within sectors, across the range of birth county population densities, within the sample period for the structural estimation (1995-2004) and out of sample. This increases our confidence in the results of the counter-factual experiment, which estimates that entry and (initial) capital stock would have declined by around 25% over the 2005-2009 period had the networks not been active. Given the dynamic increasing returns that are generated by the networks, the long-term consequences of their absence (continuing to the present day) would have been even more substantial.

Given the large network externalities that we have estimated, a natural question to ask is whether the Chinese government could have accelerated growth by providing firms with subsidized credit. To answer this question, we examine a counter-factual policy experiment in which all entering firms in the 1995-2000 period received a credit subsidy. This subsidy generates a net surplus in the 1995-2000 period; i.e. the additional private profits exceed the cost to the government. However, the direct effect of the one-time subsidy is dwarfed by the positive spillover effects on firm profits in future periods, especially in higher population density birth counties, which are generated by the additional firms that were induced to enter in 1995-1999 (and the subsequent compounding of the network effect). If the government took account of these spillovers, which are not currently internalized by individual entrants, and its objective was to maximize total profit, then the optimal policy would be to provide subsidized credit to marginal entrepreneurs (who would not enter otherwise) from birth counties with higher population densities. Given the negative selection that we have documented, these entrepreneurs would have relatively low ability and relatively small firms, and existing variation in firm size and productivity within sectors (across networks) would widen even further.

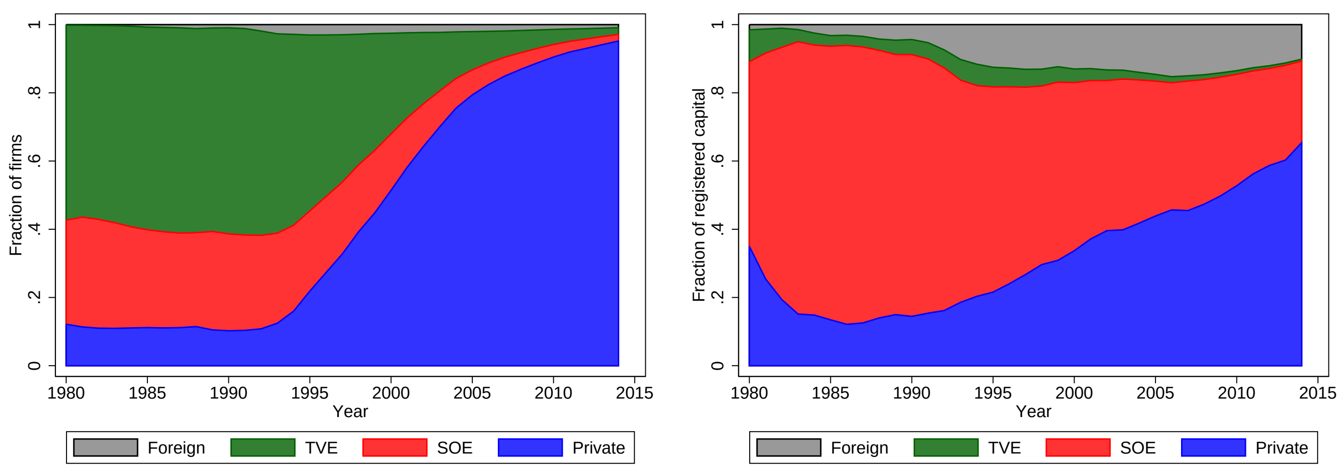
The preceding result highlights a general insight of our research, which is that while small firms and wide dispersion in firm size and productivity may be symptomatic of market inefficiencies in developing economies (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009, 2014), in a world with such inefficiencies (and the networks that emerge in response to these inefficiencies) a more successful second-best response may, paradoxically, be accompanied by smaller firms and greater dispersion. The additional implication of our research is that attempts to stimulate growth by offering subsidies to potential entrepreneurs should take account of both individual ability, which determines independent business success, and social affiliation, which determines the spillovers that lead to further entry, in economies where networks are active. Existing efforts to stimulate entrepreneurship through business training programs or business plan competitions, as described in (McKenzie and Woodruff, 2014) and (McKenzie, 2017), do not take account of these spillovers, potentially resulting in a substantial loss in efficiency. At the same time, the distributional consequences of policies that attempt to exploit network externalities must also be considered, as discussed in the concluding section.

2 Institutional Setting

2.1 The Growth of the Private Sector

The core administrative data set that we utilize for the empirical analysis comprises the universe of registered firms in China, regardless of their size, from 1980 onwards. These firms are classified as township-village enterprises (TVE's), state owned enterprises (SOE's), foreign owned firms, and private (domestically owned) firms. Our interest is in the last category. New firms enter the data each year, while a fraction of incumbents exit. We can thus trace the growth of the private sector in China from its inception in the early 1990's, which coincided with the phasing out of the TVE's. As documented in Figure 1a, private firms accounted for approximately 10% of all firms in the early 1990's. Subsequently, they grew extremely rapidly and by 2014 they accounted for over 90% of all firms.

Figure 1. Distribution of Firms, by Type



(a) Number of Firms

(b) Registered Capital

Source: SAIC registration database.

While private firms may have increased substantially in numbers, what was their contribution to overall output? To answer this question, we examine the evolution of registered capital, by firm-type, over time.[4] Figure 1b reports the share of total registered capital, by firm-type, over the 1980-2014 period. As with their numbers, the share of registered capital held by private firms grows steeply from the early 1990's onwards and by 2014 they hold 60% of total registered capital in the Chinese economy.

[4]The initial registered capital represents the total amount paid up by the shareholders. This amount is deposited with the State Administration of Industry and Commerce (SAIC) in China, and can be used to pay the firm's operating expenses before it becomes cash flow positive. Access to bank credit is also dependent on the firm's registered capital, which is why firms will often choose to increase their registered capital over time.

2.2 Community Networks and Private Enterprise

The preceding descriptive analysis suggests that private firms played an important role in China's rapid growth over the past decades. But what forces allowed private firms to enter and, thereafter, to grow? It is generally believed that governments at the local (county), provincial, and central level played a critical role in China's economic transformation. Local governments provided the infrastructure to support production clusters located throughout the country, which are a distinctive feature of the Chinese economy (Long and Zhang, 2011). Provincial governments and the central government supported firms by giving them subsidized credit and by aggressively promoting exports (Wu, 2016). Our research shifts focus from the government to the role played by informal community networks in spurring private enterprise in China.

Case studies of Chinese production clusters; e.g. Huang, Zhang, and Zhu (2008), Ruan and Zhang (2009), and Fleisher, Hu, McGuire, and Zhang (2010) consistently find that the impetus for their formation came from within, with groups of entrepreneurs setting up firms with little external support. The involvement of local governments is found to come later, through the provision of infrastructure such as roads, markets, and quality control. What supporting mechanisms allowed these early entrants to come together and commence production in an often initially rural location? There is an emerging literature on the role played by social networks or *guanxi* in facilitating China's historically unprecedented rural-urban labor migration over the past decades; e.g. Zhao (2003), Zhang and Li (2003), Hu (2008). This literature describes how migrant networks are organized around the rural hometown, complementing a well established body of work that takes the position that ethnicity in China is defined by the native place; e.g. Honig (1992, 1996), Goodman (1995). Migrants from the same rural origin move to the city in groups and most migrants end up living and working with *laoxiang* or "native-place fellows" (Cai Fang, 1997; Ma and Xiang, 1998). In Chinese cities, migrant-peasant enclaves are often named after a sending province, but as Ma and Xiang (1998) note, this nomenclature is misleading because the enclave typically consists of peasants from a single county or two neighboring counties. If the sending county is the domain around which migrant labor networks are organized, then we expect that this will also be the natural domain around which business networks supporting (migrant) county-born entrepreneurs are organized.[5]

The registration database includes a list of key individuals; i.e. major shareholders and managers in each firm. The citizenship ID is reported for each listed individual, from which the first six digits reveal the birth county.[6] The firm's legal representative is treated as the key individual or "entrepreneur" in the analysis that follows.[7] China is divided into approximately 2,000 counties and 250 prefecture-level and province-level cities (which are further divided into urban districts). Our analysis focuses on county-born entrepreneurs, of which 60% operate their business outside their birth county and for whom the hometown networks are relevant. City-born entrepreneurs will also have access to networks, but these networks will be organized around classmates or other peer groups.[8] Figure 2 describes the growth in the number of firms owned by county-born entrepreneurs, as well as their share of the total capital registered by private firms. County-born entrepreneurs make up about two-thirds of all entrepreneurs in China, with their firms accounting for a

[5]A similar argument has been made in past research in India, where the endogamous caste or *jati* is the common domain around which networks supporting rural-urban migration, business, and other functions are organized (Munshi and Rosenzweig, 2006, 2016; Munshi, 2011).

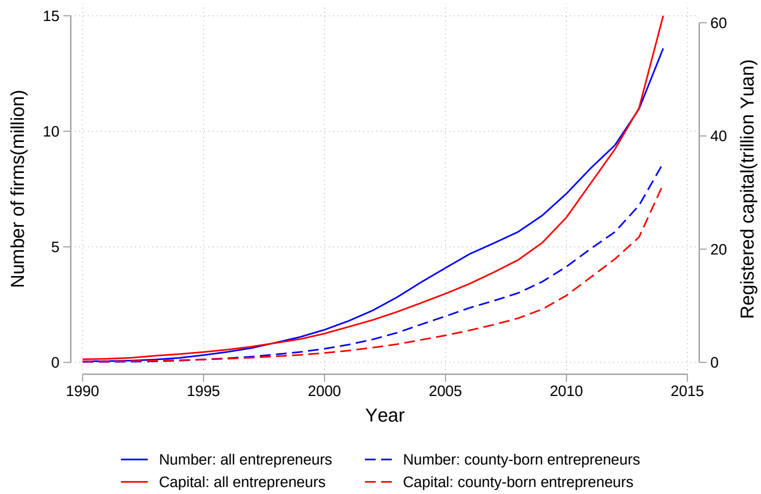
[6]Citizenship ID's were first issued in September 1985 and people born after that date are given an ID at birth. Those born before that date were registered in the county or city where they resided at the time. Given the limited opportunities for labor migration in that period and the cost of moving due to the Hukou System, almost all rural-born individuals resided in their birth-counties in 1985. The only exceptions were college students, college graduates, and soldiers, but these numbers were small. The first six digits of the citizenship ID thus reveals the county of birth, with few exceptions, even for those born before September 1985.

[7]This individual is legally responsible for the firm's liabilities. 75% of legal representatives are shareholders in their firms. The legal representative and the largest shareholder belong to the same birth county in over 90% of firms.

[8]Empirical support for this distinction, based on the nature of social interactions, will be provided below. The city-born entrepreneurs will, nevertheless, form a useful placebo group in the empirical analysis that follows.

slightly smaller share of registered capital.

Figure 2. Growth of Private Enterprise, by Birthplace of Entrepreneurs



Source: SAIC registration database.

Our objective in the analysis that follows is to identify and quantify the role played by hometown networks in supporting the entry and subsequent growth of these firms. The model that we develop derives predictions for firm entry, sectoral concentration, and firm size when hometown networks are active. We test these predictions, and rule out alternative explanations, using data from multiple sources. Estimation of the model's structural parameters and accompanying counter-factual simulations subsequently allow us to quantify the role played by the community networks in China's dramatic growth over the past decades.

3 A Model of Network Dynamics

3.1 Assumptions

There are many counties of origin; each generates a network of entrepreneurs with an independent trajectory. We assume the absence of any interactions across networks of different origins, and will later justify this assumption empirically. Hence, we focus on the dynamics of a single network originating in a given origin. Each origin has a given, exogenous level of 'social connectedness' represented by a real-valued parameter $p \geq 0$; this will determine the speed of learning or productivity spillover within networks of entrepreneurs born there. Although we 'black-box' how p increases productivity in the model, micro-foundations for the network effect are provided in Appendix A. The resulting variations in network dynamics with respect to the p parameter plays a key role in our analysis.

At every origin there are equal-sized cohorts of new agents born at dates $t = 1, 2, ...,$ who live for ever thereafter. Every cohort t agent makes a once-for-and-all choice of occupation at t. Everyone has the option to enter a traditional, non-entrepreneurial (T) sector; some of them will have an opportunity to enter either of two business sectors B_1, B_2. The two business sectors have identical 'fundamentals' but will typically end up with unequal levels of entry owing to network complementarities and asymmetric initial conditions. At the beginning of the process ($t = 0$), there is a small, exogenous number n_{i0} of older entrepreneurs (from cohorts preceding $t = 1$) who have already entered B_i. These represent the initial conditions for the dynamics. Generically these historical entry levels will not exactly be balanced across the two sectors; without loss of generality suppose $n_{10} > n_{20}$ so B_1 has a slight edge.

The dynamics operate as follows. Consider any cohort t; denote total entry into sector B_i by past cohorts by $n_{i,t-1}$. Let $N_{t-1} \equiv n_{1,t-1}+n_{2,t-1}$ denote the aggregate number of business entrepreneurs from past cohorts from the same origin, and $s_{i,t-1} \equiv \frac{n_{i,t-1}}{N_{t-1}}$ the share of B_i in this stock.

A fixed fraction $k \in (0,1)$ of agents in every cohort receive an opportunity to become an entrepreneur. Each such agent receives an opportunity to enter one of the two sectors. The fraction that get an opportunity to enter B_i equals $s_{i,t-1}$, which is the share of incumbent entrepreneurs already in that sector. This is the first important source of network complementarity, operating via aspirations, access to information, or referrals provided by older members from the same origin in a given sector.

The second source of network complementarity is through post-entry productivity spillovers from incumbents. To explain this, we characterize agent abilities and the production function in each sector. Each agent is born with a random ability draw ω from an i.i.d. distribution. To obtain closed form solutions, we assume that the distribution of $\log \omega$ is uniform on $[0,1]$. Profits are increasing and concave in ability in the traditional sector: the profit of an agent with ability ω in the T sector is ω^σ, where $\sigma \in (0,1)$. It will turn out to be linear in ability in each business sector. So low ability agents would be better off staying in the traditional sector, while high ability agents might want to switch to becoming entrepreneurs.

In sector B_i at date t, an entrepreneur with ability ω who selects capital size K has a production function

$$y = A_{it}\omega^{1-\alpha}K^{\alpha} \tag{1}$$

where $\alpha \in (0,1)$ is the capital elasticity, and A_{it} denotes 'community' TFP, given as follows:

$$A_{it} = A_0 \exp(n_{i,t-1}\theta(p)) \tag{2}$$

Note that the TFP of each agent actually depends on his own ability draw, besides A_{it}. Hence 'community TFP' A_{it} can be interpreted as the TFP of the highest ability agent in the community: we shall refer to it as CTFP from now onwards. CTFP is a dynamically evolving community level variable. It depends on the size ($n_{i,t-1}$) of the incumbent network, and on its quality $\theta(p)$ which is rising in p. The exponential relationship implies that the rate of growth of CTFP over time is proportional to the (quality-weighted) increase in network size. This represents the second source of network complementarity in the model, reflecting learning spillovers or gains from intra-network cooperation in improving productivity or market size. An explicit micro-foundation for this specification is provided in Appendix A. Briefly, the idea is that productivity improvements require help from others with experience in the industry. This requires both access to people with the appropriate skills or connections, as well as their willingness to provide help. A larger network implies a higher likelihood of finding the right person who has the capacity to provide the help, while a better quality network raises the nature of help provided.

All agents incur the same cost of capital r which is exogenous and fixed across all t, and all origins. We are thus abstracting from possible network complementarities operating via internal capital markets, as in Banerjee and Munshi (2004), which arise in response to financial market imperfections. To the extent that larger and higher quality networks lower borrowing costs for their members, the resulting dynamics turn out to be very similar to those generated via productivity spillovers, and would thus amplify the dynamics generated by the latter alone.

We also assume a fixed price of the product (normalized to unity) which is unaffected by supply from the network. This abstracts from price collusion among network members, as well as limits to market size in a competitive context. These seem plausible in the Chinese setting, where most sectors are comprised of a large number of individual origin county networks, and both domestic and international market opportunities were large.[9]

[9] Based on the registration data, firms from a given origin county account for 13% of firms at the destinations where they locate, on average (within narrow two-digit sectors). This included entrepreneurs who locate their firms in their county of birth.

Finally, we assume that agents receiving an entrepreneurial opportunity make their decision *selfishly* and *myopically.* The former assumption implies that they ignore the consequences of their entry decisions on the profits of other agents. The latter states that they make their choice solely to maximize their date–t profits, ignoring consequences at later dates. This enables us to compute the entry dynamics recursively, simplifying the analysis considerably. As network size cannot ever decrease over time and its quality does not change, and neither do profits in the traditional sector, this amounts to a conservative bias in entry decisions: those deciding to enter based on a myopic calculation would also want to enter if farsighted, while some others deciding to stay out on myopic grounds may wish to enter when they anticipate future network growth which will raise the returns to entrepreneurship. If agents were more far-sighted, they would have to forecast current and future levels of entry from the same origin county, generating strategic complementarity of entry decisions within each cohort. This extension is considered in Appendix A, where entry decisions at t are based on the discounted sum of profits at t and $t+1$, rather than t alone. We show there under some natural conditions that a unique rational expectations equilibrium exists, whose comparative statics are similar to those in the simpler myopic model. Intuitively, the network complementarities would be further amplified with farsighted behavior: higher network quality (represented by p) or historical size will induce even higher levels of entry.

3.2 Dynamics of Entry and Concentration

We first calculate the profits a new agent in any cohort with a given ability ω expects to earn upon entering a given sector, when the CTFP in that sector is expected to be A. The latter is a sufficient statistic for the specific date, sector in question, existing network size and quality (which determine CTFP as per (2)). The optimal capital size K must maximize $A\omega^{1-\alpha}K^{\alpha} - rK$, and thus satisfies:

$$\log K(\omega, A) = \log\omega + \log\phi + \frac{1}{1-\alpha}\log A - \frac{1}{1-\alpha}\log r \tag{3}$$

(where $\phi \equiv \alpha^{\frac{1}{1-\alpha}}$). The resulting profit satisfies

$$\log\Pi(\omega, A) = \log\omega + \log\psi + \frac{1}{1-\alpha}\log A - \frac{\alpha}{1-\alpha}\log r \tag{4}$$

(where $\psi \equiv \phi^{\alpha} - \phi$).

Of those new agents receiving an offer to enter this sector, the ones that will decide to enter are those whose ability exceeds a threshold $\underline{\omega}$:

$$\log\omega > \log\underline{\omega} \equiv \frac{1}{1-\sigma}[\log\frac{1}{\psi} - \frac{1}{1-\alpha}\log A + \frac{\alpha}{1-\alpha}\log r] \tag{5}$$

This threshold lies in the interior of the ability distribution if

$$(1-\alpha)\log\frac{1}{\psi} + \alpha\log r - (1-\sigma)(1-\alpha) < \log A < (1-\alpha)\log\frac{1}{\psi} + \alpha\log r \tag{6}$$

We shall assume that this is true at the beginning of the process for each sector, i.e., $\log A_0$ satisfies this inequality, and we will restrict attention to 'early phases of industrialization' when it continues to be true (i.e., until T such that $\log A_{iT}$ also satisfies it). At later dates it may fail to be true if CFTP rises sufficiently; then *all* agents will want to become entrepreneurs irrespective of how low their own ability is. Such later stages will be characterized by some slowing down of the entry process, as the range of abilities that prefer to enter ceases to expand (having reached the maximum limit). In the Chinese data we see no such tendency for the entry growth to slow down until 2014. Hence it seems reasonable to suppose that the 'early stage' assumption is adequate for our purposes. We shall continue to assume this for the rest of the paper.

Combining this with the fraction $ks_{i,t-1}$ of new agents that have an opportunity to enter (as well as expression (2) for CFTP), we can derive the volume of entry e_{it} in cohort t into B_i as a function of the state variables $n_{i,t-1}, s_{i,t-1}$:

$$e_{it} \equiv n_{it} - n_{i,t-1} = ks_{i,t-1}[1 - \frac{1}{1-\sigma}\log\frac{1}{\psi} - \frac{\alpha}{(1-\sigma)(1-\alpha)}\log r + \frac{1}{(1-\sigma)(1-\alpha)}\{\log A_0 + \theta(p)n_{i,t-1}\}] \quad (7)$$

which can be written as

$$e_{it} = ks_{i,t-1}[B' + C'\theta(p)n_{i,t-1}]$$

for constants B', C' that are functions of model parameters. This expression reduces to

$$e_{it} = Ls_{i,t-1} + \kappa(p)s^2_{i,t-1}N_{t-1} \quad (8)$$

where L denotes $k[1 - \frac{1}{1-\sigma}\log\frac{1}{\psi} - \frac{\alpha}{(1-\sigma)(1-\alpha)}\log r + \frac{1}{(1-\sigma)(1-\alpha)}\log A_0]$, and $\kappa(p)$ denotes $\frac{1}{(1-\sigma)(1-\alpha)}k\theta(p)$ which is rising in p.

Equation (8) captures different sources of dynamic increasing returns generated by the network effects. One source is the interaction between network quality $\kappa(p)$ and size N_{t-1}, also present in the one sector version of the model in Munshi (2011) where the right-hand-side is replaced by $L + \kappa(p)N_{t-1}$. Solving recursively, $e_t = (1+\kappa(p))^{t-1}(L + \kappa(p)N_0)$, and thus entry is increasing over time, more steeply in $\kappa(p)$.[10] In the current multisectoral model, this dynamic multiplier, associated with post-entry network benefits, is augmented by the additional interaction with the network effect operating through the offer process. The probability of receiving an offer is proportional to the share $s_{i,t-1}$ of the sector, thus generating an interaction of N_{t-1} with $s^2_{i,t-1}$ in the last term on the right-hand-side of (8). This means that entry into i rises *at an increasing rate* as its share $s_{i,t-1}$ among incumbents from the same origin grows. The two network effects, operating through the entry offer process and post-entry network benefits, thus complement each other. This phenomenon is more intense for origins with high connectedness p.

Aggregating (7) across the two sectors $i = 1, 2$, we obtain an expression for the dynamic of the aggregate stock of entrepreneurs operating at the end of t:

$$N_t - N_{t-1} \equiv E_t \equiv e_{1t} + e_{2t} = L + \kappa(p)N_{t-1}H_{t-1} \quad (9)$$

where $H_{t-1} \equiv s^2_{1,t-1} + s^2_{2,t-1}$ denotes the Herfindahl Hirschman Index for concentration at $t-1$. This, once again, shows the additional role of sectoral concentration as a compounder of the dynamic increasing returns generated by network size N_{t-1}. Greater sectoral concentration generates higher aggregate entry: new entrants are increasingly channelled into the dominant sector, as its share grows, where they are better able to take advantage of the increasing returns in the post-entry network benefits.

Next we derive an expression for the dynamics of concentration. Given the simplifying assumption of there being only two business sectors, the Herfindahl Hirschman Index can be expressed as a function of the share of sector 1 alone: $H_{t-1} = s^2_{1,t-1} + [1 - s_{1,t-1}]^2$. Hence it suffices to trace the evolution of sector 1's share. Observe that the concentration index is locally monotone increasing in sector 1's share if it exceeds $\frac{1}{2}$: the dominant sector becomes even more dominant. The (reciprocal of the) share at t can be expressed as[11]

$$\frac{1}{s_{1t}} = 1 + [\frac{1 - s_{1,t-1}}{s_{1,t-1}}]\{\frac{L + N_{t-1} + \kappa(p)N_{t-1}(1 - s_{1,t-1})}{L + N_{t-1} + \kappa(p)N_{t-1}s_{1,t-1}}\} \quad (10)$$

[10] In Munshi's model, the heterogeneity across populations is in the option outside business, which is embedded in L, whereas in our model it is in $\kappa(p)$.

[11] This expression is derived by manipulating the expression for the reciprocal of sector 1's share: $\frac{1}{s_{1,t}} = \frac{N_{t-1}+E_t}{n_{1,t-1}+e_{1t}}$, and then using expressions derived above for E_t, e_{1t}.

which in turn can be rewritten as

$$\frac{1}{s_{1t}} - \frac{1}{s_{1,t-1}} = [1 - \frac{1}{s_{1,t-1}}] \frac{\kappa(p) N_{t-1}(2s_{1,t-1} - 1)}{L + N_{t-1} + \kappa(p) N_{t-1} s_{1,t-1}} \tag{11}$$

From this it is evident that sector 1 once ahead (an initial share exceeding a half) will forever remain ahead. Hence concentration must rise over time. This effect is greater when there is a larger stock of incumbent firms, N_{t-1}, and is amplified in origins with higher p.[12] As discussed above, the increase in concentration raises entry and, thus, the dynamics of concentration and the dynamics of entry are mutually reinforcing.

We are now in a position to derive some comparative dynamics properties which can be tested empirically, viz, how entry and concentration varies with respect to network connectedness p and over time t. First, as explained above, concentration rises over time. Hence equation (9) implies aggregate entry rises over time. Second, at any given t, entry and concentration must be higher from origins with higher p. Observe that this must be true (given initial conditions N_0, H_0) at date $t = 1$ for entry E_1 (using (9)) and concentration (i.e., s_{11} using (10)). Applying the same argument, the same result obtains at $t = 2$: e.g., the effect of higher $\kappa(p)$ is compounded by the higher H_1, N_1.[13] And so on: induction implies the same is true for all succeeding dates.

What does the model imply about how the slope of entry or concentration with respect to p varies with t? Equivalently, is the change from t to $t+1$, and not just the level, increasing in p? Observe that from (9)

$$E_{t+1} - E_t = \kappa(p)[N_t H_t - N_{t-1} H_{t-1}] = \kappa(p)[E_t H_t + N_{t-1}(H_t - H_{t-1})] \tag{12}$$

Since $\kappa, E_t, H_t, N_{t-1}$ are all rising in p, this result would hold for entry if it were also true for concentration (i.e., $H_t - H_{t-1}$ is rising in p). A sufficient condition for the property to hold for concentration is that it holds for the share of the dominant sector (i.e., if $s_{1,t} - s_{1,t-1}$ is rising in p, since $H_t - H_{t-1} = 2(s_{1t} - s_{1,t-1})(s_{1t} + s_{1,t-1} - 1)$, and we have already shown that $s_{1t}, s_{1,t-1}$ are rising in p). However, this sufficient condition cannot hold for large enough values of t, for the simple reason that the share of the dominant sector B_1 is bounded above by 1. If it rises at a fast rate initially, it approaches 1 and then must 'flatten out'. So after a sufficient amount of time has passed, a high p origin may be almost entirely concentrated in sector B_1, while the share of this sector is continuing to rise for lower p origins. We therefore ask the question whether the result holds for concentration at early stages of the process, when the share of sector 1 is not too large. This indeed turns out to be the case: we show below that it certainly holds if this share is below 75%.

The following Proposition summarizes the comparative dynamics of entry and concentration.

Proposition 1 *(a) Entry E_t, the stock of entrepreneurs N_t and concentration H_t are rising in t (for any given p) and in p (at any given t).*

(b) $E_t - E_{t-1}$ and $H_t - H_{t-1}$ are both rising in p if $\kappa(p) < 1$ for all p and the share of the larger sector at $t-1$ is not too close to 1 (e.g., below $\frac{3}{4}$).

Proof of Proposition 1: The argument for part (a) has already been provided in the text. For part (b), we show that $s_{1t} - s_{1,t-1}$ is rising in p as long as $s_{1,t-1} < \frac{3}{4}$. Using (11), we obtain

$$s_{1t} - s_{1,t-1} = \frac{\kappa(p) N_{t-1}(2s_{1,t-1} - 1)(1 - s_{1,t-1}) s_{1t}}{L + N_{t-1} + \kappa(p) N_{t-1} s_{1,t-1}} \tag{13}$$

[12] This is because the second term on the right-hand-side of (11) is increasing in N_{t-1}.

[13] For concentration, it is straightforward to show that the right-hand-side of (10) is decreasing in $s_{1,t-1}$ and N_{t-1}, with the latter effect amplified by a larger $\kappa(p)$. Because $s_{1,t-1}$ and N_{t-1} are increasing in p, the result follows.

and then using (10):

$$s_{1t} - s_{1,t-1} = \kappa(p) N_{t-1} \frac{(2s_{1,t-1} - 1)(1 - s_{1,t-1}) s_{1,t-1}}{(L + N_{t-1})(2 - s_{1,t-1}) + \kappa(p) N_{t-1}(s_{1,t-1}^2 + 1 - s_{1,t-1})} \tag{14}$$

It is easily checked that $\kappa(p) < 1$ implies that the denominator of the right-hand-side of (14) is decreasing in $s_{1,t-1}$. And the numerator is increasing in $s_{1,t-1}$ if $s_{1,t-1} < \frac{3}{4}$ (since this implies $s_{1,t-1}(1 - s_{1,t-1}) > \frac{1}{6}$). Then $s_{1t} - s_{1,t-1}$ is rising in $s_{1,t-1}$, as well as in N_{t-1} and κ. The result then follows from the fact that $s_{1,t-1}, N_{t-1}$ are rising in p. ■

3.3 Firm Size Dynamics

Next we turn to predictions concerning entrepreneurial ability, productivity and capital size of firms. Observe that the ability threshold for entrants is decreasing in CTFP. Hence marginal entrants in high p areas will be of lower ability. Substituting from (5) in (3), initial log capital size of the marginal entrant is

$$\log K_{it}^m = U - \frac{\sigma}{(1-\sigma)(1-\alpha)} \log A_{it} \tag{15}$$

where $U \equiv \log \phi - \frac{1}{1-\sigma} \log \psi - \frac{1}{1-\alpha} \log r$, and $\log A_{it}$ denotes log CFTP at t ($= \log A_0 + \theta(p) n_{i,t-1}$). This too is decreasing in CFTP and thus falling over time. To understand this, observe that there are two conflicting effects of a higher CFTP: the direct effect would induce a higher firm size by raising firm level TFP, but this would be offset by the negative selection on ability, which lowers firm TFP and size. In our model the latter effect dominates. The reason is that the TFP $A\underline{\omega}^{1-\alpha}$ of the marginal entrepreneur is decreasing in CTFP A:

$$\log A + (1-\alpha) \log \underline{\omega} = -\frac{\sigma}{1-\sigma} \log A + \frac{1-\alpha}{1-\sigma} [\log \frac{1}{\psi} + \frac{\alpha}{1-\alpha} \log r] \tag{16}$$

This also implies that at any given t, high p regions are characterized by firms of smaller size and productivity at the lower end. The preponderance of small and seemingly unproductive firms often noted in developing countries may just be a manifestation of strong network effects! According to our model, their own productivity understates their contribution via spillovers to their network.

Next consider size of the average (rather than marginal) entrant. Substituting from (5) in (3), and noting that the average entrant has (log) ability $\frac{1 + log \omega_{it}}{2}$, the log capital size of the average entrant is:

$$\log K_{it}^a = W + \frac{1 - 2\sigma}{2(1-\alpha)(1-\sigma)} \log A_{it} \tag{17}$$

where $W \equiv \log \phi + \frac{1}{2} + \frac{1}{2(1-\sigma)} \log \frac{1}{\psi} - \frac{2-\alpha-2\sigma}{2(1-\alpha)(1-\sigma)} \log r$. Expression (17) shows that higher CFTP is also negatively correlated with the size of the average entrant, if and only if $\sigma > \frac{1}{2}$. The value of σ now needs to be high enough, whereupon the drop in the minimum ability threshold (when CFTP rises) is steep enough to outweigh the direct positive effect of CFTP. The model therefore allows network effects to generate lower average firm size and productivity among entrants, which accompanies the larger entry flows among high p community origins. It implies that the average size of entrants from cohort t at entry date t is falling in p, and across successive cohorts (for given p). Additionally, Proposition 2 below verifies that the decline in initial capital with respect to p also becomes steeper across successive cohorts.

However, subsequent growth of firm size for any given cohort exhibits the opposite pattern: while high p communities enter with smaller firms on average, they grow faster subsequently. This is because the negative selection, which pertains to initial sizes, does not apply to subsequent growth rates, which are fueled by the entry of subsequent cohorts. To see this, observe that (3) implies that the capital at date $t' > t$ of an average cohort t entrepreneur is given by

$$\log K^a_{itt'} = \frac{1 + \log \underline{\omega}_{it}}{2} + \log \phi - \frac{1}{1-\alpha} \log r + \frac{1}{1-\alpha} [\log A_{it} + \theta(p) \sum_{l=t}^{t'-1} e_{il}] \tag{18}$$

implying a growth rate at period t' which is rising in p and entry into B_i at t':

$$\log K^a_{it,t'} - \log K^a_{it,t'-1} = \frac{1}{1-\alpha} \theta(p) e_{it'} \tag{19}$$

Proposition 2 *(a) Initial capital size of marginal entrants (and of average entrants if $\sigma > \frac{1}{2}$) in cohort t is decreasing in p, and decreasing across successive cohorts for any p, in every sector. Averaging across sectors, the initial capital size of marginal entrants (and of average entrants if $\sigma > \frac{1}{2}$) is decreasing more steeply in p across successive cohorts.*

(b) Averaging across sectors, the growth rate of capital size of incumbent entrepreneurs of any past cohort t from $t' - 1 (> t)$ to t' is rising in t' and in p.

Proof of Proposition 2: The first result in part (a) follows from (15) and (17), as $\log A_{it}$ is rising in p and in t for every i. The decline in initial log capital size of marginal entrants from cohort $t-1$ to t in sector B_i equals $\frac{\sigma}{(1-\sigma)(1-\alpha)} \theta(p) e_{it}$. Averaging across sectors, this is proportional to $\frac{\sigma}{(1-\sigma)(1-\alpha)} \theta(p) E_t$, which is increasing in p by Proposition 1. A similar argument applies for average entrants size if $\sigma > \frac{1}{2}$. This establishes the second result in (a). Finally, (b) follows from equation (19), as aggregate entry E'_t is rising in p and in t'. ∎

4 Empirical Analysis

4.1 Measuring Social Connectedness

One reason why economic networks are often organized within pre-existing social groups is that members of these groups are connected to each other in many ways. Exclusion from social interactions, for example, is an effective sanctioning device, which can be used to support trust and economic cooperation. News about deviations from cooperative behavior travels faster through the population when individuals interact more frequently with each other and across a wide range of social partners (Kandori, 1992; Ellison, 1994). This is the underlying logic for why groups that are more "socially connected" or "tightly knit" can support higher levels of cooperation; e.g. Granovetter (1985), Coleman (1988), Glaeser, Laibson, and Sacerdote (2002).

We assume that all residents of a county belong to the same social group. Everything else equal, social interactions within the group are increasing in spatial proximity, which, in turn, is increasing in population density. Migrant entrepreneurs remain within the group. If they move back and forth between the home county and their place of work, then they can be sanctioned directly. If they move permanently, then sanctions can be applied to the family that remains behind.[14] We thus use population density in the entrepreneur's birth county to measure social connectedness in his network (the θ parameter in the model).

While this reasoning makes sense for entrepreneurs born in counties, it may not for those who are born in cities for two reasons. First, if there is an upper bound to an individual's social interactions, then population density may not have a positive effect beyond a certain point. Based on the 1982 population census, the population density in Chinese counties was 30 individuals per square km. on average, whereas the corresponding

[14] Munshi and Rosenzweig (2016) make the same argument in the context of rural-urban migration in India.

statistic in the cities was 140 individuals per square km. Spatial proximity thus may not be a constraint to social interactions in the city. Second, multiple social groups co-exist in a city. If the number of groups is increasing sufficiently steeply in population density and social interactions do not take place exclusively within a single group, then the frequency of within-group interactions (which is what matters for cooperation) could even be declining with population density.

Table 1a provides empirical support for our distinction between county-born and city-born entrepreneurs by comparing social interactions in counties and cities. The family module of the China Family Panel Survey (2010) includes the frequency of social interactions between the primary respondent and local residents; i.e. individuals living in the local area. The social interactions are divided into entertainment, visits, and chatting.[15] Population density in the county and in the urban district (for city residents), is derived from the 1982 population census and is measured as a Z-score. The China Family Panel Survey is designed to survey a representative sample of households drawn from both counties and cities. We see in Table 1a that the frequency of visits and chatting is increasing significantly in population density for county residents, whereas the sign of this relationship is reversed for city residents.

Table 1b, which is based on the adult individual module of the China Family Panel Survey (2010) looks at who the respondents chat with most, distinguishing between local residents, relatives, classmates, and colleagues. In counties, higher population density is associated with a significantly increased likelihood that the respondent reports chatting most with a local resident or a colleague and a significantly reduced likelihood of chatting with a relative. In cities, in contrast, higher population density is associated with a decline in the likelihood that the respondent chats most with a local resident and an increase in the likelihood of chatting most with classmates and colleagues, although none of the effects are significant at conventional levels.

Table 1a. Frequency of Local Social Interactions and Population Density

Dependent variable:	number of interactions per month with local residents					
Family's location:	county			city		
Purpose:	entertainment	visits	chat	entertainment	visits	chat
	(1)	(2)	(3)	(4)	(5)	(6)
Population density	0.208	2.006	5.420***	-0.065	-0.408*	-0.343
	(0.335)	(1.268)	(1.808)	(0.102)	(0.237)	(0.523)
Constant	1.246***	3.613***	18.013***	0.999***	3.391***	13.595***
	(0.193)	(0.724)	(0.930)	(0.216)	(0.442)	(1.158)
Observations	8,675	8,675	8,675	3,256	3,256	3,256

Source: family module of China Family Panel Survey (2010).
Population density in the county or city (urban district) based on the 1982 population census.
Population density is measured in units of 10,000 people per square Km, and then converted to Z-score.
Standard errors clustered at the birthplace level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

The differences in social interactions that we have documented map into differences in trust, which is closely associated with economic cooperation. The adult individual module of the China Family Panel Survey (2012) collected information on trust in local residents and in outsiders; i.e. individuals living outside the local area. Trust is measured as an ordinal variable, taking values from 0 to 10. We see in Table 2, Column 1 that trust in local residents is increasing in population density for respondents in counties. However, population density has no impact on trust in outsiders for those respondents, nor does it have an impact on trust in either local residents or outsiders for city residents, in Columns 2-4. The results on social interactions and trust,

[15]Entertainment includes playing Mahjong or cards, reading newspapers, listening to radio,and watching TV. Chatting is defined as a face-to-face interaction without any other activity.

Table 1b. Interaction Partners and Population Density

Dependent variable:	indicates whether the respondent chats with most with a							
Respondent's location:	county				city			
Interaction partner:	local resident	relative	classmate	colleague	local resident	relative	classmate	colleague
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population density	0.063**	-0.077**	-0.002	0.064***	-0.011	-0.001	0.006	0.007
	(0.029)	(0.036)	(0.017)	(0.009)	(0.008)	(0.013)	(0.004)	(0.005)
Constant	0.223***	0.515***	0.146***	0.077***	0.129***	0.605***	0.154***	0.080***
	(0.016)	(0.018)	(0.009)	(0.005)	(0.013)	(0.019)	(0.009)	(0.009)
Observations	20,357	20,357	20,357	20,357	7,070	7,070	7,070	7,070

Source: adult individual module of China Family Panel Survey (2010).
Dependent variables indicate whether the respondent chats most with a local resident, a relative, a classmate, and a colleague, respectively, on a daily basis.
Excluded interaction partners include interactions with social workers, babysitters, property managers, and teachers.
Population density in the county or city (urban district) based on the 1982 population census.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Standard errors clustered at the birthplace level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

taken together, indicate that population density is a valid measure of social connectedness for entrepreneurs born in counties. The tests of the model will, therefore, be applied to them. Entrepreneurs born in cities will also be used, for placebo tests, in the analysis that follows.

Table 2. Trust and Population Density

Respondent's location:	county		city	
Dependent variable:	trust in local residents	trust in outsiders	trust in local residents	trust in outsiders
	(1)	(2)	(3)	(4)
Population density	0.482***	0.027	0.001	0.015
	(0.153)	(0.208)	(0.056)	(0.046)
Constant	6.610***	2.202***	6.312***	2.180***
	(0.082)	(0.109)	(0.097)	(0.090)
Observations	19,637	19,637	6,248	6,248

Source: adult individual module of China Family Panel Survey (2012).
Trust is an ordinal variable, and takes value from 0 to 10.
Population density in the county or city (urban district) based on the 1982 population census.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Standard errors clustered at the birthplace level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

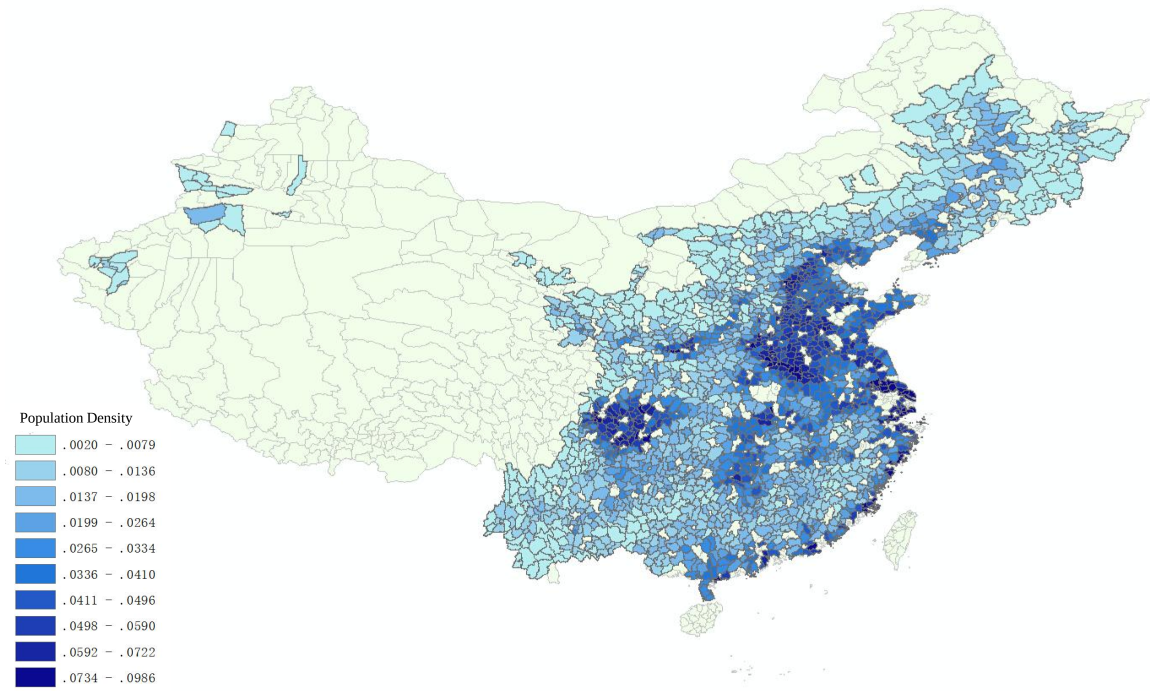
Population density must satisfy two conditions to be a valid proxy for social connectedness. First, it must be positively associated with social connectedness. We have verified that this is the case for entrepreneurs born in counties. Second, there must be sufficient variation in population density across counties to test the model's predictions. Figure 3 plots population density across Chinese counties, excluding counties below a threshold density.[16] The population density in this figure, and in all the analyses, is derived from the 1982 population census. This is before the first wave of privatization in the early 1990's and the accompanying rural-urban labor migration.[17] There is substantial variation in this pre-determined measure of social connectedness,

[16]Population density is measured in units of 10,000 people per square km. The threshold density is set at 0.002; i.e. 20 people for square km. This excludes sparsely populated regions such as Western China, Inner Mongolia, and Tibet, which supply few entrepreneurs, from the analysis.

[17]The advantage of using a predetermined measure of social connectedness in the analysis is that it avoids the possibility that changes in population density in later time periods are generated by endogenously determined network-based migration. We nevertheless verified, using successive rounds of the population census, that the ranking of counties with respect to population density was time invariant.

which ranges from 20 to 1000 people per square km across counties, and we will see that the tests of the model's predictions consequently having sufficient statistical power. Nevertheless, population density is not a statistical instrument for social connectedness. It is possible that population density is (accidentally) correlated with other direct determinants of entry, concentration, and firm size. We will account for this possibility in different ways in the analysis that follows.

Figure 3. Population Density across Counties



Source: 1982 population census.

4.2 Firm Data

The primary data source that we use to test the model is the firm registration database maintained by the State Administration of Industry and Commerce (SAIC). We have received special permission to use the entire database, with firm and individual identifiers, for our research. The following information is available for each firm: establishment date (and exit date if relevant), 4-digit sector, location, ownership-type, registered capital (initial and subsequent changes), and the list of major shareholders and managers, with their citizenship ID. Although the database includes the universe of Chinese firms registered since 1980 in China, the analysis that follows is restricted to private firms and covers the 1990-2009 period (up to the financial crisis). Starting with a relatively small number of private firms in 1990, there were 7.3 million registered private firms in 2009. Our unique data thus allows us to trace the growth of private enterprise in China in its entirety. In contrast, previous analyses of firms in China have relied on a publicly available database of manufacturing firms with sales above a threshold level (5 million Yuan) over the shorter 1998-2008 period; e.g. Hsieh and Klenow (2009), Song, Storesletten, and Zilibotti (2011), Brandt, Van Biesebroeck, and Zhang (2012), Aghion, Cai, Dewatripont, Du, Harrison, and Legros (2015). These firms account for less than 15% of the private firms in the registration database in 2008.

Although the registration database is well suited to examine entry, sectoral choice, and initial capital investments, it is not as useful for analyses of capital growth. Registered capital does change, but given the administrative costs involved, these changes will not track perfectly with changes in the firm's assets over time. For the analysis of firm growth, we thus turn (separately) to the industrial census, which was conducted in 1995, 2004, and 2008 and the SAIC's inspection database, which includes annual firm-level information on assets and sales from 2004 onwards.

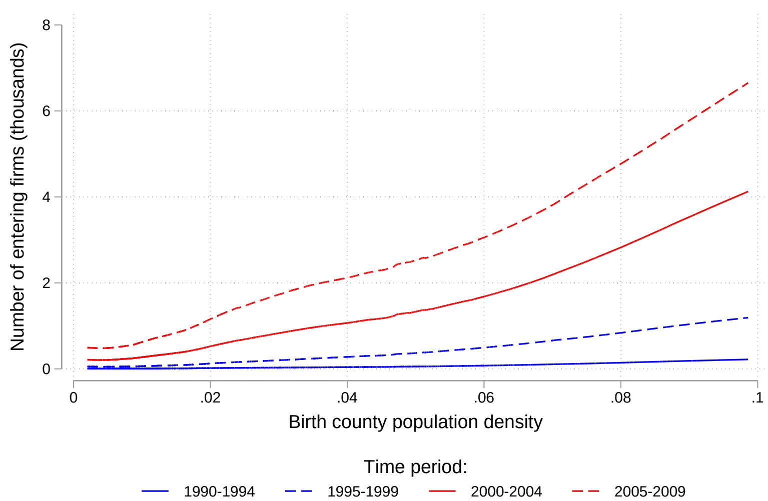
To verify the quality of the SAIC administrative data, which are used for much of the analysis, we linked firms by their ID's across the industrial census and the registration and inspection databases. The correlation in reported firm assets between the 2008 industrial census and the inspection database is 0.63. The correlation in the total number of firms, by sector and birth county of the entrepreneurs, in the 2008 industrial census and the registration database is 0.85. The SAIC data are reported by firms. The industrial census information is collected by enumerators. Despite the fact that the data are collected independently, there is a relatively high degree of consistency across the difference data sources that we use in the analysis.

4.3 Evidence on Firm Entry

The model predicts that firm entry is (i) increasing in social connectedness at each point in time, (ii) increasing over time, and (iii) increasing more steeply in social connectedness over time. This is a statement about the flow of firms rather than the stock, and so we will measure entry in five-year windows over the 1990-2009 period. As discussed, population density proxies for social connectedness for entrepreneurs born in counties but not in cities. The analysis that follows thus focuses on their firms, although we will later examine the relationship between population density and the model's outcomes with a sample of entrepreneurs born in cities as a placebo test. Because social connectedness is not increasing in population density in the urban sample, as shown above, we do not expect to validate the predictions of the model with this sample. Note that the analysis places no restrictions on where firms locate; there are 3,235 counties and urban districts where firms locate in our data and they are included in all the analysis.

Figure 4 reports nonparametric estimates of the relationship between the entry of firms from each birth county in five-year windows and 1982 population density in that county. As predicted by the model, entry is (i) increasing in population density at each point in time, (ii) increasing over time, and (iii) increasing more steeply in population density over time.

Figure 4. Firm Entry



Source: SAIC registration database and 1982 population census.

Table 3 reports parametric estimates corresponding to Figure 4, separately by time period. This allows us to test the prediction that entry is increasing in population density at each point in time. Although population density may be positively correlated with social connectedness in a county, it could potentially be correlated with other variables that directly determine the model's outcomes. We deal with this concern in different ways. One approach is to control for some of the most obvious correlates of population density in the estimating equation. Population density is positively associated with the county's population and a larger population is mechanically associated with greater entry. Counties with higher population densities also have more educated populations, based on literacy rates from the 1982 census. If education is a proxy for ability, then there will be greater entry in more educated counties because there is positive selection on ability into entrepreneurship in the model. We thus include 1982 population and 1982 literacy in an augmented specification of the estimating equation in Table 3. Population density, measured in 1982, has a positive and significant effect on entry in each time period. the augmented specification, both population and education have a positive and significant (with one exception) effect on entry in each time period. However, the pattern of coefficients on the birth county population density variable remains unchanged.

Table 3. Firm Entry (with controls)

Dependent variable:	number of entering firms							
Time period:	1990-1994		1995-1999		2000-2004		2005-2009	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth county population density	0.028***	0.021***	0.176***	0.128***	0.592***	0.393***	1.008***	0.531***
	(0.004)	(0.004)	(0.020)	(0.022)	(0.063)	(0.060)	(0.103)	(0.083)
Birth county population	–	0.038***	–	0.290***	–	1.211***	–	3.002***
		(0.011)		(0.061)		(0.171)		(0.298)
Birth county education	–	0.001***	–	0.003***	–	0.007***	–	0.003
		(0.000)		(0.001)		(0.002)		(0.003)
Constant	0.031***	-0.037***	0.208***	-0.089*	0.787***	-0.231*	1.560***	-0.021
	(0.002)	(0.011)	(0.011)	(0.050)	(0.031)	(0.139)	(0.048)	(0.220)
Observations	1,624	1,624	1,624	1,624	1,624	1,624	1,624	1,624

Note: number of entering firms from each birth county in each time period is measured in thousands.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Population is measured in millions and education is measured by the percent of the population that is literate.
Number of firms is obtained from the SAIC registration database and population density, population and education are derived from the 1982 population census.
Robust standard errors are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Another explanation for variation in entry across birth counties is that entrepreneurs born in high population density counties fortuitously have preferred access to sectors or destinations that grew relatively fast. We account for this by taking advantage of the fact that firms from multiple birth counties will select into a single sector or a single destination. This allows us to include sector fixed effects and destination fixed effects in the estimating equation in Table 4. This equation is estimated separately in each time period and so the fixed effects capture the changing fortunes of sectors and destinations over time.[18] Birth county population density continues to have a positive and significant effect on entry in each time period in Table 4. A comparison of the results obtained with the benchmark specification in Table 3 and the specification in Table 4 indicates that the inclusion of the fixed effects actually increases the point estimates. This indicates that entrepreneurs born in high population density counties select sectors and destinations that had relatively *low* entry on average.

[18]Entry in Table 4 is measured at the birth county-sector-destination level in each time period. The number of entrants is multiplied by the county-specific product of the number of sectors and the number of destinations so that the dependent variable reflects average entry at the level of the county.

Table 4. Firm Entry (with sector and destination fixed effects)

Dependent variable:	number of entering firms			
Time period:	1990-1994	1995-1999	2000-2004	2005-2009
	(1)	(2)	(3)	(4)
Birth county population density	0.050***	0.344***	1.123***	1.749***
	(0.009)	(0.081)	(0.233)	(0.278)
Mean of dependent variable	0.0725	0.483	1.673	3.024
Observations	1,085,169	1,085,169	1,085,169	1,085,169
Sector Fixed Effects	Yes	Yes	Yes	Yes
Destination Fixed Effects	Yes	Yes	Yes	Yes

Note: firm entry, in thousands, measured at the birth county-time period-sector-destination level.
For a given birth county, all sectors and destinations that ever have entrants are included in all time periods (assigned zero entry where necessary). To adjust for differences in the number of sectors and destinations across birth counties, the number of entrants is multiplied by the number of sectors X the number of destinations.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Population is measured in millions and education is measured by the percent of the population that is literate.
Number of firms is obtained from the SAIC registration database and population density, population and education are derived from the 1982 population census.
Standard errors clustered at the birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

4.4 Evidence on Concentration

The model predicts that sectoral concentration, measured by the Herfindahl Hirschman Index (HHI) is (i) increasing in social connectedness at each point in time, (ii) increasing over time, and (iii) increasing more steeply in social connectedness over time. As with firm entry, we proxy for social connectedness with birth county population density, restricting attention to entrepreneurs born in counties.

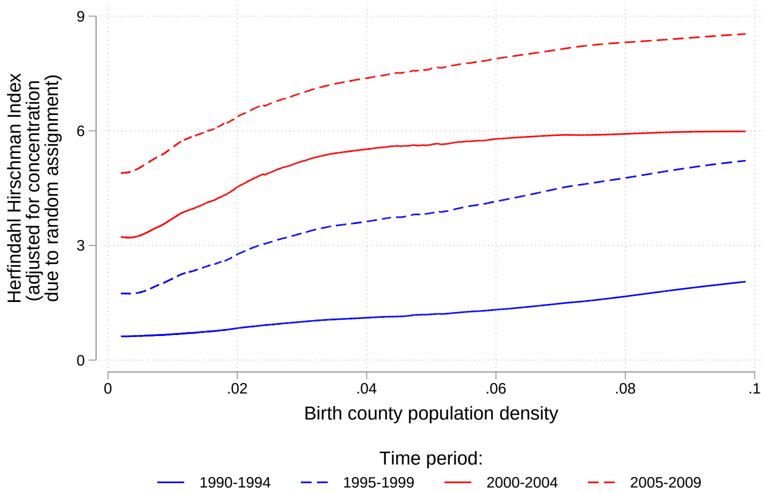
Figure 5 reports nonparametric estimates of the relationship between sectoral concentration at the two-digit level and 1982 population density in the birth county in five-year windows over the 1990-2009 period. The HHI is adjusted for the fact that measured concentration could vary with the number of firms and the number of sectors just by chance, using a normalization that is derived in Appendix B.[19] The adjusted HHI is (i) increasing in population density at each point in time, (ii) increasing over time, and (iii) increasing more steeply in population density over time.

Table 5 reports parametric estimates corresponding to Figure 5, separately by time period. Although the link between sectoral concentration and the county's population and education is not immediately obvious, we include these variables in an augmented specification, just as we did with firm entry as the dependent variable in Table 4. Population density in the birth county has a positive and significant effect on (adjusted) sectoral concentration in each time period. These results are obtained with or without the inclusion of county population and education in the estimating equation.

Although the firm's location plays no role in the model, it is possible that the network spillovers are (in part) geographically constrained. If that is the case, then the model's predictions for sectoral concentration apply, without modification, to spatial concentration (within sectors) for firms that are located outside their birth county. Figure 6 reports nonparametric estimates of the relationship between spatial concentration, within one-digit sectors, and population density in the birth county in five-year intervals. As with the analysis of sectoral concentration, the spatial concentration within each sector for a given birth county is adjusted for the number of firms that are located outside the birth county and the number of external destinations, which

[19] Previous attempts to examine the spatial distribution of production; e.g. Ellison and Glaeser (1997), Duranton and Overman (2005), have also taken account of this feature of all concentration statistics.

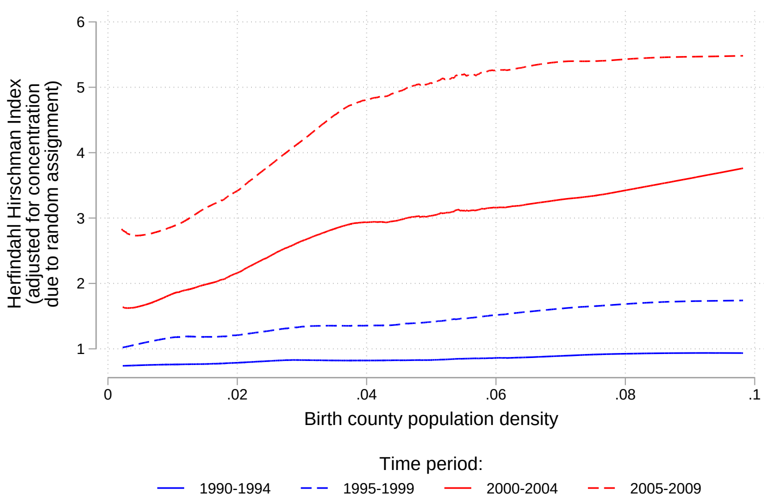
Figure 5. Sectoral Concentration



Source: SAIC registration database and the 1982 population census.
Sectoral concentration measured by the Herfindahl Hirschman Index (HHI) across two-digit sectors divided by the expected HHI that would be obtained by random assignment, given the number of entrants and the number of sectors at each point in time.

would generate variation in the measured HHI just by chance.[20] Matching the predictions of the model, the spatial HHI is (i) increasing in birth county population density in each time period, (ii) increasing over time, and (iii) increasing more steeply over time.

Figure 6. Spatial Concentration, within Sectors



Source: SAIC registration database and 1982 population census.
Spatial concentration, within one-digit sectors, is measured by the Herfindahl Hirschman Index (HHI) across destinations (outside the birth county) divided by the expected HHI that would be obtained by random assignment, given the number of entrants and number of destinations at each point in time.
We remove the cell with only 1 entrants. We only include the origin county-sector which have positive entrants in all 4 period.

Table 6 collects results with all three outcomes; i.e. firm entry, sectoral concentration, and spatial concentration. We have already documented a positive and significant relationship between birth county population

[20]To maintain consistency across time periods, we only include birth county-sectors that have multiple entrants in all time periods. This is not a constraint in the sectoral analysis because all birth counties have multiple entrants in each time period.

Table 5. Sectoral Concentration

Dependent variable:	adjusted HHI across sectors							
Time period:	1990-1994		1995-1999		2000-2004		2005-2009	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth county population density	0.273***	0.186***	0.840***	0.493***	0.850***	0.448***	0.964***	0.669***
	(0.022)	(0.027)	(0.059)	(0.065)	(0.054)	(0.055)	(0.071)	(0.073)
Birth county population	–	0.480***	–	2.034***	–	2.332***	–	1.675***
		(0.104)		(0.252)		(0.238)		(0.274)
Birth county education	–	0.009***	–	0.017***	–	0.024***	–	0.024***
		(0.001)		(0.004)		(0.004)		(0.006)
Constant	0.933***	0.157	2.978***	0.923***	4.668***	2.032***	6.564***	4.254***
	(0.017)	(0.106)	(0.051)	(0.277)	(0.054)	(0.342)	(0.066)	(0.442)
Observations	1,624	1,624	1,624	1,624	1,624	1,624	1,624	1,624

Note: sectoral concentration measured by the Herfindahl Hirschman Index (HHI) across two-digit sectors divided by the expected HHI that would be obtained by random assignment, given the number of entrants and the number of sectors at each point in time.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Population is measured in millions and education is measured by the percent of the population that is literate.
The sector of each firm is obtained from the SAIC registration database and population density, population and education are derived from the 1982 population census.
Robust standard errors are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

density and both entry and concentration, in each time period. The focus now is on testing the model's predictions for changes over time and for changes in the birth county population density effect over time. Data from all time periods are pooled and the estimating equation now includes birth county population density, time period, and the interaction of these variables. To preserve space, we only report the coefficient on the time period variable and the interaction coefficient in Table 6.

Table 6. Changes over Time and Interaction Effects

Birth location:	county			city district		
Dependent variable:	number of entrants	sectoral HHI	spatial HHI	number of entrants	sectoral HHI	spatial HHI
	(1)	(2)	(3)	(4)	(5)	(6)
Time period	0.517***	1.858***	1.381***	0.661***	0.366***	1.886***
	(0.016)	(0.021)	(0.036)	(0.026)	(0.008)	(0.070)
Birth place population density × time period	0.353***	0.208***	0.306***	0.355***	-0.036***	-0.148***
	(0.029)	(0.022)	(0.034)	(0.041)	(0.005)	(0.033)
Observations	6,496	6,496	9,848	3,224	3,224	8,590

Note: the estimating equation includes, in addition, birthplace population density and a constant term.
Number of entering firms from each birth county in each time period is measured in thousands.
Sectoral concentration measured by Herfindahl Hirschman Index (HHI) across two-digit sectors divided by the expected HHI that would be obtained by random assignment.
Spatial concentration, within one-digit sectors, is measured by the Herfindahl Hirschman Index (HHI) across destinations (outside the birth county) divided by the expected HHI that would be obtained by random assignment.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Time period is an ordinal variable taking value from 1 to 4 corresponding to successive five-year time windows over the 1990-2009 period.
Number of entrants and concentration statistics are derived from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birthplace level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

The model predicts that entry and concentration should be increasing over time and that the interaction coefficient should be positive with both outcomes. Restricting the sample to county-born entrepreneurs in Table 6, Columns 1-3, the time period coefficient and the interaction coefficient are positive and significant with the number of entrants, sectoral concentration, and spatial concentration as the dependent variables, as predicted by the model. As a placebo test, we restrict the sample to entrepreneurs born in urban districts in Table 6, Columns 4-6. Population density is not positively associated with social connectedness in cities and thus we do not expect to find support for the model's predictions with this sample. Reassuringly, the interaction coefficient is *negative* and significant with sectoral concentration and spatial concentration as the dependent variables, contrary to the predictions of the model.

4.5 Testing the Mechanism

In our model, initial entry generates subsequent entry through a compounding effect over time, which is reinforced by increasing concentration in particular sectors and destinations. These dynamic increasing returns to initial entry are stronger for firms from more socially connected origins; i.e. from higher population density birth counties.

We test this mechanism in Table 7 by estimating the effect of initial entry on subsequent entry. Entry is measured at the birth county-sector-destination level, restricting attention to destinations outside the birth county. Initial entry is measured in the 1990-1994 period, when private firms were just starting to emerge, and subsequent entry is measured separately in 2000-2004 and 2005-2009. The estimating equation includes birth county-sector fixed effects.[21] The benchmark specification, reported in Columns 1-2, includes, in addition, the total number of initial entrants, regardless of their birth county, in each sector-destination in the estimating equation. The number of initial entrants from the birth county in a given sector-destination has a positive and significant effect on the number of subsequent entrants; one additional initial entrant generates seven additional entrants in the 2000-2004 period and nine additional entrants in the 2005-2009 period. Conditional on the number of initial entrants from the birth county, the total number of initial entrants in a given sector-destination has no effect on subsequent entry from that birth county in that sector-destination. This last result provides empirical support for the key assumption in the model that the birth county is the domain within business networks are organized in China and that these networks operate independently. It also provides support for the assumption that individual networks cannot influence the price of the product. If the members of the network could collude (depending on their market share) or there were limits to market size, then the total number of entrants, conditional on the number of entrants from the birth county, would also be relevant.[22]

It has been argued that Chinese cities are too small (Au and Henderson, 2006a). This stylized fact has been explained by restrictions on migration due to China's *hukou* system (Au and Henderson, 2006b; Desmet and Rossi-Hansberg, 2013) and by competition between local governments, giving rise to multiple production clusters in the same sector (Long and Zhang, 2012). Our model and the supporting empirical evidence provide an alternative explanation for China's low spatial concentration, which is that agglomeration effects are restricted to firms from the same birth county. Although firms from a given birth county will concentrate, in particular sectors and locations, over time, there is no tendency for firms from different birth counties to agglomerate. There would appear to be gains from expanding the scope of inter-firm spillovers, and recent experimental evidence indicates that this is indeed the case (Cai and Szeidl, 2016).

While the model predicts that initial entry will have a compounding effect on future entry when networks are active, it also generates the stronger prediction that the effect of initial entry should be larger for higher population density birth counties. We test this prediction in Table 7, Columns 3-4 by interacting the number of initial entrants from the birth county, within sector-destinations, with its population density. The interaction coefficient is positive and significant, as predicted by the model, for entry in 2000-2004 and in 2005-2009 as the dependent variable.

[21] All destinations outside the birth county which had a positive number of entrants by 2000-2004 and 2005-2009, respectively, for a given birth county-sector are included in the estimating equation.

[22] If all firms from the same origin collude, but there is competition across (origin-based) networks at the destination, then we would expect to see a negative effect of entry from other origins. While pricing may be non-competitive in China (see, for example, Brooks, Kaboski, and Li (2016)), the origin-based networks do not appear to be directly associated with these distortions.

Table 7. Initial Effects

Dependent variable:	subsequent entrants from the birth place							
Birth location:	county		county		county		city district	
Time period:	2000-2004	2005-2009	2000-2004	2005-2009	2000-2004	2005-2009	2000-2004	2005-2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Initial entrants from the birth place	7.120***	8.935***	5.928***	6.756***	5.895***	6.705***	7.768***	6.928***
	(0.711)	(0.972)	(0.844)	(1.063)	(0.842)	(1.058)	(0.958)	(0.978)
All initial entrants	0.054	-0.020	–	–	–	–	–	–
	(0.050)	(0.057)						
Initial entrants from the birth place × birth place population density	–	–	1.843**	2.928**	1.856**	2.947**	-0.067	-0.432**
			(0.838)	(1.283)	(0.836)	(1.281)	(0.238)	(0.218)
Distance to the birth place	–	–	–	–	-0.002***	-0.002***	-0.003***	-0.003***
					(0.000)	(0.000)	(0.000)	(0.000)
Observations	413,452	804,918	413,452	804,918	413,452	804,918	313,520	449,207

Note: number of entrants is measured at the (two-digit) sector-destination level.
All specifications include birth place-sector fixed effects.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Number of entrants is obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birthplace-sector level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Although we assume in the model that initial entry is randomly assigned, entrepreneurs will, in practice, locate relatively close to their home counties. The persistence in this distance effect could explain the correlation between initial entry and subsequent entry. If locations that are proximate to high population density counties grew relatively fast, then this would also explain why the interaction of birth county population density and initial entry has a positive effect on subsequent entry. We address this concern in two ways. First, we control directly for distance in the estimating equation. Second, we check whether locations that were initially selected by entrepreneurs from high population density counties grew relatively fast.

Table 7, Columns 5-6 include the distance from the birth county to the destination under consideration in the estimating equation. The coefficient on the distance variable is negative and significant, indicating that transportation costs are salient when deciding where to locate outside the birth county. Notice, however, that the inclusion of the distance variable has virtually no impact on the estimated coefficients on the initial entrants and on the interaction of the initial entrants with population density in the birth county. These effects are evidently not being driven by transportation costs. As a placebo test, we restrict the sample in Table 7, Columns 7-8 to entrepreneurs who are born in cities. Population density is not positively associated with social connectedness in cities and thus we do not expect the interaction of initial entry and population density in their birth locations to have a positive effect on subsequent entry. The distance effect continues to be negative and significant for city-born entrepreneurs. However, while initial entry also continues to have a positive and significant effect on subsequent entry, the coefficient on the interaction term is *negative* (and significant with 2005-2009 entry as the dependent variable).

As a final check, we verify that the destinations that were selected by the initial entrepreneurs from high population density counties did not have a larger number of total entrants (from all origins) in subsequent periods. For this test, a counter-factual total entry variable is constructed for each birth county, within each sector-time period. This variable is a weighted average of the total number of entrants in that sector-time period across all destinations, where the weight is the fraction of initial entrants from the birth county (in that sector) in each of those destinations. The counter-factual total entry variable is regressed on birth county population density, with sector fixed effects, separately in each time period. The coefficient on the birth county population density variable in Table 8 is small in magnitude and statistically insignificant in each time period. This implies that entrepreneurs from high population density counties did not fortuitously select initially into destinations that subsequently grew relatively fast and received many entrants (from different origins).

Table 8. Counter-Factual Total Entry

Dependent variable:	Counter-factual total number of entrants			
Time period:	1990-1994	1995-1999	2000-2004	2005-2009
	(1)	(2)	(3)	(4)
Birth county population density	-0.001	0.014	0.027	0.012
	(0.003)	(0.010)	(0.020)	(0.024)
Constant	0.088***	0.252***	0.599***	0.800***
	(0.003)	(0.009)	(0.018)	(0.024)
Observations	8,811	8,279	8,565	8,573

Note: counter-factual total firm entry, in thousands, is a weighted average of the total entry(from all origins) across all destinations in a given time period, with the weight based on initial (1990-1994) entry from the birth county.
All specification include sector fixed effects.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Number of entrants is obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

4.6 Evidence on Firm Size

The model predicts that the initial capital of the marginal entrant is (i) decreasing in social connectedness at each point in time, (ii) decreasing over time, and (iii) decreasing more steeply in social connectedness over time. If the negative selection on ability that accompanies a stronger network dominates the positive productivity effect of that network for inframarginal firms, then the preceding predictions apply to the average entrant as well. However, only the positive network productivity effects are relevant for growth rates.

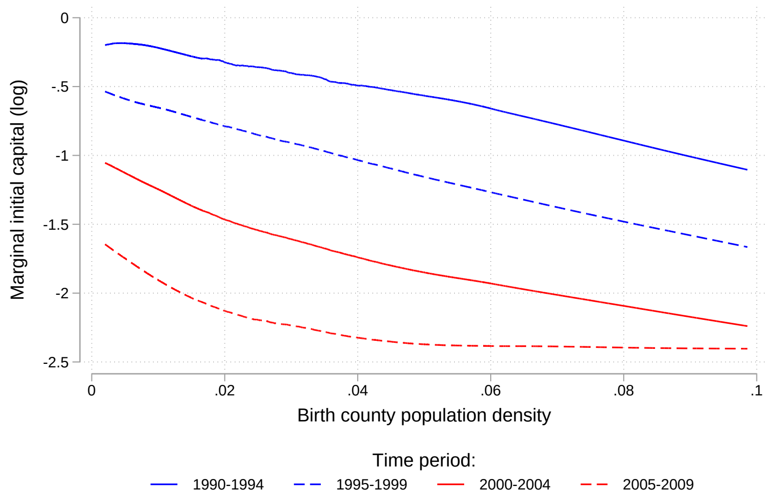
Figure 7 reports nonparametric estimates of the relationship between marginal initial capital, measured in logs, and 1982 population density in the birth county in five-year windows over the 1990-2009 period. Marginal initial capital is defined as the bottom one percentile of the initial capital distribution at the birth county-sector-time period level.[23] As predicted by the model, marginal initial capital is decreasing in birth county population density in each time period, and decreasing over time. A similar pattern is obtained with average initial capital (not reported).[24]

Notice from Figure 7 that the decline in initial capital with birth county population density does not grow steeper over time (as implied by the model). Figure 7 is designed to be consistent with Figures 4-6, where entry and concentration are the outcomes. However, the empirical analysis of capital size must also take account of the fact that the capital requirement will vary across sectors, particularly since the sectoral composition could vary with birth county population density. Table 9 allows for this possibility by studying the change in the ability of entering entrepreneurs and their capital investments over time within birth county-sectors. Recall that the decline in initial capital of the marginal entrant is driven in the model by the fact that the ability threshold declines as networks strengthen over time. It is standard practice to proxy for ability with education. In a developing economy, the level of education will vary across birth cohorts and in the cross-section (across birth counties) for the same level of ability, depending on the supply of schooling. Entering firms in each birth county-sector in a given time period are thus ranked on the basis of

[23] The initial capital for a firm is determined by its initial registered capital, which can be recovered from the SAIC registration database.

[24] Appendix Table A.1 reports parametric estimates corresponding to Figure 7, separately by time period. Estimates with average initial capital, measured in logs, as the dependent variable are also reported in the table. Appendix Table A.2 repeats this exercise with the estimating equation including, in addition, the county's population and average education in 1982. The population density coefficient is negative and significant (with a couple of exceptions) across all specifications and all time periods.

Figure 7. Marginal Initial Capital



Source: SAIC registration database and 1982 population census.
Initial capital in million Yuan is measured in logs.
Marginal initial capital is defined by the bottom one percentile of the initial capital distribution at the birth county-sector-time period level.

the entrepreneurs' education, which is available in the SAIC registration database, adjusted for the level of education in his birth county-birth cohort.[25] The marginal entrant is the entrepreneur who is placed at the bottom one percentile of this ranking. We see in Table 9, Column 1, which includes birth county-sector fixed effects, that the marginal entrant is drawn from lower down in his birth county-cohort education distribution over time and that this decline is significantly steeper for entrants from higher population density counties, as predicted by the model.

Table 9, Columns 2-3 use the distribution of initial capital (in logs) in each entering cohort of firms, in five-year windows over the 1990-2009 period, to identify the marginal entrant (the bottom one percentile) and the average entrant by birth county-sector. Including birth county-sector fixed effects in the estimating equation, we see that both the marginal entrant's initial capital and the average entrant's initial capital are decreasing significantly over time. While the coefficient on the time period-birth county population density interaction is also negative and significant with the marginal entrant's initial capital as the dependent variable, the interaction coefficient is positive (albeit small in magnitude and statistically insignificant) with average initial capital as the dependent variable. The analysis of firm size thus far has not accounted for destination choices, and the possibility that variation in these choices across birth counties could be driving the results. Table 9, Columns 4-5 thus includes destination fixed effects, in addition to birth county-sector fixed effects in the estimating equation. The marginal entrant's initial capital and the average entrant's initial capital are now based on the distribution of capital in each birth county-sector-destination-time period.[26] Both marginal initial capital and average initial capital are declining significantly over time, as in Columns 2-3. Moreover, the coefficient on the time period-birth county population density interaction is now negative and significant with both dependent variables, as predicted by the model. As with the analysis of firm entry, accounting for

[25]The education distribution is constructed in each county for birth cohorts from 1920 to 1989 in five-year intervals, based on data from the 2000 population census. Each entrepreneur is assigned to a birth cohort interval based on his birth year, which is available from the registration database, and his position in the relevant education distribution is determined on the basis of his education, which is also obtained from the registration database. The coverage for the education variable is not complete in the SAIC registration database, with a significant minority of entrepreneurs not reporting this information. This is not a concern with the complementary analysis of firm size, which includes all registered firms.

[26]The sample in Columns 5-6 is restricted to birth county-sector-destinations with entrants in the initial period. Similarly, the sample in Columns 3-4 is restricted to birth county-sectors with entrants in the initial period.

Table 9. Negative Selection

Dependent variable:	marginal adjusted education	marginal initial capital	average initial capital	marginal initial capital	average initial capital
	(1)	(2)	(3)	(4)	(5)
Time period	-18.532***	-0.882***	-0.115***	-0.655***	-0.109***
	(0.409)	(0.012)	(0.008)	(0.009)	(0.007)
Birth county population density × time period	-1.040***	-0.028**	0.002	-0.069***	-0.022***
	(0.394)	(0.012)	(0.008)	(0.010)	(0.007)
Mean of dependent variable	49.36	-1.744	-1.744	-1.223	-1.223
Observations	21,028	43,579	43,579	46,417	46,417
Origin-sector fixed effects	Yes	Yes	Yes	Yes	Yes
Destination fixed effects	No	No	No	Yes	Yes

Note: adjusted education is the entrepreneur's rank in his birth county-birth cohort education distribution (obtained from the 2000 population census).
Marginal adjusted education is defined by the bottom one percentile of the adjusted education distribution of entering entrepreneurs at the birth county-sector-time period level.
Initial capital in million Yuan is measured in logs.
Marginal initial capital defined by the bottom one percentile of the initial capital distribution at the birth county-sector-time period level.
Average initial capital is the mean of the distribution.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Time period is an ordinal variable taking value from 0 to 3 corresponding to successive five-year time windows over the 1990-2009 period.
Education and initial capital are obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birth county-sector level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

destination effects only strengthens our results.

Conditional on having entered, the model predicts that firms from higher population density counties will grow faster. As noted, the registration data are not suitable for computing growth rates because the firm's registered capital does not track perfectly with changes in its assets. Annual information on firm assets is available, however, in the industrial census and the SAIC inspection database. To maintain consistency in the growth analysis across the two data sources, we compute growth rates over the 2004-2008 period.[27] Table 10, Columns 1-3 report growth estimates with industrial census data, while Columns 4-6 report the corresponding estimates with SAIC inspection data. Firm-specific average annual growth rates, computed between 2004 and 2008, are averaged up to the birth county-sector level in the benchmark specification, reported in Column 1 and Column 4, which includes birth county population density and sector fixed effects in the estimating equation. As predicted by the model, the coefficient on birth county population density is positive and significant. Our next specification, reported in Column 2 and Column 5 computes growth rates at the birth county-sector-destination level and adds destination fixed effects to the estimating equation. The coefficient on birth county population density continues to be positive and significant. One alternative explanation for these findings is mechanical convergence; firms from counties with higher population densities have lower initial capital and thus grow faster. The specification in Column 3 and Column 6 accounts for this by adding the firm's initial capital to the estimating equation.[28] The coefficient on initial capital is positive and significant with both industrial census data and SAIC inspection data. The coefficient on birth county population density, nevertheless, remains positive and significant.

[27] The industrial census reports firm assets in the 1995, 2004, and 2008 rounds. The SAIC inspection database has reasonable coverage from 2004 onwards. The average annual growth is computed as the difference in log assets in 2008 and 2004 divided by four. The growth analysis with SAIC inspection data is restricted to manufacturing firms to be consistent with the industrial census data. Data coverage for seven provinces is poor with the inspection data and these provinces are thus dropped from the analysis.

[28] Initial capital is obtained from the SAIC registration database. This database can be merged with both the industrial census and the SAIC inspection database at the firm level.

Table 10. Growth of Assets

Dependent variable:	average annual growth of assets 2004-2008					
Data source:	manufacturing census			annual inspection		
	(1)	(2)	(3)	(4)	(5)	(6)
Birth county population density	0.004**	0.004**	0.004**	0.004***	0.002**	0.002**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Average initial capital	–	–	0.001***	–	–	0.001***
			(0.000)			(0.000)
Mean of dependent variable	0.133	0.136	0.136	0.106	0.110	0.110
Observations	31,234	64,258	64,258	18,708	43,494	43,494
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Destination fixed effect	No	Yes	Yes	No	Yes	Yes

Note: average annual growth of assets is computed as the difference in log assets in 2008 and 2004 divided by 4.
Firm-level average annual growth is averaged up to the birth county-sector level in specifications with sector fixed effects and to the birth county-sector-destination level in specifications with sector fixed effects and destination fixed effects.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Initial capital (in million Yuan) obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

5 Alternative Explanations

Could there be alternative explanations for the facts established above that do not require networks to be active? To answer this question, consider a simple model that retains the essential features of our model, but without the network component; i.e. with community TFP replaced by an exogenously increasing productivity term at the destination. For simplicity, assume that there is a single business sector and a single destination. To systematically examine alternative explanations, we work with this model, retaining the notation of the parent model and adding different sources of heterogeneity, which are correlated with population density, in turn.

1. Heterogeneity at the Origin (affecting entry at the intensive margin)
The model assumes that the number of potential entrepreneurs, k, and the ability distribution, $log\omega \sim U[0,1]$, in each entering cohort, from all origin counties, and constant over time. However, higher population density counties have larger populations and higher levels of education, and data from the population census indicates that this variation has widened over time. We thus allow the number of potential entrepreneurs, $k(p,t)$, in the alternative model without networks to be a function of p and t; i.e. $k_p > 0$, $k_t > 0$, $k_{pt} > 0$. Productivity at the destination, $exp(A(t))$, is assumed to be increasing over time; $A_t > 0$, but is independent of p. This retains the key assumption of the model, which is that the only source of exogenous heterogeneity is at the origin.

From equation (5) in the model, the ability of the marginal entrant from any given origin in cohort t is

$$log\underline{\omega}(t) = \frac{1}{1-\sigma}\left[log\frac{1}{\psi} - \frac{1}{1-\alpha}A(t) + \frac{\alpha}{1-\alpha}logr\right],$$

which is independent of p. However, our assumption that $k(p,t)$ is a function of p and t implies that entry, nevertheless, satisfies the predictions of the model.

$$e(p,t) = k(p,t)\left[1 - log\underline{\omega}_t\right],$$

which implies that $e_p > 0$, $e_t > 0$, $e_{pt} > 0$. Notice that this result would be obtained even if k was a function of p alone; i.e. $k(p)$, with $k_p > 0$, because $A(t)$ is increasing in t.

While heterogeneity in k can explain all the patterns for entry without requiring networks to be active, it cannot explain our model's predictions for initial capital and firm growth because $\underline{\omega}_t$ and $A(t)$, respectively, are independent of p. Once we allow firms from a single origin to select into multiple sectors and destinations, it also cannot explain why sectoral and spatial concentration is increasing in birth county population density. We would expect a larger population and higher education to be associated with greater occupational diversity in the origin county, due to greater diversity in skills or learning opportunities. If the (traditional) occupational pattern in the birth county determines, in part, the sectors and locations that its entrepreneurs choose, then this would result in *lower* concentration for firms from higher population density birth counties.

2. Heterogeneity at the Origin (affecting entry at the extensive margin)
The model assumes that productivity in the traditional sector is the same in all origin counties and constant over time. It also ignores moving costs, which could vary across origins. We relax these assumptions of the model by specifying the production function in the origin as $exp(v(p,t))\omega^{\sigma}$. The $exp(v(p,t))$ term can be interpreted as productivity in the traditional sector (this was normalized to one in our model). Once we take logs, and examine occupational choice, $v(p,t)$ can also be interpreted as a moving cost. Continuing to assume that productivity at the destination, $A(t)$, is increasing over time but independent of p, the marginal entrant's ability in cohort t can be expressed as

$$log\underline{\omega}(p,t) = \frac{1}{1-\sigma}\left[log\frac{1}{\psi} - \frac{1}{1-\alpha}A(t) + \frac{\alpha}{1-\alpha}logr + v(p,t)\right].$$

If $v_p < 0$, $v_t < 0$, $v_{pt} < 0$, then it is straightforward to verify that all the parent model's predictions for entry and initial capital go through. However, with the productivity interpretation for $v(p,t)$, the required assumption that productivity in the traditional sector is *lower* in higher p locations is quite strong; if anything, we would expect the opposite in origin counties with larger and more educated populations. There is also no reason why productivity should be declining over time, more steeply in higher-p counties. When $v(p,t)$ is interpreted as a moving cost, $v_p < 0$ may be more reasonable because higher p origins are likely to be more spatially connected (and this advantage could potentially have strengthened over time with economic development). Suppose that we allow firms from a single origin to select into multiple sectors and destinations. If moving costs for higher-p origins are decreasing especially rapidly in particular destinations (and their associated sectors), then this would also explain our finding that concentration is increasing in p, more steeply over time. What the alternative model still cannot explain, however, is why firm growth at the destination (which depends on $A(t)$) is increasing in p.

3. Heterogeneity at the Destination
Suppose we relax the assumption that there is a single destination and allow firms from different origins to end up at different destinations, each with its own exogenously determined productivity growth. The marginal entrant's ability in cohort t can then be expressed as

$$log\underline{\omega}(p,t) = \frac{1}{1-\sigma}\left[log\frac{1}{\psi} - \frac{1}{1-\alpha}A(p,t) + \frac{\alpha}{1-\alpha}logr\right].$$

If $A_p > 0$, $A_t > 0$, $A_{pt} > 0$, then there is little to distinguish the alternative model from the parent model with networks in which $A(p,t)$ is endogenously determined. Our model's predictions for entry, initial capital, and now even firm growth can be generated by the alternative model. Once we allow firms from a single origin to select into multiple sectors and destinations, the alternative model will also predict higher concentration for higher p origins.

Based on the results we have presented, however, there are a number of reasons why the assumption that firms from higher-p counties end up at more favorable destinations may not be valid. First, there is no evidence that these firms ended up at destinations (or in sectors) that received a larger number of entrants overall. Entry analysis in Tables 3 and 4, with and without sector and destination fixed effects, indicates that firms from higher p origins ended up in sectors/destinations that received fewer entrants in each time period. The counter-factual entry analysis in Table 8, based on initial destination choices within sectors, also finds that firms from higher p origins did not select into destinations that received more entrants overall. Second, the initial capital and firm growth results in Tables 9 and 10 are robust to the inclusion of destination fixed effects. These results are generated by variation in origin county population density *within* destinations, and would not be obtained if firms from higher p origins simply had access to destinations with superior infrastructure, cheaper capital, or larger agglomeration externalities. Complementing this result, and providing direct evidence that community networks are active, recall that entry into particular sector-destinations in Table 7 is driven entirely by initial entry into those sector-destinations by firms from the same birth county. If all firms in a given sector-destination were benefiting from the same, exogenous, economic environment, then this result would not be obtained.

6 Structural Estimation and Quantification

The model generates predictions for two independent outcomes: entry at the sector level and firm size. Sector-specific entry is aggregated up to compute total entry and is also used to construct concentration statistics. Firm-specific capital investment, together with the assumed distribution of ability in the population and the endogenously generated pattern of entry, is used to derive marginal initial capital, average initial capital, and capital growth in each birth county-sector. The structural estimation is thus based on two fundamental equations, with sector-specific entry and average initial capital (by sector) for each birth county in each time period as the dependent variables. We choose average initial capital as the dependent variable in the capital equation because it is conveniently aggregated up in the quantification exercise to derive a measure of total production.

Based on the corresponding equations in the model, (7) and (17), and retaining its notation, we estimate the following structural equations:

$$e_{ci,t} = G(\alpha, \sigma, r, A_0)k_c S_{ci,t-1} + \frac{\theta}{(1-\sigma)(1-\alpha)} k_c S_{ci,t-1} \cdot pn_{ci,t-1} \tag{20}$$

$$logK^a_{ci,t} = H_t(\alpha, \sigma, r, A_0, f_t) + \frac{\theta(1-2\sigma)}{2(1-\sigma)(1-\alpha)} pn_{ci,t-1} \tag{21}$$

$e_{ci,t}$ measures the number of entrants and $logK^a_{ci,t}$ measures average initial capital (in logs) for birth county c and sector i in time period t. $n_{ci,t-1}$ is the stock of firms from that birth county that are already established in that sector at the beginning of the time period. k_c measures the number of potential entrepreneurs from the birth county. k_c was set to be equal across birth counties and time periods in the model. For the structural estimation, the number of potential entrepreneurs in each birth county is derived from the 1990 population census, based on the characteristics of actual entrepreneurs when they established their firms (which is derived, in turn, from the SAIC registration database).[29] Capital is measured in the model in physical units, whereas in the data it is measured in monetary units. The mapping from physical units to monetary units changes over time owing to changes in the relative price of capital goods. This is especially relevant in the structural estimation because the objective is to match predicted and actual firm size in each time period. f_t thus measures the time period specific mapping from physical capital to capital in monetary

[29] We see in Appendix Figure A.1 that most entrepreneurs in the registration database have at least high school education and that most were aged 25-44 when their firm was established. k_c is thus specified to be the number of men born in county c, aged 25-44, with at least high school education, as reported in the 1990 population census.

units. We parameterize the $\theta(p)$ function to be increasing linearly in p in the equations above, with the restriction that $\theta(0) = 0$.

The structural equations are linear in variables; (i) $k_c S_{ci,t-1}$ (ii) $k_c S_{ci,t-1} pn_{ci,t-1}$ (iii) $pn_{ci,t-1}$, with four reduced-form coefficients.[30] One of these coefficients, H_t, cannot be used to identify the structural parameters because f_t is unobserved by the econometrician. This leaves three reduced-form coefficients and five structural parameters: $\alpha, \sigma, r, A_0, \theta$. The model is evidently under-identified, and our solution to this problem is to assign values to some of these parameters. Based on evidence from recent surveys of firms in China, we set the interest rate, r to 0.3.[31] The productivity multiplier is set to one in all sectors; i.e. $A_0 = 1$. The variation in productivity across sectors in the structural model is thus generated entirely by the network effect.

To accommodate differences in the capital requirement across sectors, we allow the α parameter, which measures the marginal returns to capital, to vary across four broad sector categories: high-tech services, wholesale and retail services, manufacturing and transportation, and heavy industry (mining, electricity, and construction). This increases the number of structural equations to eight, with two equations in each sector category, and the number of structural parameters to six; $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma, \theta$. The structural parameters can, nevertheless, be estimated by matching on entry and average initial capital in each birth county-sector-time period. Initial entry in each birth county-sector is based on the number of entrants in 1990-1994 and, if there is no entry in that time period, on the number of entrants in 1995-1999. Sectors are defined at the one-digit level in the structural estimation to ensure that there is positive initial entry in (almost) all birth county-sectors. Subsequent entry is derived endogenously in the model, which is estimated over the 1995-2004 period; i.e. over two time periods.

Parameter estimates, with bootstrapped standard errors in parentheses, are reported in Table 11, Column 1. The σ coefficient lies between 0.5 and 1, satisfying the condition, derived in the model, which ensures that average initial capital is decreasing in birth county population density. The adjustment from physical capital to capital in monetary units, f_t, appears additively in the H_t function and, thus, can be estimated separately in 1995-1999 (period 1) and 2000-2004 (period 2). Table 11, Column 2 examines the sensitivity of the parameter estimates to a change in the interest rate from 0.3 to 0.2. The interest rate parameter is not identified, and we verified that this change has no effect on predicted entry or initial capital. However, the estimated parameters will change in magnitude to accommodate the change in r while leaving the predicted outcomes unchanged; although the σ parameter continues to be lie between 0.5 and 1, the θ parameter declines in magnitude. The latter parameter plays a key role in generating variation across birth county-sectors and over time through the underlying community networks. When the interest rate declines, both entry and capital investment shift up and so the θ parameter must decline to retain their levels. The θ parameter declines even further in Table 11, Column 3 as a compensatory response when we allow agents to be forward looking (one period ahead).[32]

Figure 8a assesses the goodness of fit of the model by comparing actual and predicted entry across birth counties in each time period. The model that we estimate is extremely parsimonious, with just six parameters. Nevertheless, it does a good job of predicting entry across nearly 2,000 birth counties and over two time periods. Figure 8b repeats this exercise with average initial capital and, once again, we see that

[30] The functional forms for $G(\alpha, \sigma, r, A_0)$ and $H_t(\alpha, \sigma, r, A_0, f_t)$ are obtained directly from (7) and (17), with the addition of the separable f_t term in the H_t function.

[31] The interest rate, r, is the sum of the real interest rate and the depreciation rate. Hsieh and Klenow (2009) assume that the real interest rate is 0.05 in an economy, such as the U.S., with perfect financial markets and that the depreciation rate is 0.05. Using the same production function as Hsieh-Klenow and data from the industrial census, Brandt, Kambourov, and Storesletten (2016) estimate the real interest rate to be 0.15 in 1995 and 2004 and 0.18 in 2008. To be conservative, we set r to 0.3, but also report estimates with $r = 0.2$.

[32] The extension to the model that allows for forward looking behavior is derived in Appendix A. Entry must now be derived as the solution to a nonlinear equation, satisfying a fixed point condition, in each birth county-sector-time period. The discount factor per year, δ is set to 0.8 when estimating the model with foresight. Because one time period is five years, this works out to $0.8^5 = 0.33$.

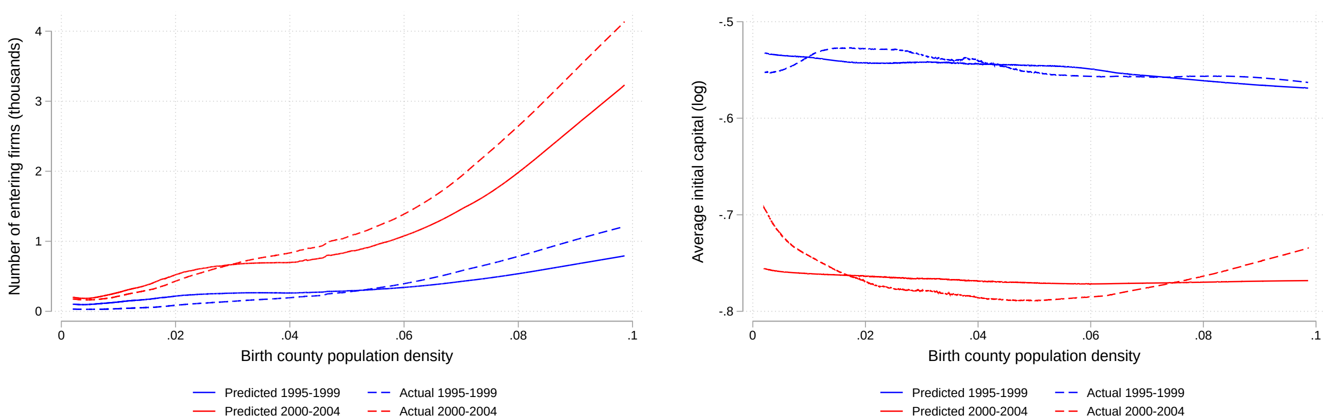
Table 11. Structural Estimates

Model:	myopic		forward looking
Interest rate:	r = 0.3	r = 0.2	r=0.3
σ	0.6418	0.7762	0.6437
	(0.0005)	(0.0003)	(0.0014)
θ	0.5511	0.3902	0.2191
	(0.1025)	(0.0726)	(0.0208)
α_1	0.1946	0.1238	0.1946
	(0.0020)	(0.0013)	(0.0023)
α_2	0.2741	0.1787	0.2777
	(0.0027)	(0.0017)	(0.0044)
α_3	0.2785	0.1822	0.2737
	(0.0026)	(0.0017)	(0.0033)
α_4	0.3437	0.2324	0.3438
	(0.0019)	(0.0013)	(0.0025)
f_1	0.2409	0.2437	0.2408
	(0.0037)	(0.0035)	(0.0041)
f_2	0.2004	0.2027	0.2001
	(0.0025)	(0.0023)	(0.0030)

Note: the parameters are estimated by matching on entry and average initial capital (log), measured at the birth county-sector-time period level.
The model is estimated over two time periods, 1995-1999 and 2000-2004, taking entry in the first period, 1990-1994, as given.
Sectors are defined at the one-digit level when measuring entry and capital, but the α parameter is estimated at the aggregate sector level: (1) new technological services; (2) wholesale, retail and business service; (3) manufacturing and transportation; (4) heavy industry (mining, electricity & construction)
The discount factor per year δ is set to 0.8 in the forward looking model.
Bootstrapped standard errors in parentheses.
Number of entrants and average initial capital are derived from the SAIC registration database and population density is derived from the 1982 population census.

the model predicts variation across birth counties fairly accurately. Once the capital adjustment factor is included in each time period, note that actual and predicted average initial capital will match on levels by construction.

Figure 8. Actual and Predicted, Entry and Initial Capital



(a) Entry **(b)** Initial Capital

Source: SAIC registration database, model generated data, and 1982 population census.

In an alternative test of the goodness of fit, we compare reduced form estimates obtained with the actual data and the data generated by the (estimated) model, based on both myopic behavior and forward looking behavior. Given that the structural estimates are based on two time periods, the estimating equation includes birth county population density and a binary time period variable, which takes the value one in 2000-2004 and zero in 1995-1999, as the only regressors. The dependent variable is either entry or average initial capital, measured at the birth county-sector-time period level. If the model generated data match the actual data across the range of birth county population densities, then the estimated coefficients will be the same, regardless of the data that are used for estimation. The estimated population density coefficients in Table 12a are statistically indistinguishable with actual data and (myopic) model generated data. While the corresponding estimates with the forward looking model are statistically indistinguishable from the estimates with the myopic model, the point estimates with the myopic model are, if anything, closer to those obtained with the actual data.

Table 12a. Estimates based on Actual and Model Generated Data (In-Sample)

Dependent variable:	number of entrants			average initial capital		
Data:	actual	model generated (myopic)	model generated (forward looking)	actual	model generated (myopic)	model generated (forward looking)
	(1)	(2)	(3)	(4)	(5)	(6)
Birth county population density	0.336***	0.226***	0.138***	-0.014***	-0.004***	-0.004***
	(0.034)	(0.030)	(0.016)	(0.005)	(0.001)	(0.001)
Time period	0.477***	0.359***	0.350***	-0.232***	-0.223***	-0.223***
	(0.020)	(0.021)	(0.017)	(0.010)	(0.003)	(0.003)
Constant	-0.327***	-0.138***	-0.206***	-0.307***	-0.319***	-0.319***
	(0.018)	(0.018)	(0.014)	(0.019)	(0.005)	(0.005)
Observations	3,032	3,032	3,032	19,486	19,486	19,486

Note: firm entry and initial capital (in million Yuan) obtained from the SAIC registration database and population density is derived from the 1982 population census.
Number of entrants and average initial capital (in logs) computed at the birth county-sector-time period level.
Time period is a binary variable taking the value one for the 2000-2004 period and zero for the 1995-1999 period.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Standard errors clustered at the birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Table 12b. Estimates based on Actual and Model Generated Data (Out of Sample)

Dependent variable:	number of entrants			adjusted HHI across sectors		
Data:	actual	model generated (myopic)	model generated (forward looking)	actual	model generated (myopic)	model generated (forward looking)
	(1)	(2)	(3)	(4)	(5)	(6)
Birth county population density	0.919***	0.821***	0.707***	0.217***	0.851***	0.606***
	(0.083)	(0.127)	(0.179)	(0.018)	(0.039)	(0.080)
Constant	1.331***	0.704***	0.992***	2.708***	2.670***	5.102***
	(0.041)	(0.049)	(0.081)	(0.016)	(0.030)	(0.079)
Observations	1,516	1,516	1,516	1,516	1,516	1,516

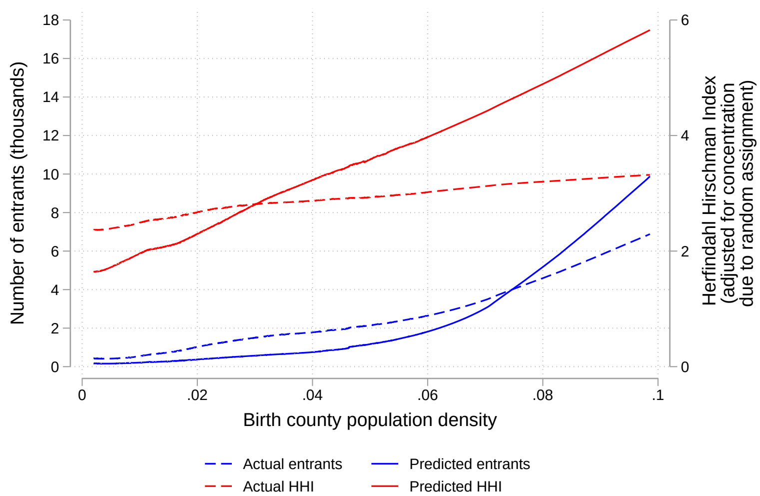
Note: number of entrants (in thousands) is measured at the birth county level over the 2005-2009 period.
Sectoral concentration measured by the Herfindahl Hirschman Index(HHI) across one-digit sectors at the birth county level over the 2005-2009 period is divided by the expected HHI that would be obtained by random assignment, given the number of entrants and the number of sectors.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Firm entry and HHI are obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at the birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

A more stringent test of the model is to assess its out-of-sample predictions. The structural model is estimated over the 1995-1999 and 2000-2004 periods. Figure 9 compares actual and predicted entry and sectoral concentration in the subsequent 2005-2009 period.[33] The model does a good job of predicting firm entry and sectoral concentration, although it does over-shoot at higher birth county population density

[33]We cannot test the model's ability to predict capital investments beyond the sample period because the mapping from physical capital to capital in monetary units is unavailable.

levels. As with the in-sample predictions, we independently assess the model's ability to match the data by comparing reduced form estimates obtained with actual data and model generated data, based on both myopic and forward looking behavior. Because there is now a single period, 20005-2009, the estimating equation includes a single regressor; birth county population density. The estimated coefficient on birth place population density is statistically indistinguishable with actual and (myopic) model generated data when firm entry is the dependent variable in Table 12b, Columns 1-2. In contrast, it is the level; i.e. the intercept that is statistically indistinguishable when sectoral HHI is the dependent variable in Table 12b, Columns 4-5. Once again, the myopic model does at least as good a job of matching the data as the forward looking model.

Figure 9. Out of Sample Tests C Entry and Sectoral Concentration, 2005-2009

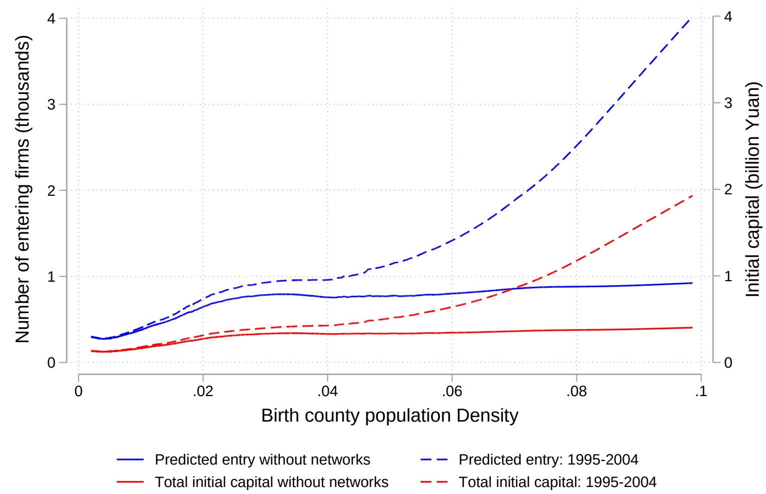


Source: SAIC registration database, model generated data, and 1982 population census.
Number of entrants and sectoral concentration computed at the birth county level in 2005-2009 with actual and model generated data.

With a single θ parameter common to all birth counties and sectors in the model, cross-sectional variation is generated by differences in population density and initial entry alone. Nevertheless, the model is able to match the data quite well, across counties and sectors, even out of sample. The estimated parameters can thus be used for counter-factual simulations. A major objective of our research is to quantify the role played by community networks in the growth of private enterprise in China. This is accomplished by setting the θ parameter to zero and then generating counter-factual entry and capital investment over the sample period. The results of this exercise are reported in Figure 10. It is evident from the figure that the number of entrants and the accompanying initial capital investment would have been substantially reduced in the absence of community networks, particularly in higher population density birth counties. Based on our estimates, the total number of entrants would have declined by 26.5% and the total initial capital investment would have declined by 24.6% over the 1995-2004 period if the networks had not been active.

An important objective of industrial policy in any developing economy is to stimulate entrepreneurship. It has been claimed that the government played a critical role in accelerating China's growth by providing firms with subsidized credit; e.g. Wu (2016). In the absence of a market-based allocation mechanism, a natural question to ask is which firms should have been targeted for the subsidy. To answer this question, we examine a counter-factual policy experiment in which all entering firms in the 1995-1999 period received credit at an interest rate of 0.25; i.e. with a subsidy of 0.05. This subsidy would have had two effects; it would have induced additional firms to enter at the margin and it would have increased the profit of all (marginal and infra-marginal) entrants. As observed in Figure 11, the total profit increase generated by the subsidy far exceeds its cost to the government, across all birth counties. More interestingly, the spillover effect of the

Figure 10. Counter-Factual Simulation: Effect of Community Networks on Entry and Total Initial Capital



Source: Model generated data, and 1982 population census.

one-time subsidy in future time periods, 2000-2004 and 2005-2009, dwarfs the direct effect in the 1995-1999 period in higher population density birth counties. This is because the credit subsidy induces additional entry, which through the compounding network effect generates large profit increases in the more socially connected counties. If the government took account of these spillovers, which are not currently internalized by individual entrants, and its objective was to maximize total profit, then the optimal policy would be to provide subsidized credit to marginal entrepreneurs (who would not enter otherwise) from birth counties with higher population densities. Given the negative selection that we have documented, these entrepreneurs would have relatively low ability and relatively small firms, and existing variation in firm size and productivity within sectors (across networks) would widen even further.

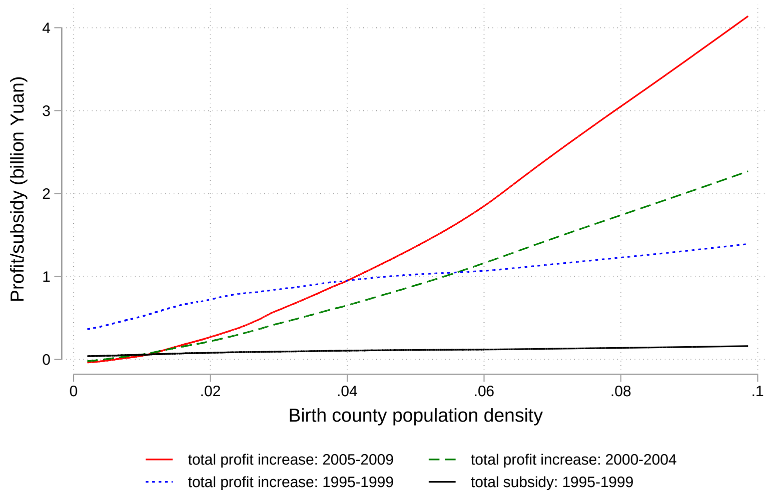
The preceding result highlights a central message of our paper, which is that while small firms and wide dispersion in firm size and productivity may be symptomatic of market inefficiencies, in a world with such inefficiencies (and the networks that emerge in response to these inefficiencies) a more successful second-best response may, paradoxically, be accompanied by smaller firms and greater dispersion.[34] The additional message is that attempts to stimulate growth by offering subsidies to potential entrepreneurs should take account of both individual ability, which determines independent business success, and social affiliation, which determines the spillovers that lead to further entry, in economies where networks are active.

7 Conclusion

In this paper, we identify and quantify the role played by community networks, organized around the birth county, in the growth of private enterprise in China. The model that we develop generates dynamic predictions for firm entry, sectoral and spatial concentration, and firm size across birth counties with different levels of social connectedness (measured by population density) when networks are active. We validate each of these predictions with unique administrative data that covers the universe of registered firms and provides information on entrepreneurs' birth counties. The rich set of results that we obtain, taken together, allow

[34]Two stylized facts motivate a large and growing macro-development literature on misallocation: (i) that the variation in marginal productivity within narrow sectors is too wide in developing economies, and (ii) that firms in those economies are too small and grow too slowly (Peters, 2016). Although a number of mechanisms have been proposed to explain these facts; e.g. Caunedo (2016), Asker, Collard-Wexler, and De Loecker (2014), Peters (2016), and Akcigit et al. (2016), perhaps the simplest is based on a neoclassical model with wedges in factor prices (Restuccia and Rogerson (2008), Hsieh and Klenow (2009, 2014)).

Figure 11. Counter-Factual Simulation: Effect of Interest Rate Subsidy on Profits



Source: Model generated data, and 1982 population census.

us to rule out alternative non-network explanations. Having validated the model, we estimate its structural parameters and conduct counter-factual simulations.

The first simulation shuts down the community networks. Our estimates indicate that this would have reduced the total number of entering firms and their (initial) capital stock by around 25% over the 1995-2004 period. This was a critical period in China's economic development, and given the network multiplier effect, the long-term consequences of shutting down the networks at this time would have been even more significant. Individual firms do not internalize the positive externality they generate for other members of the network, which implies that a credit subsidy to stimulate entry could be efficiency enhancing. Our second counter-factual experiment considers the effect of a one-time credit subsidy for entrants in the 1995-1999 period. This subsidy would have increased profits for infra-marginal firms (who would have entered regardless) but, more importantly, would have induced additional firms to enter at the margin. The effect of this additional entry would have compounded over time due to the networks, and our estimates indicate that the direct effect of the subsidy on profits in the 1995-1999 period would have been dwarfed by the indirect spillover effects (in later periods) particularly in higher population density birth counties.

If the government's objective was to maximize total profits, then the optimal policy would be to provide subsidized credit to marginal entrepreneurs (who would not otherwise enter) from higher population density birth counties so as to take advantage of the network externalities. Given the negative selection that we have documented, however, these entrepreneurs would have relatively low ability among all marginal entrepreneurs. If individual merit is the only consideration for selection into entrepreneurship training programs and business plan competitions, which have grown increasingly popular in recent years, then the individuals with the greatest potential to stimulate further entry would not be selected.

There are, however, two caveats to the policy prescription we have recommended. First, a policy that places weight on both social affiliation and individual merit will only be effective in a population where community networks are already active or have the potential to be activated. More generally, the Chinese development experience will not be replicated in other countries by simply providing infrastructure and credit. This is relevant for Chinese overseas development assistance policy, which has largely focussed on infrastructure construction and industrial development, with a special emphasis on Africa (Zhang, 2016). This policy is explicitly motivated by the Chinese domestic experience, and the belief that infrastructure construction is the key to development (see, for example, China's second Africa policy paper; Xinhua, December 4, 2015) . Chinese development assistance has grown exponentially in recent years (Lin and Wang (2017)), but our analysis indicates that the expected returns will only be realized if community networks in the recipient countries evolve in parallel with the infrastructure construction, just as they did in China. The second caveat is that even if the social structure is amenable to network-based growth, there are important consequences for inequality that must be considered. By bringing in less able entrepreneurs at the margin, community networks are redistributive within their populations. However, a policy that targets individuals from more socially connected populations to take advantage of the positive externalities that their stronger networks provide will only exacerbate existing inequalities across communities. Given the dynamic increasing returns that are generated by the networks, these inequalities will persist and, if anything, worsen over time. Absent other redistributive mechanisms, any policy that attempts to exploit network externalities must pay attention to the potentially enduring consequences for inequality.

References

Aghion, P., J. Cai, M. Dewatripont, L. Du, A. Harrison, and P. Legros (2015): "Industrial policy and competition," *American Economic Journal: Macroeconomics*, 7(4), 1–32.

Allen, F., J. Qian, and M. Qian (2005): "Law, finance, and economic growth in China," *Journal of Financial Economics*, 77(1), 57–116.

Au, C.-C., and J. V. Henderson (2006a): "Are Chinese Cities Too Small?," *The Review of Economic Studies*, 73(3), 549–576.

——— (2006b): "How Migration Restrictions Limit Agglomeration and Productivity in China," *Journal of Development Economics*, 80(2), 350–388.

Banerjee, A., and K. Munshi (2004): "How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur," *The Review of Economic Studies*, 71(1), 19–42.

Brandt, L., G. Kambourov, and K. Storesletten (2016): "Firm Entry and Regional Growth Disparities: the Effect of SOEs in China," *University of Toronto mimeo.*

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012): "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing," *Journal of Development Economics*, 97(2), 339–351.

Cai, J., and A. Szeidl (2016): "Interfirm Relationships and Business Performance," Discussion paper, National Bureau of Economic Research.

Cai Fang (1997): "Recent Trends of Migration and Population in Shandong," *Working Papers on Chinese Politics, Economy and Society*, (10).

Coleman, J. S. (1988): "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94, S95–S120.

Desmet, K., and E. Rossi-Hansberg (2013): "Urban Accounting and Welfare," *American Economic Review*, 103(6), 2296–2327.

Duranton, G., and H. G. Overman (2005): "Testing for Localization Using Micro-Geographic Data," *The Review of Economic Studies*, 72(4), 1077–1106.

Ellison, G. (1994): "Cooperation in the prisoner's dilemma with anonymous random matching," *The Review of Economic Studies*, 61(3), 567–588.

Ellison, G., and E. L. Glaeser (1997): "Geographic Concentration in US Manufacturing Industries: a Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

Fleisher, B., D. Hu, W. McGuire, and X. Zhang (2010): "The evolution of an industrial cluster in China," *China Economic Review*, 21(3), 456–469.

Glaeser, E. L., D. Laibson, and B. Sacerdote (2002): "An Economic Approach to Social Capital," *The Economic Journal*, 112(483), F437–F458.

Goodman, B. (1995): *Native place, city, and nation: Regional networks and identities in Shanghai, 18531937.* University of California Press.

Granovetter, M. (1985): "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology*, 91(3), 481–510.

Honig, E. (1992): *Creating Chinese Ethnicity: Subei People in Shanghai, 1850-1980*. Yale University Press, New Haven.

——— (1996): "Regional identity, labor, and ethnicity in contemporary China," in *Putting class in its place : worker identities in East Asia*, ed. by E. J. Perry, no. 48 in China research monographs. Institute of East Asian Studies, University of California, Berkeley, California.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

——— (2014): "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 129(3), 1035–1084.

Hu, B. (2008): "People's Mobility and Guanxi Networks: A case study," *China & World Economy*, 16(5), 103–117.

Huang, Z., X. Zhang, and Y. Zhu (2008): "The role of clustering in rural industrialization: A case study of the footwear industry in Wenzhou," *China Economic Review*, 19(3), 409–420.

Kandori, M. (1992): "Social norms and community enforcement," *The Review of Economic Studies*, 59(1), 63–80.

Long, C., and X. Zhang (2011): "Cluster-based industrialization in China: Financing and performance," *Journal of International Economics*, 84(1), 112–123.

——— (2012): "Patterns of China's industrialization: Concentration, specialization, and clustering," *China Economic Review*, 23(3), 593–612.

Ma, L. J. C., and B. Xiang (1998): "Native Place, Migration and the Emergence of Peasant Enclaves in Beijing," *The China Quarterly*, 155, 546.

McKenzie, D. (2017): "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition," *American Economic Review*, 107(8), 2278–2307.

McKenzie, D., and C. Woodruff (2014): "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?," *The World Bank Research Observer*, 29(1), 48–82.

Munshi, K. (2011): "Strength in Numbers: Networks as a Solution to Occupational Traps," *The Review of Economic Studies*, 78(3), 1069–1101.

Munshi, K., and M. Rosenzweig (2006): "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 96(4), 1225–1252.

——— (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 106(1), 46–98.

Restuccia, D., and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11(4), 707–720.

Ruan, J., and X. Zhang (2009): “Finance and cluster-based industrial development in China,” *Economic Development and Cultural Change*, 58(1), 143–164.

Song, Z., K. Storesletten, and F. Zilibotti (2011): “Growing Like China,” *American Economic Review*, 101(1), 196–233.

Summers, L. (2007): “The Rise of Asia and the Global Economy,” *Research Monitor (the bi-annual newsletter of the Global Development Network)*, (2).

Wu, M. (2016): “The China, Inc. Challenge to Global Trade Governance,” *Harv. Int’l LJ*, 57, 261.

Zhang, X., and G. Li (2003): “Does guanxi matter to nonfarm employment?,” *Journal of Comparative Economics*, 31(2), 315–331.

Zhao, Y. (2003): “The Role of Migrant Networks in Labor Migration: the Case of China,” *Contemporary Economic Policy*, 21(4), 500–511.

Appendix A. Social Connectedness and Population Density

Here we provide a microfoundation for the assumption that social connectedness is increasing in population density at the origin.

1. Cooperation at the Origin

The neighborhood at the origin has a given geographic area normalized to unity, and has p residents. If the origin is a county, the residents constitute a homogenous social group of size $n = p$. If it is a city instead, it is comprised of s different social groups of size $n = \frac{p}{s}$ each. Information spreads perfectly and instantly within social groups, but does not spread at all between groups. Cities are typically characterized by higher values of p, but this is mainly on account of coexistence of a greater number of social groups rather than higher size per group. Indeed, p and n are likely to be negatively correlated in cities: i.e., social groups are smaller on average compared to counties.

At every date residents are subject to an i.i.d. shock wherein with a given probability a, each resident needs help from others in the community with respect to some economic or social enterprise. When this event occurs, every other local resident is chosen with probability $\frac{k}{p}$ as a potential help-giver. Hence on average there are k potential help-givers. The level of help provided by any neighbor is a real number $e \geq 0$: this generates benefit be to the recipient, while the giver incurs a personal cost of $\frac{e^2}{2}$. A norm of effort e provided by any help-giver, if it could be sustained by some means, would generate an expected per capita payoff of $\Pi(e) \equiv ak[be - \frac{e^2}{2}]$.[35] The (utilitarian) optimal level of help is $e^* = b$, while myopic Nash equilibrium entails zero help.

We now examine what help norms can be sustained in an incentive compatible manner. At the end of each date, a resident that experienced the shock observes (costlessly and without error) the aggregate help received from others, but not necessarily the level of help provided by any specific help-giver. The latter could be monitored with some probability by other residents in the neighborhood: whatever information is available to any other resident spreads instantaneously within the social group of that resident.

Consider now the level of effort that can be sustained in a county, which has a single social group. The probability that the action of any specific help-giver (at any given date) is directly observed by *someone* within the county is $q(n)$ which is rising in the size $n = p$ of the county.[36] This is the probability that this action becomes known to all others in the community; with probability $1 - q(n)$ it remains hidden for ever.

Any downward deviation from a group norm of effort e that becomes known to the community, is punished by grim trigger strategies: the miscreant is denied help forever thereafter. With a common discount factor $\delta \in (0, 1)$, the maximal supportable level of help in the range $(0, b]$ is $e(n) = b \min\{1, \frac{2q(n)}{1-\delta+q(n)}\}$. The associated per period expected payoff $P(n) \equiv \Pi(e(n))$ of any given resident is strictly increasing in n over the range $q(n) < 1 - \delta$. Assuming that $q(\bar{p}) < 1 - \delta$ where p is an upper bound for the size of a county, it follows that counties with higher p will achieve greater cooperation.

In a city with multiple social groups, potential help-givers are chosen randomly and so will frequently contain people from other social groups. Information about actual help provided by specific help-givers is available (if at all) only to those in the same social group, and spreads within that group only. The strongest possible punishments of deviations by help-givers involve those from the same social group. Hence the same expression $e(n)$ represents the maximal help that can be sustained in a city. The key point to note is that —

[35] With probability a each resident needs and receives help of ke. Hence the expected benefit of help received is $akbe$. Each resident also provides help at a cost of $\frac{e^2}{2}$ for an expected number ak other residents, since the likelihood of having to provide help any other specific resident is $\frac{ak}{p}$ and there are p residents in all.

[36] For example, suppose each resident observes help provided by i at t with respect to any other resident with probability z (independently of any other member of the group). Then a deviation remains hidden with probability $(1 - z)^{n-1}$, and $q(n) = 1 - (1 - z)^{n-1}$.

unlike a county — n and p are not the same in a city, and could be negatively correlated. Hence cooperation levels in the city may be independent or declining in p.

2. Cooperation at the Destination

Now we explain how a set of entrepreneurs in a given sector from a common home-county can achieve higher (revenue) productivity. The basic idea is simple: productivity improvements frequently require the help of other entrepreneurs in the same sector in improving product quality or securing new marketing orders. These requirements often arise randomly, as well as the identity of other entrepreneurs in the sector who have the required qualification or connections who can potentially provide the specific help needed. For simplicity consider the case where whenever help is needed by some entrepreneur, it can be provided by only one other entrepreneur in the sector, and the level of help actually provided can be observed perfectly by the entrepreneur needing it. Markets for 'help' are missing, owing to standard problems of verifiability and enforcement: hence it is not possible to obtain help by paying for it. This generates a potential free-riding problem in the provision of such help: in the absence of any social connectedness, selfish behavior will not be consistent with any provision of help. However, if the entrepreneurs hail from the same county with a high level of connectedness, failure to provide requisite help in the entrepreneurial arena can be punished by 'ostracism' back in the home county (i.e., in social and economic interactions of the entrepreneurs or their families 'back home'). So entrepreneur networks from home counties with high p can sustain greater help among one another.

Here is a formal elaboration. Let the size of the network (members from the same origin who have already entered) at a given destination be denoted N_t, and let the TFP of a representative member at date t be denoted A_t. Every entrepreneur receives an opportunity to improve his own productivity, which requires the help of another entrepreneur operating in the same sector with a suitable skill or contact. As explained above, non-negligible help will be provided only if the latter hails from the same origin county.[37]

Consider the productivity of any given entrepreneur, denoted by E_1. Let N_t denote the number of entrepreneurs in the same sector hailing from the same home county as E_1. The probability that someone within that set can be found who has the capacity to provide the help needed is γN_t (where γ is some positive number small enough to ensure that γN is smaller than one for all relevant values of N). Once such an entrepreneur (E_2, say) is found, a help request is sent by E_1. If $h \geq 0$ denote the help subsequently provided by E_2, the resulting TFP growth rate from t to $t+1$ achieved by E_1 equals h, while generating a cost of $c.h$ to E_2. The expected TFP of E_1 at the next date equals

$$A_{t+1} = A_t \exp(\gamma N_t h) \tag{22}$$

Failure by E_2 to provide a stipulated norm for the level of help is punished via sanctions at the origin (in the game they or their respective families play there). Then the maximal per member effort that can be supported at the destination is

$$h(p) = \frac{\delta}{c(1-\delta)} P(p) \tag{23}$$

for a network with a county origin of population density p. Combining with (22), we obtain the expression for productivity growth used in Section xx:

$$A_{t+1} = A_t \exp(\theta(p) N_t) \tag{24}$$

where $\theta(p) \equiv \gamma \frac{\delta}{c(1-\delta)} P(p)$ is increasing in p.

[37] With a large number of entrepreneurs in each sector, combined with the randomness of who can provide it, strategies of bilateral reciprocity can sustain only negligible levels of cooperation.

3. Extension: Entry with Foresight

Consider the consequences of allowing entrepreneurs to look ahead and incorporate profits they would expect to make after the first period they enter. We suppose cohort t agents look ahead one additional period, i.e., make their entry decision based on anticipated present value profits in periods t and $t+1$. The equilibrium can no longer be computed recursively, owing to the need for entrants to coordinate their expectations of entry decisions of one another. We shall consider equilibria where these expectations are fulfilled. We continue to assume that incumbents are committed to their previous entry decisions.

Let ζ denote $\psi r^{-\frac{\alpha}{1-\alpha}}$, and $\delta \in (0,1)$ denote the common discount factor of agents. Then expected present value of entering B_i at t for a cohort t agent of talent ω is

$$P_{it}(\omega) = \omega\zeta A_0^{\frac{1}{1-\alpha}} \exp(\theta p n_{i,t-1}\frac{1}{1-\alpha})[1 + \delta \exp(\theta(p)e_{it}\frac{1}{1-\alpha})] \tag{25}$$

while of staying in T is

$$N_{it}(\omega) = \omega^{\sigma}[1+\delta] \tag{26}$$

The agent will enter if

$$\log\omega > \frac{1}{1-\sigma}[-\log\zeta - \frac{1}{1-\alpha}\log A_0 + \log(1+\delta) - \frac{1}{1-\alpha}\theta(p)n_{i,t-1} - \log\{1+\delta\exp(\theta(p)e_{it}\frac{1}{1-\alpha})\}] \tag{27}$$

Define the function

$$\begin{aligned} g(e|s_{i,t-1}, n_{i,t-1}, A_{i0}) &= ks_{i,t-1}\{1 + \tfrac{1}{1-\sigma}[\log\zeta + \tfrac{1}{1-\alpha}\log A_0 - \log(1+\delta) \\ &\quad + \tfrac{1}{1-\alpha}\theta(p)n_{i,t-1} + \log\{1+\delta\exp(\theta(p)e\tfrac{1}{1-\alpha})\}]\} \end{aligned}$$

Then equilibrium entry decisions form a fixed point of this function, i.e., $e_{it} = e(s_{i,t-1}, n_{i,t-1}, A_{i0})$ solves

$$g(e|s_{i,t-1}, n_{i,t-1}, A_{i0}) = e \tag{28}$$

The intercept of this function is exactly the entry that results in the myopic equilbrium with $\delta = 0$. The function is increasing in e, with a slope

$$g'(e|s_{i,t-1}, n_{i,t-1}, A_{i0}) = s_{i,t-1}\frac{\delta\exp(\frac{\theta(p)e}{1-\alpha})}{1+\delta\exp(\frac{\theta(p)e}{1-\alpha})}\frac{k\theta(p)}{(1-\alpha)(1-\sigma)} \tag{29}$$

Hence if

$$\frac{k\theta\bar{p}}{(1-\alpha)(1-\sigma)} < 1 \tag{30}$$

where $\bar{p}$ is an upper bound for p, an equilibrium exists and is unique. Computing the equilibrium is easy, as it involves solving for fixed points of a contracting mapping defined recursively by past entry decisions. It can be easily verified that entry is rising in $s_{i,t-1}, \theta(p)$ and $n_{i,t-1}$, just as in the myopic entry case.

Appendix B. Derivation of the Adjusted HHI

Suppose that there are n trials, that each outcome j from the set of k possible outcomes has an independent probability of occurring p_j, and that the random variable X_j is the number of occurrences of outcome j. Then the multivariate random variable $\mathbf{X} = (X_1, \cdots, X_k)$ has a multinomial distribution with parameters $(n, k, p_1, \cdots, p_k)$. Applied to our context, (i) n is the total number of firms that enter from a given birth county in a given period, (ii) k is the total number of sectors or destinations that they select into, and (iii) $p_1, \cdots, p_k$ are the probabilities that a firm choosing independently would select each of those sectors or destinations. We assume that there is an equal probability of choosing any sector or destination; $p_j = \frac{1}{M}, \forall j$.

The expected HHI when migrants move independently can be expressed as,

$$E(HHI) = E\left(\frac{1}{n^2}\sum_{i=1}^{k} X_i^2\right) = E\left(\frac{1}{n^2}\mathbf{X}^T\mathbf{X}\right).$$

Based on the general properties of the multinomial distribution,

$$E(HHI) = \frac{1}{n^2}\left([E(\mathbf{X})]^T E(\mathbf{X}) + tr[cov(\mathbf{X})]\right).$$

It follows that,

$$E(HHI) = \frac{1}{n^2}\left(k\left(\frac{n}{k}\right)^2 + k\left[n\frac{1}{k}\left(1 - \frac{1}{k}\right)\right]\right) = \frac{1}{k} + \frac{1}{n}\frac{k-1}{k}.$$

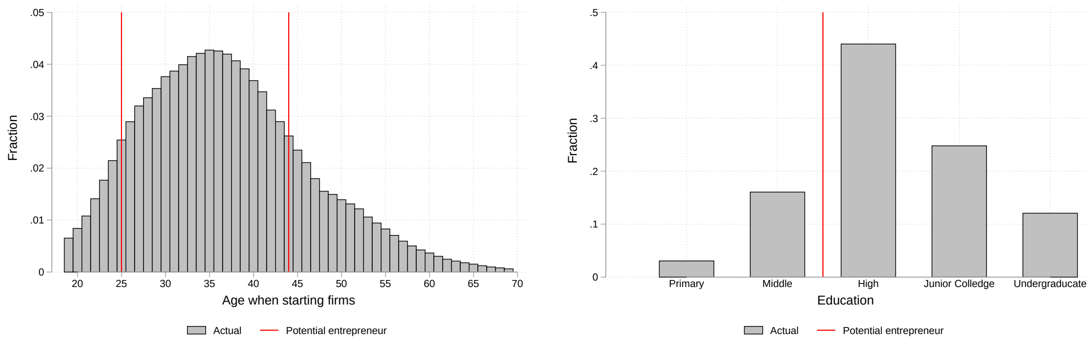
For large k, the fraction $\frac{k-1}{k}$ is close to one, and

$$E(HHI) \approx \frac{1}{k} + \frac{1}{n}. \tag{31}$$

For large n, $E(HHI) \approx \frac{1}{k}$. For small n, $E(HHI)$ is decreasing in n. We account for this by constructing a normalized HHI statistic, which is simply the unadjusted HHI, based on the observed distribution of firms across sectors or destinations, divided by $E(HHI)$. If firms make decisions independently, then the adjusted HHI will be close to one, providing us with a useful benchmark for this statistic.

Appendix C. Figures and Tables

Figure A.1. Education and Age Distribution of Entrepreneurs



(a) Age **(b)** Education

Source: SAIC registration database.

Table A.1. Marginal and Average Initial Capital

Dependent variable:	marginal initial capital				average initial capital			
Time period:	1990-1994	1995-1999	2000-2004	2005-2009	1990-1994	1995-1999	2000-2004	2005-2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth county population density	-0.242***	-0.294***	-0.277***	-0.180***	-0.031	-0.044***	-0.019***	0.034***
	(0.030)	(0.026)	(0.017)	(0.010)	(0.021)	(0.011)	(0.006)	(0.007)
Constant	-0.725***	-1.063***	-1.676***	-2.239***	-0.267***	-0.339***	-0.541***	-0.722***
	(0.031)	(0.021)	(0.015)	(0.010)	(0.025)	(0.011)	(0.007)	(0.007)
Observations	15,601	46,877	83,276	99,877	15,601	46,877	83,276	99,877

Note: marginal initial capital defined by the bottom one percentile of the initial capital distribution at the birth county-sector-time period level. Average initial capital is the mean of the distribution.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Initial capital is obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Table A.2. Marginal and Average Initial Capital (with controls)

Dependent variable:	marginal initial capital				average initial capital			
Time period:	1990-1994	1995-1999	2000-2004	2005-2009	1990-1994	1995-1999	2000-2004	2005-2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth county population density	-0.209***	-0.197***	-0.176***	-0.044***	-0.035	-0.012	-0.003	0.063***
	(0.033)	(0.024)	(0.017)	(0.010)	(0.024)	(0.012)	(0.007)	(0.008)
Birth county population	-0.226	-0.632***	-0.636***	-0.829***	0.024	-0.211***	-0.104***	-0.178***
	(0.141)	(0.107)	(0.064)	(0.037)	(0.101)	(0.048)	(0.028)	(0.026)
Birth county education	-0.003	-0.000	-0.006***	-0.008***	0.002	0.002**	-0.001	-0.003***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Constant	-0.402**	-0.704***	-0.977***	-1.308***	-0.405**	-0.384***	-0.446***	-0.455***
	(0.195)	(0.119)	(0.084)	(0.056)	(0.160)	(0.069)	(0.044)	(0.045)
Observations	15,601	46,877	83,276	99,877	15,601	46,877	83,276	99,877

Note: marginal initial capital defined by the bottom one percentile of the initial capital distribution at the birth county-sector-time period level. Average initial capital is the mean of the distribution.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Population is measured in millions and education is measured by the percent of the population that is literate.
Initial capital is obtained from the SAIC registration database and population density is derived from the 1982 population census.
Standard errors clustered at birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.