The Community Origins of Private Enterprise in China*

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

February 15, 2020

Abstract

This paper identifies and quantifies the role played by birth-county-based community networks in the growth of private enterprise in China. The starting point for the analysis is the observation that population density is positively associated with social homogeneity, local social interactions, and enforceable trust in counties (but not cities). This motivates a model of network-based spillovers that predicts how the dynamics of firm entry, concentration, and firm size vary with birth county population density, independently of other factors such as government infrastructure and agglomeration effects. The predictions of the model are validated over the 1990-2009 period with administrative data covering the universe of registered firms. Competing non-network-based explanations can explain some, but not all of the results. We subsequently estimate the structural parameters of the model and conduct counter-factual simulations, which indicate that entry over the 1995-2004 period and capital stock in 2004, for the economy as a whole, would have been 11% and 12.5% lower without the rural hometown networks that we analyze. Additional counter-factual simulations shed light on misallocation and industrial policy.

Keywords. Community Networks. Trust. Entrepreneurship. Misallocation. Informal Institutions. Growth and Development.
JEL. J12. J16. D31. I3.

*We are grateful to numerous seminar and conference participants and discussants for their constructive comments. Research support from the Economic Development and Institutions (EDI) RA4 program, Cambridge-INET, China National Science Foundation (project number 71874008) and the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010 is gratefully acknowledged. We are responsible for any errors that may remain.

†Peking University dairuochenpku@gmail.com

‡Boston University dilipm@bu.edu

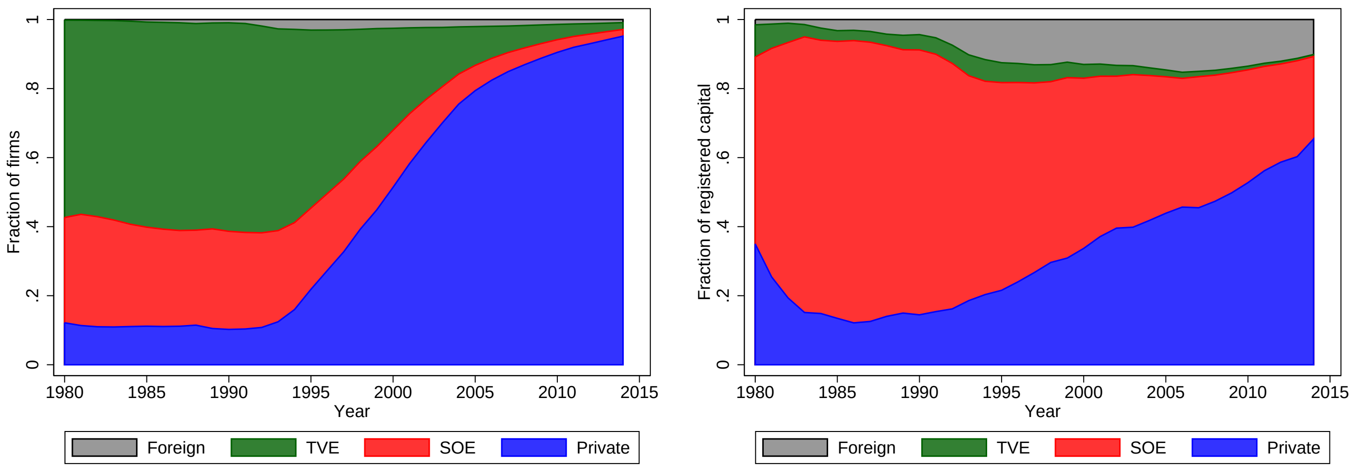
§Yale University kaivan.munshi@yale.edu

¶Peking University x.zhang@nsd.pku.edu.cn

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 (as documented in Figure 1a). As with their numbers, the share of registered capital held by private firms grew steeply from the early 1990's onwards and by 2014 they held 60% of total registered capital in the economy (see Figure 1b).

Figure 1. Distribution of Firms, by Type



(a) Number of Firms

(b) Registered Capital

Source: State Administration of Industry and Commerce (SAIC) registration database.

What is perhaps most striking about the growth of private enterprise in China is that it occurred without the preconditions generally believed to be necessary for market-based development; i.e. without effective legal systems or well functioning financial institutions (Allen *et al.*, 2005). While the government played an important role in China's economic transformation by providing infrastructure and credit (Long and Zhang, 2011; Wu, 2016), it has been argued that informal mechanisms based on reputation and trust must have been at work to allow millions of entrepreneurs, most of whom were born in rural areas, to establish and grow their businesses (Peng, 2004; Allen *et al.*, 2005; Song *et al.*, 2011; Greif and Tabellini, 2017; Zhang, 2017). Case studies of production clusters, which are a distinctive feature of the Chinese economy, indicate that long-established relationships among relatives and neighbors (from the rural origin) substitute for legally enforced contracts between firms and fill in for missing markets (Fleisher *et al.*, 2010; Nee and Opper, 2012). Our paper advances this line of research by identifying and quantifying the role played by informal community-based business networks in the growth of private enterprise for the economy as a whole.

Our analysis proceeds in four steps. First, we construct and validate a population-based measure of trust in Chinese counties, which we posit will determine the quality of the business networks that emerge from these rural origins and locate in different destinations. Second, we develop a model of network-based spillovers that predicts how the dynamics of firm entry, sectoral and spatial concentration, and firm size will vary with the strength of origin-based networks. Third, we utilize administrative data covering the universe of registered firms, rather than just the largest firms as in previous studies, to test these predictions over the 1990-2009 period. We argue that the results, taken together, cannot be explained by alternative non-network mechanisms based on origin or destination heterogeneity. Finally, we estimate key parameters of the model and then use these structural estimates to quantify the role of the hometown networks in the growth of private enterprise, as well as to assess the impact of possible policy interventions that exploit their presence.

The first step in the analysis is to construct a population-based measure of trust. We argue that population density in rural areas is positively associated with the frequency of local social interactions, which under plausible assumptions on the matching process, give rise to more inter-connected social networks (Coleman, 1988; Jackson *et al.*, 2012). Inter-connected networks sustain greater economic cooperation via norms based on community enforcement (Greif, 1993, 1994; Greif and Tabellini, 2017). The localized trust that we focus on is distinct from the generalized trust that has received much attention in the rapidly growing economics literature on culture (Alesina and Giuliano, 2015). Localized trust is restricted to neighbors rather than the general population, and is sustained via external enforcement rather than by internalized cultural values. We provide evidence linking population density to localized (but not generalized) trust, at the country level with data from the World Values Survey (WVS) and at the county level with nationally representative data from the China Family Panel Survey (CFPS). While we uncover a robust positive relationship between population density and both social interactions and trust in Chinese counties, this relationship does not extend to cities. This can be explained by the offsetting effect of social heterogeneity, which is increasing in urban (but not rural) population density. The empirical analysis, linking birthplace population density to entrepreneurship, consequently focuses on county-born businessmen; their firms account for two-thirds of all registered private firms, and a comparable share of private registered capital, in China. Although the majority of county-born businessmen establish their firms outside their birth counties, our assumption is that business networks drawn from higher population density counties will support higher levels of mutual cooperation among their members regardless of where they are located.[1]

The second step in our analysis is to develop a model of community-based entrepreneurship whose predictions can be subsequently tested. The model features successive cohorts of agents that make a choice between a traditional occupation (such as farming or wage labor) and entrepreneurship. Individual abilities are drawn from an i.i.d. process and affect returns to both occupations. The payoff from entrepreneurship depends additionally on the contribution of the network to productivity – what we refer to as community TFP (CTFP) – via mutual help. Help provided by different network members from the same origin is complementary, which implies that there are increasing returns to network size. When sector-location choice is incorporated in the model, an additional channel for network-based spillovers opens up, via a pre-entry referral process which increasingly directs entering firms from a given origin into an initially favored destination (a term we use to denote either sectors or locations). The interaction between the two types of spillovers generates dynamic increasing returns to network size in any given destination, resulting in increased entry overall and increased

[1]The implicit assumption is that county-born entrepreneurs remain connected to, and can be sanctioned by, their origin communities. This assumption is supported by a well established sociological literature on migration in China; e.g. Honig (1992), Goodman (1995) and the economics literature on temporary migration; e.g. Morten (2019).

sectoral and spatial (within sector) concentration over time. Once network quality, denoted by the level of help for given network size, is introduced in the model, the additional prediction is that entry and concentration will be rising in network quality at each point in time, with a slope that increases over time at early stages of the industrialization process. Network quality is measured by origin population density in the empirical analysis.

With regard to capital investment, the network-based spillovers that raise CTFP over time have two conflicting effects on the initial size of the marginal entrant's firm: the direct effect, for a given level of ability, is to increase firm size by raising the firm's TFP, but an increase in CTFP also lowers the ability threshold for entry into entrepreneurship and this negative selection works in the opposite direction to lower TFP. We show that the latter effect dominates; the marginal entering firm from a higher population density birth county will be unambiguously smaller, with this negative relationship growing stronger over time as networks get larger. Under specific conditions on the model's parameters, this result is shown to hold for average initial 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 over time and is the same for all firms in a network at a given point in time, regardless of their cohort or the ability of the entrepreneur. Because (higher quality) networks from higher population density birth counties are growing faster, firms from those counties will start small but subsequently grow faster.

The third step in the analysis tests these predictions over the 1990-2009 period, with administrative data obtained from the State Administration of Industry and Commerce (SAIC). These data, which cover the universe of registered firms in China, are linked to the industrial census and the SAIC inspection database for the analysis of firm growth. The registration database includes the following information for each private firm: establishment date, sector, location, registered capital, 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" for our analysis. The model generates cross-sectional and dynamic predictions for the relationship between birth county population density, which we measure in 1982 (prior to the onset of privatization) and a rich set of outcomes – firm entry, sectoral and spatial concentration, initial firm size, and firm growth – and the data match each of them. While these results are consistent with the presence of underlying community networks (of varying quality) the empirical challenge is to establish that birth county population density is not correlated with other factors that independently determine the outcomes of interest. Population, education, and the occupational structure in the birth county (measured in 1982) are included as exogenous controls in the estimating equations, but such conditioning may not account for all relevant factors. An important feature of our analysis is that we comprehensively examine alternative non-network explanations by systematically relaxing different assumptions of the model. We show that no other explanation can account for all of our results simultaneously.

In general, alternative explanations are generated by introducing additional sources of heterogeneity at the origin and the destination. Starting at the origin, the alternative explanation that matches our results most closely is one in which options outside business are inferior in higher population density counties and worsening disproportionately over time. This will generate the increased entry and the worsening marginal ability (with consequences for initial firm size) from denser birth counties that are predicted by the model and documented in the data. However, this explanation and, for that matter, any explanation based on origin heterogeneity, cannot explain the increased clustering over time, in particular sectors and locations, that is predicted and documented for firms from higher population density birth counties. Moreover, origin heterogeneity, even if it is changing over time, cannot explain why firms from those counties start smaller but then grow faster. There must be some force at the destination that is giving these firms a boost. In our model, it is the endogenously evolving network, but an alternative non-network mechanism based on destination heterogeneity could also generate this result if firms from higher population density birth counties have preferred access to destinations that were exogenously growing faster for other reasons, such as improved infrastructure or destination-based productivity spillovers, as emphasized in the literatures on endogenous growth and agglomeration.[2] To address this concern, we exploit the fact that firms from many birth counties are established in the same sector-location. Sector-time period and location-time period effects can thus be included as controls in the estimating equations. We find that the estimated relationships get stronger when sector and location effects are incorporated in the analysis. This implies that entrepreneurs from higher population density birth counties are, if anything, selecting into less favorable sectors and locations.

In our model, new entrants are channelled into particular destinations (sectors and locations) where their network is established. Although the results on entry and concentration are consistent with this mechanism, we verify it directly by estimating the effect of initial entry in 1990-1994 on subsequent entry, separately in 2000-2004 and 2005-2009, at the level of the birth county-sector-location. Consistent with the dynamic network multiplier effect that is implied by the model, we find that initial entry has a positive effect on subsequent entry, which is increasing over time. Moreover, the initial entry effect is increasing in population density. This last result, together with the reduced form evidence that firms from denser counties do not have preferred access to faster growing destinations (which could otherwise generate a spurious inter-temporal correlation) provides further support for the network mechanism. In addition, we find no evidence of cross-community externalities: conditional on the number of initial entrants from the birth county, the total number of initial entrants in a sector-location (aggregating across all origins) has no predictive power. This is consistent with the hypothesis that the birth county is the domain within which business networks are organized in China, rather than destination sectors or locations, and with the model's assumptions that networks are operating independently and cannot influence the product price. If networks were using their market power to extract monopoly rents (Brooks *et al.*, 2016) or competing with each other for subsidized credit from the local government (Bai *et al.*, 2019), then additional entry from other origins would have lowered the returns for their members, resulting in lower entry and a negative cross-community effect.

Having tested and validated the network-based model, the final step in the analysis seeks to quantify the impact of these networks on aggregate firm entry and capital stocks by estimating the structural parameters of the model. Given that firms from multiple origins were established in each destination (sector-location) and that the structural equations are linear in variables, it is possible to control completely flexibly for local government and agglomeration effects by including destination-time period dummies in the estimating equation.

[2] For example, Faber (2014) describes how the construction of China's National Trunk Highway System provided unanticipated economic benefits to particular local areas. Similarly, a voluminous theoretical and empirical literature describes the positive impact of inter-firm spillovers, within sectors and locations, on firm productivity; e.g. Romer (1990, 1986); Segerstrom *et al.* (1990); Aghion and Howitt (1992); Jones (1995); Segerstrom (1998); Ciccone and Hall (1996); Ellison and Glaeser (1997); Au and Henderson (2006) and Combes *et al.* (2012).

Consistent with the reduced form results, which find no evidence that entrepreneurs from higher population density birth counties had preferred access to faster growing destinations, the destination controls have absolutely no effect on the estimated structural parameters. Although the model is extremely parsimonious, it does a good job of matching entry and initial capital across the range of birth county population densities, during a period (1995-2004) in which the Chinese economy was growing at an explosive rate. This increases our confidence in the results of a counter-factual experiment, which estimates that entry from county origins would have declined by as much as 22% over the 1995-2004 period, with an even larger 28.5% decline in the total capital stock in 2004, had the rural hometown networks not been active. Adjusting for exogenous initial entry and the fact that firms from county origins account for approximately two-thirds of all registered firms in China, this amounts to an impact of approximately 11% and 12.5% on entry flows and capital investment, respectively, for the entire economy. Given the dynamic increasing returns that are generated by the networks, the long-term consequences of their absence would have been even more substantial.

We conclude by discussing the implications of the networks for industrial policy, and for the growing literature on firm misallocation in developing countries. This literature documents two sets of stylized facts: (i) the variation in marginal productivity and, hence, firm size within narrow sectors is especially wide in developing economies, and (ii) firms in those economies are small (Peters, 2016). Although a number of mechanisms can explain these facts; e.g. Caunedo (2016), Asker *et al.* (2014), Akcigit *et al.* (2016), Haltiwanger *et al.* (2018), perhaps the simplest is based on a model with mark-ups in output prices and wedges in factor prices (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009, 2014; Peters, 2016). There are no price distortions in our model. Dispersion in firm size and productivity are a consequence, instead, of community-based inter-firm spillovers. The results of a second counter-factual policy experiment, which exploits these spillovers by providing a temporary credit subsidy to entering firms, suggests that optimal second-best policies should target subsidies to more socially connected communities. This would increase dispersion and the proportion of small firms even further. In light of these findings, we would not want to infer that one developing economy is less efficient than another developing economy because it has smaller firms or greater dispersion in firm size. Indeed, these characteristics may be symptomatic of a more dynamic economy in which underlying community networks are responding more effectively to market frictions, turning the usual tests of misallocation on their head.

2 Trust and Population Density

The point of departure for our analysis is the assumption that localized trust in relatively sparsely populated rural areas is increasing in population density. The underlying idea is that higher population density, which is mechanically correlated with greater spatial proximity, raises the frequency of social interactions and facilitates communication with neighbors. This, in turn, helps sustain higher levels of mutual cooperation, supported by the threat of social sanctions, as argued in early papers on social norms and community enforcement (Greif, 1993, 1994; Kandori, 1992; Ellison, 1994). To make this argument more precise, consider a random graph model in which the probability that an individual is connected to any other individual in a local population, γ, is rising in population density. A higher γ directly raises the degree of the social network (the number of links per capita), and indirectly also network inter-connectedness i.e. the probability that friends of friends are linked, and so on. For example, the rate of triadic closure – the probability that any three individuals are directly linked – is increasing in γ. Coleman (1988) argues that network closure is a necessary condition for economic cooperation, enforced by social sanctions. Jackson *et al.* (2012) make a similar argument based on a related network property, which they refer to as "support." These results do not necessarily rely on the random matching assumption, and are also likely to hold if the matching process exhibits homophily.[3]

The preceding discussion implicitly assumes that the population is socially homogeneous. Matters are more complex when the population is fragmented into smaller communities. Suppose that individuals only interact within their communities and that social sanctions consequently only apply within those communities. If communities are not (perfectly) spatially segregated, then the frequency of social interactions, the effectiveness of social sanctions, and resulting enforceable trust are all decreasing in social heterogeneity. If social heterogeneity is (weakly) decreasing in population density, as documented below in Chinese counties, then both social interactions and enforceable trust will be unambiguously increasing in population density. However, social heterogeneity could potentially be increasing in population density in cities. Most urban residents in developing countries are recent arrivals, typically from many different origins. Greater population density in an urban neighborhood might well be associated with a larger migrant presence, and a greater diversity of origins, in which case the relationship between population density and both social interactions and enforceable trust will be ambiguous.

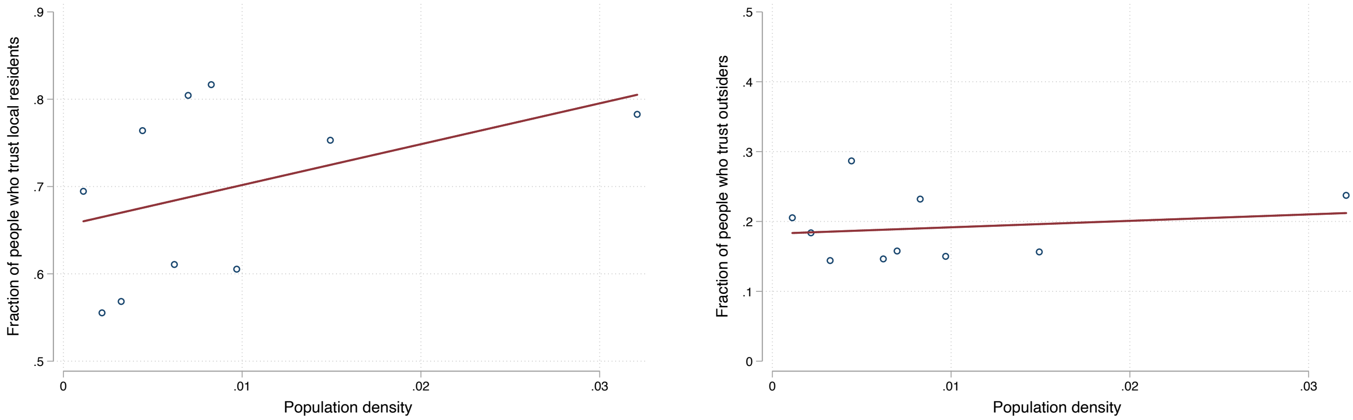
To provide empirical support for each component of the preceding argument, we begin by estimating the relationship between trust and population density. This relationship is not China-specific and, hence, we expect it to be observed across a wide cross-section of countries. We first measure trust with data from the World Values Survey (WVS); while the advantage of these data is that they cover many countries, one limitation is that responses from rural and urban residents cannot be distinguished. We (partially) address this limitation by only including large developing countries, with a large rural population, in the sample.[4] Figure 2a presents a binned scatter plot describing the relationship between trust in local residents, which measures localized enforceable trust, and population density (obtained from the World Development Index) for the 31 countries in our restricted sample. This relationship is strongly positive and statistically significant (based on regression estimates not reported). Figure 2b presents a binned scatter plot describing the relationship between trust in outsiders; i.e. individuals that the respondent would meet for the first time, which measures generalized trust, and population density. No relationship can now be detected and this is also true for the companion regression estimates (not reported).

Figures 3a and 3b subject the preceding results to closer scrutiny by repeating the analysis with data from the China Family Panel Survey (CFPS). The advantage of the CFPS, apart from the obvious relevance for the China-based analysis of entrepreneurship that follows is that the relationship between trust and population

[3]Jackson and Rogers (2007) describe a model of matching in which some links are formed randomly, whereas other links are formed strategically. The latter process results in clustering or homophily, with individuals having higher degree (more links) more likely to be matched to each other. While adding a strategic element to the matching process will generally increase the inter-connectedness of the network, the relationship between population density and inter-connectedness now becomes more complex. Nevertheless, we continue to expect this relationship to be positive by a limit argument; as γ goes to zero, no one is connected and the rate of triadic closure goes to zero. As γ goes to one, everyone is connected, and the rate of triadic closure goes to one.

[4]The WVS provides the fraction of respondents for a given country in the following categories: trust completely, trust somewhat, trust not very much, trust not at all. We combine the first two categories to construct a measure of trust. Countries with annual GDP per capita exceeding \$20,000, with an area less than $100,000 km^2$, or with missing population data in the World Development Index are dropped from the analysis.

Figure 2. Trust and Population Density: Cross-Country Comparison



(a) Localized trust

(b) Generalized trust

Source: World Values Survey and World Development Index.

density can be estimated separately in counties and cities.[5] Population density in the counties and cities covered by the CFPS, and in all the analysis that follows, is computed from the 1982 population census. The Chinese economy was largely agrarian at that time and, hence, the spatial distribution of population density would have been determined by crop suitability. Rural-urban migration only commenced with privatization in the 1990's, and thus 1982 population density can be treated as predetermined in our analysis of the evolution of private enterprise.[6] The adult individual module of the 2012 CFPS collected information on trust. As with the WVS, the analysis distinguishes between trust in local residents, which measures localized trust, and trust in outsiders, which measures generalized trust.[7] The bin scatter plot presented in Figure 3a indicates that localized trust is increasing in population density in (rural) counties, but not in cities. In contrast, there is a (mildly) declining relationship between generalized trust and population density, in counties and cities, in Figure 3b. Table 1 reports the estimated relationships between trust and population density, controlling for population, education, and occupational structure in the county or city (also measured in 1982). These covariates, which are correlated with population density and could independently determine the outcomes of interest, are included in all specifications in the paper that estimate the direct effect of population density. We see that the only economically and statistically significant parameter estimate in the table is the coefficient on county population density, with localized trust as the outcome.

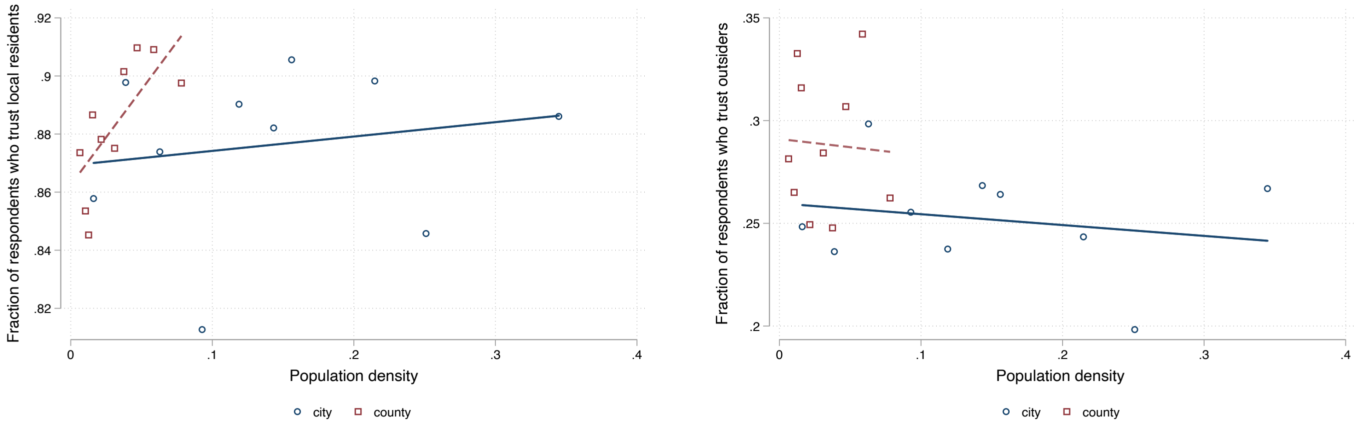
In our framework, localized trust is determined by social interactions. In light of the preceding results, we would expect local social interactions to be increasing in population density in counties but not cities. Table 2 reports the relationship between the frequency of social interactions per month with local residents, obtained from the family module of the 2010 CFPS, and population density, separately for counties and

[5]There are approximately 2,000 counties and 250 prefecture-level and province-level cities (which are further divided into urban districts) in China.

[6]Population density is measured in units of 10,000 people per square km. The threshold density in our analysis is set at 0.002; i.e. 20 people per square km. This excludes sparsely populated regions such as Western China, Inner Mongolia, and Tibet, which are inhabited by ethnic minorities with a different culture than the majority Han Chinese.

[7]Trust is measured as an ordinal variable, taking values from 0 to 10 in the CFPS. To construct a binary trust measure that is consistent with the WVS, we selected a cutoff, which turns out to be 5, such that trust in neighbors and strangers obtained from the CFPS matches most closely with the corresponding statistics for China from the WVS.

Figure 3. Trust and Population Density: China



(a) Localized trust **(b)** Generalized trust

Source: China Family Panel Survey and 1982 population census.

Table 1. Trust and Population Density

Dependent variable:	trust in local residents	trust in outsiders	trust in local residents	trust in outsiders
Respondent's location:	county		city	
	(1)	(2)	(3)	(4)
Population density	0.026***	0.004	-0.001	-0.006
	(0.005)	(0.012)	(0.010)	(0.015)
Mean of dependent variable	0.883	0.289	0.876	0.252
Observations	93	93	39	39

Source: Trust measures are obtained from the adult individual module of China Family Panel Survey (2012).
Trust is a binary variable which takes the value one if the value of ordinal variable in the CFPS exceeds a cutoff.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Control variables include population, education and occupation distribution in the county or city.
Population is measured in millions and education is measured by the percent of the population that is literate.
Occupation distribution is measured by the share of workers in agriculture and industry with services the excluded category.
Population density, population, education and occupation distribution are computed from the 1982 population census.
Standard errors clustered at the county or city level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

cities. We divide the different types of social interactions reported in the CFPS into planned interactions; i.e. group entertainment, visits to neighbors' homes, and dining together, and unplanned interactions, which are defined as meetings or conversations without other background activities.[8] The population density coefficient is positive (and statistically significant with unplanned interactions as the dependent variable) for county residents, whereas it is negative (and statistically insignificant) for city residents.

One explanation for the weak relationship between urban population density and both social interactions and localized trust is that dense urban neighborhoods are more socially heterogeneous. We begin in Table 3, Columns 1-2, by establishing that population density at the aggregate – city or county – level, obtained from the 1982 population census, is positively associated with population density at the local – neighborhood or village – level, obtained from 2010 CFPS. Next, we compare the fraction of locally born residents in counties and cities in Columns 3-4. Based on the mean of the dependent variable, these fractions are very different;

[8]Group entertainment includes playing *mahjong* or cards, reading newspapers, listening to the radio, or watching TV with others.

Table 2. Frequency of Local Social Interactions and Population Density

Dependent variable:	average frequency of social interactions per month with local residents			
Respondent's location:	county		city	
Type of social interactions	planned	unplanned	planned	unplanned
	(1)	(2)	(3)	(4)
Population density	0.391	1.640**	-0.774	-0.098
	(0.415)	(0.821)	(0.505)	(1.184)
Mean of dependent variable	4.027	15.70	3.700	12.92
Observations	93	93	39	39

Source: Family module of China Family Panel Survey (2010).
Planned interactions include group entertainment, visits to neighbors' homes, and dining together. Unplanned interactions are one-on-one social meetings without other background activities.
We take the average frequency of social interactions for each county/city.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Control variables include population, education and occupation distribution in the birth place.
Population density, population, education and occupation distribution are computed from the 1982 population census.
Standard errors clustered at the county or city level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

while around 90% of county residents are born there, this statistic is lower than 50% in cities. Moreover, the fraction of locally born residents is decreasing in local population density in cities but not counties.

Based on the preceding result, the weak predictive power of urban population density can be attributed to its offsetting effect on social heterogeneity. While the fraction of migrants does not increase in the same way with local population density in counties, there is another dimension of social heterogeneity – based on lineages or clans – that must be considered in rural areas. Counties in China are divided into villages, which consist, in turn, of one or more clans (Peng, 2004; Tsai, 2007). Members of a clan will share the same surname (Peng, 2004) and the 2010 CFPS provides the fraction of the village population that has the most popular surname. We see in Table 3, Column 5 that local population density has a positive and significant (at the 10 percent level) effect on this measure of social homogeneity. Clans historically supported the business activities of their members, who were bound together by mutual moral obligations. It has been argued that this role has re-emerged in the post-collectivist era (Peng, 2004; Zhang, 2017; Greif and Tabellini, 2017). We remain agnostic about the boundary of the social unit within which trust is sustained and from which business networks are drawn in this paper; i.e. whether it is the county or the clan. Given that clan concentration is increasing in population density, trust will be increasing in population density regardless of whether it is sustained at the level of the clan or the county.

Given the absence of an association between trust and urban population density, our analysis of community-based entrepreneurship will focus on county-born businessmen, with city-born businessmen serving as a useful placebo group in the analysis.[9] Figure 4 uses SAIC data to describe the growth in the total number of private registered firms and in the number of firms owned by county-born entrepreneurs, respectively, over the 1990-2014 period. County-born entrepreneurs made up about two-thirds of all entrepreneurs in China, with this ratio remaining stable over time. Firms owned by county-born entrepreneurs are just slightly smaller than the average registered firm (not reported). The contribution of these entrepreneurs, most of whom are

[9]This does not imply that city-born entrepreneurs do not have access to networks; based on the evidence provided above, population density is just not a good indicator of network quality for them.

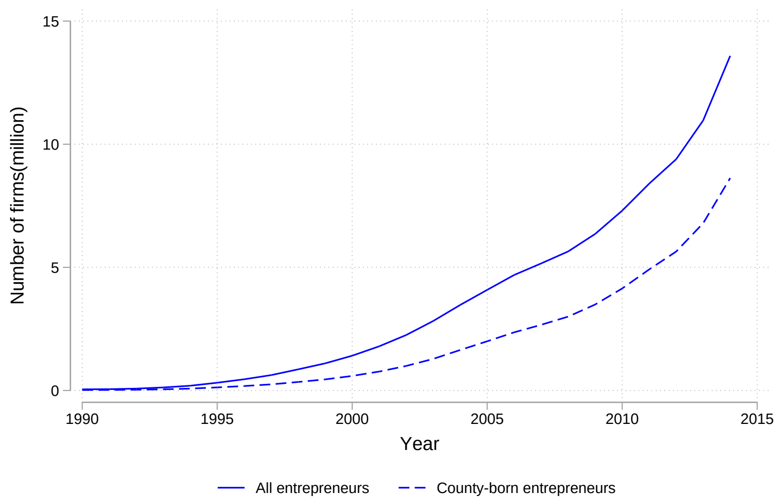
Table 3. Social Structure and Population Density

Dependent variable:	population density at neighborhood/village level		fraction of residents born locally		fraction of households using the most common surname
Respondent's location:	county	city	county	city	county
	(1)	(2)	(3)	(4)	(5)
Population density at city/county level	2.076***	2.763***	–	–	–
	(0.548)	(1.016)			
Population density at neighborhood/village level	–	–	1.550	-7.866***	18.601*
			(2.109)	(2.468)	(9.609)
Mean of dependent variable	0.112	0.794	90.16	47.90	42.46
Observations	265	134	265	134	265

Source: China Family Panel Survey (2010).
Clan affiliation is measured by surname. Population density at neighborhood/village level is computed from the community module of the CFPS. Population density at city/county level is computed from the 1982 population census.
Population density is measured in units of 10,000 people per square km.
Robust standard errors are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

first-generation businessmen, to Chinese economic growth has thus been substantial.

Figure 4. Growth of Private Enterprise, by Birthplace of Entrepreneurs



Source: SAIC registration database.

Entrepreneurs in developing countries often rely on other entrepreneurs operating in the same sector and location for different forms of help. Chinese production clusters, for example, are characterized by a high degree of specialization by firms in specific stages of production, extensive exchange of intermediate goods, and flexible adjustment to workloads and product specifications in response to volatile market demand. Entrepreneurs also rely on each other for connections to buyers and suppliers (who often provide credit) as well as for information about new technologies and markets. This type of informal support is difficult to sustain through the market mechanism, due to the inherent problem of verifying help sought and received, coupled with a weak legal environment. Cooperation is based instead on community norms, backed by social ties among the entrepreneurs in question (Nee and Opper, 2012). The key assumption in our analysis is that these social ties are based on the birth county, or the clan within the birth county, despite the fact that a majority of county-born entrepreneurs establish their firms elsewhere.

There are many accounts of 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). These accounts describe how migrant networks are organized around the rural hometown, complementing an older sociological literature that takes the position that ethnicity in China is defined by the native place; e.g. Honig (1992, 1996), Goodman (1995).[10] If the home place is the domain around which migrant labor networks are organized, then it is plausibly the domain around which business networks supporting county-born entrepreneurs are organized.[11] Migrant entrepreneurs typically have close family members, such as aged parents, in the rural origin, and visit it frequently. They (or close family members) thus continue to interact socially with the origin community. It follows that cooperation at the destination (sector-location) where firms are operating can be backed effectively by social sanctions imposed by the wider community in the origin. The model that follows derives the relationship between network quality, which is measured by population density in the birth county, and a wide range of business decisions and outcomes.

3 The Model

3.1 Population and Technology

The population is comprised of a large number of communities, distinguished by their population density, p. As discussed above, the level of trust that can be supported in a community, and in the networks that are drawn from that community, is increasing in p. Each community consists of a continuum of agents, with equal sized cohorts of new agents born at successive dates t=1,2,... Agents vary in individual ability ω, which is drawn independently from a log uniform distribution on the unit interval. The ability distribution is identical across cohorts and communities.

Each agent makes a once-and-for-all choice between a traditional occupation (such as farming or wage labor) and becoming an entrepreneur. The returns to entering the traditional occupation for an agent of ability ω is ω^σ where $\sigma \in (0,1)$. There are multiple destinations $B_i, i = 1, 2..$ for entrepreneurship, with destinations denoting sectors or locations. For simplicity we assume these destinations are *ex ante* symmetric, except for entry at the initial date. In destination B_i at date t, an entrepreneur with ability ω selects capital size K, and 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 (CTFP); i.e. the contribution of the network to the firm's productivity. $A_{it} = A_0(1 + h(p))^{n_{i,t-1}}$, where $h(p)$ denotes per-member help provided by members of the network to one another, which is increasing in p, and $n_{i,t-1}$ measures the stock of entrepreneurs from the community who are already established in sector i by the end of period $t-1$ (and thus in a position

[10]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" (Fang, 1997; Ma and Xiang, 1998; Zhang and Xie, 2013). 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.

[11]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).

to provide support to the cohort of new entrants that follows).[12] This specification captures the idea that help provided by network members is mutually complementary, which implies, in turn, that there are increasing returns with respect to the size of the network, $n_{i,t-1}$ (for given $h(p)$). If the help provided by one network member to another is observed by other network members and entrepreneurs remain connected to their origin communities, then the maximal incentive compatible level of help (for given network size) will be increasing in origin population density p; i.e. $h'(p) > 0$. Letting $\theta(p)$ denote $\log(1 + h(p))$, the preceding expression reduces to

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

where network quality, $\theta(p)$, is increasing in p.

In our model, networks are based on the social origins of entrepreneurs rather than the destinations they select: θ reflects social connectedness in the birth county and n_{it} is the number of firms from that county operating in a given destination. The implicit assumption in our formulation of the production function, which we validate empirically, is that networks from different birth counties operate independently at a given destination. In the standard agglomeration model, the θ parameter would measure exogenous destination characteristics and n_{it} would be the total number of firms operating in that destination, irrespective of the social origin of their respective entrepreneurs. Ciccone and Hall (1996), for instance, use the number of workers per square km as a proxy for agglomeration effects in a given location. We will exploit this difference in the empirical analysis to distinguish between birth county network effects and agglomeration effects.

The dependence of CTFP, A_{it}, on the size of the incumbent stock represents one source of network complementarity, reflecting gains from intra-network cooperation in improving productivity for those who have already entered destination i. We add to this a second source of network complementarity, which pertains to 'referrals' or 'access' to particular business sectors or locations. A fixed fraction $k \in (0, 1)$ of new agents in every cohort receive an opportunity to become an entrepreneur. Within this group of 'potential entrants', the fraction that get an opportunity to enter destination B_i equals $s_{i,t-1}$, the share of incumbent entrepreneurs from the origin community already in that destination. This reflects the formation of aspirations, access to information, or referrals provided by older members from the same origin community in a given destination.

Apart from the decision of whether or not to enter a given destination when presented with the opportunity, an agent decides on how much capital to invest. 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.[13] We also assume a fixed price of the product, 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

[12]The A_0 term incorporates the product price and labor productivity. Labor is not included as a variable input in the production function because it is not observed in our data. With the Cobb-Douglas specification of the production function, the optimal labor input can be derived as a function of the model's parameters and is subsumed in the A_0 term.

[13]We ignore the role of labor networks in the model. The owner of the firm and the workers rarely belong to the same community, even in network-based economies. The historical and contemporary experience, across the world, indicates that incumbent workers (with a reputation to maintain within their firms) are the primary source of job referrals.

most sectors are comprised of a large number of origin county networks, and both domestic and international market opportunities are large.[14]

3.2 Occupational Choice

To determine occupational choice, we first calculate the profits a new agent in any cohort with a given ability ω expects to earn upon entering a given business destination (sector and location) when the CTFP in that destination is expected to be A. The latter is a sufficient statistic for the specific date, destination in question, and 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 $\psi \equiv \phi^{\alpha} - \phi$).[15]

Of the new agents receiving an offer, the ones that will decide to enter business are those who receive a higher profit in that destination compared to the traditional occupation. These agents will be endowed with a level of ability that exceeds a threshold $\underline{\omega}$:

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

We assume that the threshold lies in the interior of the support of the ability distribution at the beginning of the process for each destination, and we will restrict attention to 'early phases of industrialization' when this continues to be true.

Notice 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. 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. If anything, the myopic model generates a conservative bias in entry decisions. This is because a network's size cannot ever decrease over time and its quality does not change, and neither do profits in the traditional

[14] 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 statistic is based on all entrepreneurs, including those who locate their firms in their county of birth.

[15] If we allowed for credit networks organized around the origin county and parameterized the interest rate as $r = r_0 exp(-\eta(p)n_{i,t-1})$, then the productivity channel operating through the A term and the credit network channel would not be separately identified. Although the model is set up so that networks operate through the productivity channel, all the results that follow would go through if, instead, they operated through the credit channel.

sector. Those deciding to enter based on a myopic calculation would also want to enter if farsighted. Others, who decided to stay out on myopic grounds, might now wish to enter when they anticipate future network growth, which would further raise the returns to entrepreneurship.

3.3 Dynamics of Entry and Concentration

We make the simplifying assumption that the different business destinations have identical 'fundamentals'. 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. These historical entry levels will generically not be exactly balanced across destinations; without loss of generality suppose $n_{i0} > n_{i-1,0} > 0$ for all i. We show below that the initial imbalance across destinations will cumulate thereafter, with entrants in later cohorts increasingly locked-in to the destinations with higher initial presence.

To derive entry in subsequent cohorts, we start with the threshold condition (5), which determines the measure of agents from cohort t who would choose to enter destination B_i if they had the opportunity. Combining this with the fraction $ks_{i,t-1}$ of those agents that have an opportunity to enter, we derive the volume of entry e_{it} in cohort t into B_i as a function of the state variables $n_{i,t-1}, s_{i,t-1}$:

$$e_{it} = ks_{i,t-1}[B + C\theta(p)n_{i,t-1}]$$

where $B \equiv 1 - \frac{1}{1-\sigma}\log\frac{1}{\psi} - \frac{\alpha}{(1-\sigma)(1-\alpha)}\log r + \frac{1}{(1-\sigma)(1-\alpha)}\log A_0$ and $C \equiv \frac{1}{(1-\sigma)(1-\alpha)}$. This expression reduces further to

$$e_{it} = Ls_{i,t-1} + \kappa(p)N_{t-1}s_{i,t-1}^2 \tag{6}$$

where L denotes kB; $\kappa(p)$ denotes $Ck\theta(p)$ which is rising in p, and $N_{t-1} \equiv \sum_i n_{i,t-1}$ denotes the aggregate number of entrepreneurs from past cohorts from the same origin. Aggregating (6) across destinations, we obtain an expression for the dynamics of aggregate entry:

$$N_t - N_{t-1} \equiv E_t \equiv \sum_i e_{it} = L + \kappa(p)N_{t-1}H_{t-1} \tag{7}$$

where $H_{t-1} \equiv \sum_i s_{i,t-1}^2$ denotes the Herfindahl Hirschman Index for concentration at $t-1$. Equations (6,7) define the dynamics of the vector $(N_t, s_{it}, i = 1, 2..)$, where $s_{it} \equiv s_{i,t-1}\frac{N_{t-1}}{N_t} + \frac{e_{it}}{N_t}$.

Proposition 1 *Concentration H_t and aggregate entry flow E_t are both rising in t.*

The proofs of this and subsequent propositions are provided in Appendix B. The intuitive reason for concentration to rise over time is simple: a destination with higher incumbent stock is both more profitable and generates more opportunities for entry, so its share grows faster, reinforcing the higher initial presence. The network complementarity associated with post-entry productivity spillovers, embodied in the N_{t-1} term in (7), is reinforced by the network complementarity associated with the referrals; i.e. the H_{t-1} term. Entry E_t will rise over time from (7) if concentration is increasing over time.

The compounding network effect is stronger for firms from higher p origins, on account of the $\kappa(p)$ multiplier, so one would also expect the level of concentration and entry, and their growth over time, to be rising in p. Verifying this conjecture is more complicated, however, especially with respect to concentration. To illustrate this, consider the case of two destinations $i = 1, 2$. Without loss of generality, assume that destination 1 has a higher initial (and subsequent) presence, in which case $H_t = s_{1t}^2 + (1 - s_{1t})^2$ is monotonically increasing in s_{1t}. Variation in s_{1t} is thus synonymous with variation in H_t. From equations (6) and (7),

$$\begin{aligned} s_{1t} \equiv [\frac{N_t}{n_{1t}}]^{-1} &= [\frac{L + N_{t-1} + \kappa(p)N_{t-1}H_{t-1}}{Ls_{1,t-1} + n_{1,t-1} + \kappa(p)N_{t-1}s_{1,t-1}^2}]^{-1} \\ &= [1 + (\frac{1}{s_{1,t-1}} - 1)\{\frac{L + N_{t-1} + \kappa(p)N_{t-1}(1 - s_{1,t-1})}{L + N_{t-1} + \kappa(p)N_{t-1}s_{1,t-1}}\}]^{-1} \end{aligned} \tag{8}$$

We use this expression, together with equation (7), which characterizes variation in entry, to derive the following result.

Proposition 2 *With two destinations:*

(a) Entry E_t and concentration H_t are rising 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 $\frac{3}{4}$).

Part (a) confirms that the level of concentration is rising in p at any t, which in turn implies the same for entry flows. Part (b) shows that growth of entry or concentration is rising in p at 'early stages' of the industrialization process; i.e. when concentration is not too high. The qualifier is required because the share of the dominant destination B_1 is bounded above by 1. Thus, the share of the dominant destination cannot be increasing faster forever; eventually, its growth rate will flatten out as the share approaches one.

The results concerning the dynamics of concentration across destinations translate into testable predictions concerning either sectoral or spatial concentration, given that destinations correspond to sectors or locations. We partition the set of destinations into sectors, with each sector consisting of a subset of locations. Proposition 1 can then be extended to show that spatial (location) concentration within any given sector must be rising in t. The same can be shown for sectoral concentration, provided that sectors with higher initial shares are also characterized by higher intra-sectoral spatial concentration at date 0.[16] Moreover, with two locations within any sector, the results on concentration in Proposition 2 apply across sectors or across locations within sectors.

3.4 Ability Selection and Firm Size Dynamics

Next we derive predictions concerning entrepreneurial ability and firm size. From (3), an increase in CTFP, A_{it}, increases initial capital, for a given level of entrepreneurial ability, ω. However, we also know that network effects generate negative selection on ability: from (5), as CTFP increases over time, the threshold for entry falls, and entrepreneurs with lower ability start entering. This negative selection has a negative effect on initial capital from (3). Substituting from (5) in (3), we see that the latter effect dominates and, hence, that the initial capital of the marginal entrant is unambiguously decreasing in CTFP, A_{it}:

$$\log K_{it}^m = U' - \frac{\sigma}{(1-\sigma)(1-\alpha)} \log A_{it} \tag{9}$$

[16]The reason is that the expression for entry flow into sector c is modified to $e_{ct} = \kappa L s_{c,t-1} + \kappa N_{t-1} s_{c,t-1}^2 H_{c,t-1}$, so the term involving the quadratic term in lagged sectoral share is weighted by lagged intra-sectoral spatial concentration $H_{c,t-1}$.

where $U' \equiv \log \phi - \frac{1}{1-\sigma} \log \psi - \frac{1}{1-\alpha} \log r$, and $\log A_{it} = \log A_0 + \theta(p) n_{i,t-1}$. The marginal entrepreneurs that enter later in time, when their networks are stronger (with higher CTFP) are thus less productive and have smaller firms.[17] The same argument applies to comparisons at any given t across different p origins: marginal entrants from higher p origins, with stronger networks, enter with smaller firm sizes. If $\sigma \in (\frac{1}{2}, 1)$ this is true also for the average entrant: firms from high p origins enter with smaller initial capital on average, with the opposite result holding if $\sigma < \frac{1}{2}$.[18] To see this, observe that substituting from (5) in (3), the capital size of the average entrant satisfies:

$$\log K_{it}^{a} = W + \frac{1 - 2\sigma}{2(1-\alpha)(1-\sigma)} \log A_{it} \tag{10}$$

where $W \equiv \log \phi + \frac{1}{2} + \frac{1}{2(1-\sigma)} \log \frac{1}{\psi} - \frac{2-\alpha-2\sigma}{2(1-\alpha)(1-\sigma)} \log r$. All firms face the same cost of capital and there are no mark-ups in our model. The preponderance of small and seemingly unproductive firms often noted in developing countries, which is typically attributed to wedges in factor prices and mark-ups in output price in the misallocation literature, may instead just be a manifestation of strong network effects! Our model implies that their own productivity understates their contribution via spillovers to their network.

In contrast to the results for initial capital, post-entry growth rates of firm size for any given cohort can be shown to be rising in p and over time. Equation (3) implies that the capital at date $t' > t$ of a cohort t entrepreneur with ability ω is given by

$$\log K_{itt'}^{a} = \log \omega + \log \phi - \frac{1}{1-\alpha} \log r + \frac{1}{1-\alpha} [\log A_{it} + \theta(p) \sum_{l=t}^{t'-1} e_{il}] \tag{11}$$

implying a growth rate at period t':

$$\log K_{it,t'}^{a} - \log K_{it,t'-1}^{a} = \frac{1}{1-\alpha} \theta(p) e_{it'} \tag{12}$$

In our model, growth in incumbent firm size is independent of the entrepreneur's ability and cohort and is driven entirely by contemporaneous changes in CTFP.

Proposition 3 *With two destinations:*

(a) Averaging across destinations, ability and initial capital of marginal entrants (also of average entrants if $\sigma > \frac{1}{2}$) is decreasing in t (for any given p) and in p (for any given t), and decreasing more steeply in p across successive cohorts.

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

From (5) and (9), the marginal entrant's ability and initial capital are decreasing in $\log A_{it}$. From (10), the average entrant's ability and initial capital are also decreasing in $\log A_{it}$ if $\sigma > \frac{1}{2}$. $\log A_{it} \equiv \log A_0 + \theta(p) N_{t-1} s_{i,t-1}$ is increasing in N_{t-1} when it is averaged across destinations. From Propositions 1 and 2 we

[17] This result does not depend on assumptions concerning the distribution of ability. To see this, observe that expressions (3, 4) show that capital size and entrepreneurial profit depend on individual ability and CTFP in exactly the same way. The marginal entrepreneur must be of lower ability when CTFP is higher, and must be indifferent between the traditional occupation and entrepreneurship. Profits will thus be lower in the traditional occupation for an agent with lower ability. So the same is true for entrepreneurial profit, and hence for capital size.

[18] This depends on the assumption of a log uniform distribution of ability.

know that E_t and, hence, N_t is increasing in t (for any p), increasing in p (for any t), and increasing more steeply in p over time. Hence, part (a) of Proposition 3 follows immediately. A similar argument can be used to establish part (b). Averaging across destinations, $e_{it'}$ is replaced by $E_{t'}$ on the right hand side of equation (12). The result then follows from Propositions 1 and 2. Firms from high-p origins start smaller, but subsequently grow faster.[19]

3.5 Alternative Explanations

To what extent do the preceding results rely on network spillovers? Could they be obtained, instead, by relaxing different assumptions of our model, while shutting down the network component? These questions are relevant because although population density may be positively associated with trust in the birth county and, hence, with network quality, it could also be correlated with other factors that independently determine the dynamics of entry, concentration, and firm size. The discussion that follows systematically examines this possibility by introducing new sources of (possibly time-varying) heterogeneity at the origin, which are, in turn, correlated with population density, and by allowing firms from different origins to have favorable access to destinations of varying quality. Our model treats sectors and locations interchangeably. Because locational heterogeneity is an important alternative that we must consider, entrepreneurs choose between locations (which we refer to as destinations for expositional convenience) rather than sectors in the alternative models that are examined below.

3.5.1 Origin Heterogeneity

Our model assumes cohort size and the share of potential entrepreneurs, k, are constant across origin counties and cohorts. Suppose that we relax these assumptions and let $k(p,t)$, which now refers to the number of potential entrepreneurs, be a twice differentiable function satisfying $k_p > 0, k_t > 0, k_{pt} > 0$. This could be because higher population density counties simply have larger populations that are growing relatively fast over time or because their residents have greater wealth or preferred access to finance, which facilitate entry into business. An additional source of origin heterogeneity could be in payoffs in the traditional occupation across counties. Our model assumes that the payoff, ω^σ, where ω is individual ability, is the same in all counties and constant over time. However, the payoff could be lower in higher population density counties because there is a larger population for a given amount of resources (such as agricultural land). It is also possible that this population pressure is increasing over time. We allow for this possibility by representing the payoff in the traditional sector by $\omega^\sigma v(p,t)$, where $v(p,t)$ is a twice differentiable function satisfying $v_p < 0, v_t < 0, v_{pt} < 0$.[20]

Abstract for the time being from heterogeneity across destinations, so the TFP of any entrepreneur with ability ω at any destination at time t is $\omega^{1-\alpha}A_t$, with A_t growing exogenously over time. Moreover, an

[19] In a related paper, Banerjee and Munshi (2004) find that outsiders in Tirupur's garment cluster, who face a higher cost of capital because they have weaker local credit networks, start with smaller firms and then grow faster (because they are positively selected on ability). To explain Banerjee and Munshi's findings, our model would need to be augmented to allow firm growth to be increasing in the entrepreneur's ability, with the additional condition that the ability effect needs to dominate the network effect (which is stronger for the insiders).

[20] An alternative interpretation of $v(p,t)$ is that it represents the payoff from origin-based networks operating in the traditional sector.

exogenous share s_i of potential entrepreneurs at each origin have the opportunity to enter any given destination. Owing to the absence of network effects, neither A_t nor s_i depend on p.

The ability threshold for entry into destination i from an origin with population density p in this alternative model would equal

$$\log \underline{\omega}_i(p;t) = \frac{1}{1-\sigma}\left[\log\frac{1}{\psi} + \frac{\alpha}{1-\alpha}\log r - \frac{1}{1-\alpha}\log A_t + \log v(p,t)\right] \tag{13}$$

while the expression for entry flows is:

$$e_i(p,t) = s_i k(p,t)\left[1 + Z + \frac{1}{(1-\alpha)(1-\sigma)}\log A_t - \frac{1}{1-\sigma}\log v(p,t)\right] \tag{14}$$

where $Z \equiv \frac{1}{1-\sigma}\log\psi + \frac{\alpha}{(1-\alpha)(1-\sigma)}\log r$. The entry flows will be rising in p and in t, and the slope with respect to p will be rising in t. The alternative model can thus generate our model's predictions for entry. However, the share of different destinations will be constant and independent of p. In order to obtain the same predictions for spatial concentration generated by the network model, the shares s_i of different destinations would have to (exogenously) depend on p and t in a way that exactly delivers these results. Although we do not explicitly incorporate sectors in the alternative model, it would similarly need to be augmented to exactly match our model's predictions for the dynamics of sectoral concentration.

With $\sigma \in (\frac{1}{2}, 1)$ the initial capital of entrants would fall over time due to the increase in A_t and the decline in $v(p,t)$, and would also be falling in p (more steeply over time) due to the $v(p,t)$ term. However, post-entry growth of firm size would be driven entirely by rising productivity at the destinations, A_t, which does not vary with the origins of entrepreneurs. Hence this model would not generate result (b) of Proposition 3 concerning post-entry growth in firm size across origin counties.

3.5.2 Destination Heterogeneity

Now consider the implications of varying productivity levels and growth rates across destinations. This could reflect the effect of geography, support provided by local governments (through credit and infrastructure), or agglomeration spillovers. The latter depend on the total number of firms at a destination, regardless of their origin. Let A_{it} denote productivity at destination i at t, which does not vary with the origins of entrepreneurs in the absence of network effects. Suppose in addition that high p origins have better, and increasing, access to the faster growing destinations. For instance, if there are two destinations and productivity at destination 1 is higher and growing faster than at destination 2, then the share $s_1(p,t)$ is increasing in t and in p (more steeply over time). The expressions for entry thresholds and entry flows are now

$$\log \underline{\omega}_i(p;t) = \frac{1}{1-\sigma}\left[\log\frac{1}{\psi} + \frac{\alpha}{1-\alpha}\log r - \frac{1}{1-\alpha}\log A_{it} + \log v(p,t)\right] \tag{15}$$

$$e_i(p,t) = s_i(p,t) k(p,t)\left[1 + Z + \frac{1}{(1-\alpha)(1-\sigma)}\log A_{it} - \frac{1}{1-\sigma}\log v(p,t)\right] \tag{16}$$

This model which incorporates both origin and destination heterogeneity would generate the same predictions as Proposition 2 for entry and concentration. There would be greater total entry from high p origins owing to the origin heterogeneity, coupled with greater access to the faster growing destination. Concentration would rise over time for entrepreneurs from all origins, owing to faster entry growth into destination 1. This would be more pronounced for the high p origins, so concentration would rise in p and $p * t$. Entry thresholds from high-p origins would be lower due to higher A_{it} (averaged across destinations) or lower $v(p, t)$, so the initial capital size result in part (a) of Proposition 3 would also go through. The average rate of growth of firm size (where we average across destinations) would be higher for high-p origins, owing to their preferred access to the faster growing destination.

The alternative model specified above can generate the predictions of our model relating to the dynamics of entry, concentration, and firm size because the key $s_i(p, t)$, A_{it} terms are exogenously specified to match the endogenous evolution of these terms in our model. If firms from each origin locate at a unique set of destinations, then our network-based model would not be distinguishable from the alternative model with destination heterogeneity. In practice, however, firms from multiple origins will locate at the same destination. Destination-time period dummies can then be included in the estimating equation. Conditional on these dummies, the network model would imply that firms from higher-p origins will grow faster on average because growth is determined by changes in CTFP. In contrast, there is no relationship between firm growth and p in the alternative model once destination-time period dummies are included because there is no longer any variation within destinations.

One way to incorporate heterogeneity within destinations, without networks, would be to allow firm growth to vary with the entrepreneur's ability (this is not a feature of our model). A positive relationship between p in the origin county and firm growth would then be obtained even within destination-time periods if entrepreneurs from higher p origins have higher ability on average. However, this model would not explain why firms from higher p origins, with higher ability, nevertheless have lower initial capital. An alternative model that may be considered, is that entrepreneurs do not have access to external credit and have to be entirely self-financing (Song *et al.*, 2011). Suppose that for some reason entrepreneurs from high p counties have a higher shadow cost of capital, so entering firms start with lower capital size, and thereafter grow faster owing to convergence forces akin to those in the Ramsey-Solow neoclassical growth model. This model would not be able to explain the positive relationship between population density and either entry or concentration; high p origins ought then to be associated with smaller entry flows. Nor would it be able to explain why the positive relationship between firm size growth and population density is robust to controlling for initial capital size (as shown below).

4 Testing the Model

4.1 Evidence on Firm Entry

The model predicts that firm entry is (i) increasing in origin county population density at each point in time, (ii) increasing over time, and (iii) increasing more steeply in population density over time. This is a statement about the flow of firms rather than the stock. We test these predictions with data over the 1990-2009 period because, as documented below, the decline in the ability threshold predicted by the model starts to weaken beyond that point.[21] Our analysis covers the universe of registered firms, in contrast with previous analyses

[21] Recall that the model only applies to the initial, rapid growth, phase of industrialization.

of firms in China; e.g. Hsieh and Klenow (2009), Song *et al.* (2011), Brandt *et al.* (2012), Aghion *et al.* (2015) which 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. The above-scale firms account for less than 15% of all private firms in the registration database in 2008. Nevertheless, with nearly 2,000 counties, the number of entering firms from a given birth county at each point in time is relatively small. Each time period in the dynamic analysis that follows thus covers a five-year window.

Although the model assumes that there is a single entrepreneur in each firm, in practice most registered firms consist of multiple shareholders. The SAIC database lists the major shareholders and managers, with their citizenship ID, in each registered private firm. The first six digits of the citizenship ID reveal the birth county of the individual.[22] We designate the firm's legal representative as the "entrepreneur" for the purpose of the empirical analysis and his birth county thus applies to the firm as a whole. This individual is legally responsible for the firm's liabilities and typically plays a key role in its functioning; for example, 75% of legal representatives are shareholders in their firms. Given the high degree of clustering by birth county within firms, our choice of the designated entrepreneur has little bearing on the analysis in any case. The legal representative and the largest shareholder belong to the same birth county in over 90% of firms. Even among the 58% of firms that are established outside the legal representative's birth county, as many as 74% of the listed individuals belong to his birth county.[23] This is substantially higher than the statistic that would be obtained by random assignment of listed individuals in the firm's sector-location, which is just 6%, highlighting the role played by the birth county in supporting business.

Figure 5 reports nonparametric estimates of the relationship between the entry of firms from each birth county in each time period and 1982 population density.[24] The entry patterns in the figure are visually consistent with the model's predictions.[25] Table 4, Columns 1-4 report parametric estimates corresponding to Figure 5, separately by time period. This allows us to statistically validate the prediction that entry is increasing in birth county population density at each point in time. As noted, all analyses that estimate the direct effect of birth county population density will include population, education, and occupational structure (also measured at the county level in 1982) in the estimating equation. This allows for the possibility that population density is correlated with county characteristics that independently determine the outcomes of interest. We see in Table 4, Columns 1-4 that the population density coefficient is positive and significant in

[22] 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.

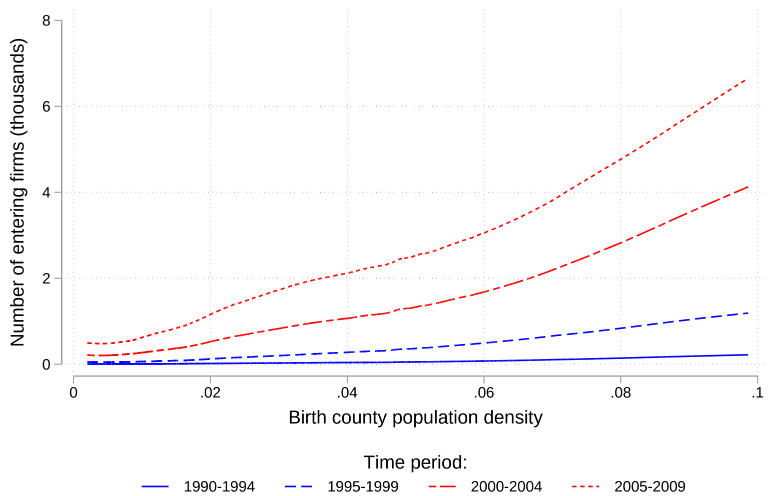
[23] Among the county-born legal representatives, 42% establish their firm in their birth county, 11% in their birth prefecture but outside the birth county, 18% in their birth province but outside the birth prefecture, and 29% outside their birth province.

[24] The advantage of measuring population density in 1982, prior to privatization, is that it avoids the possibility that changes in population density in later time periods are generated by endogenously determined network-based migration. Nevertheless, as documented in Appendix Figure D.1 using successive rounds of the population census, 1982 population density is highly correlated with population density in later time periods.

[25] Appendix Figure D.2a reports the corresponding nonparametric relationship between population density in the birth county and the stock of firms (measured at the end of each time period). The predictions of the model apply to both firm entry; i.e. the flow and the stock of firms. In practice, however, the stock will also take account of exits, which play no role in the model. We see in Figure D.2a that the model's predictions for the stock of firms go through as well, despite the exits. As an additional robustness test, Appendix Figure D.2b reports the nonparametric relationship between population density in the birth county and firm entry, restricting attention to firms that locate outside the birth county. Although the entry result is based on all locations, we see that the predictions of the model hold up with this reduced sample of locations as well.

each time period. Notice also that the mean of the dependent variable and the population density coefficient are increasing across time periods, in line with predictions (ii) and (iii) above. Formal tests of these predictions are reported later in this section.

Figure 5. Firm Entry and Population Density



Source: SAIC registration database and 1982 population census.

Table 4. Firm Entry and Population Density

Dependent variable:	number of entering firms				number of entering firms			
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	0.013***	0.092***	0.289***	0.448***	0.014**	0.118***	0.382***	0.575***
	(0.003)	(0.013)	(0.035)	(0.052)	(0.006)	(0.037)	(0.101)	(0.126)
Mean of dependent variable	0.0306	0.208	0.787	1.560	0.0725	0.483	1.673	3.024
Sector fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Location fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Observations	1,624	1,624	1,624	1,624	1,085,169	1,085,169	1,085,169	1,085,169

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.
In Columns 5-8, for a given birth county, all sectors and locations 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 locations across birth counties, the number of entrants is multiplied by the number of sectors $\times$ the number of locations.
Control variables include population, education and occupation distribution in the birth county.
Population is measured in millions and education is measured by the percent of the population that is literate.
Occupation distribution is measured as the share of workers in the birth county in agriculture and industry. Service is the excluded category.
Number of firms is obtained from the SAIC registration database and population density, population, education and the occupation distribution are derived from the 1982 population census.
Robust standard errors are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

As discussed above, both origin heterogeneity and destination heterogeneity can explain the entry results without requiring origin-based networks to be active. We account for important elements of origin heterogeneity by including additional birth county characteristics in the estimating equation. We next account for the possibility that entrepreneurs born in higher population density birth counties have access to faster growing destinations. Given that firms from multiple origin counties will enter each destination, we can flexibly accommodate this possibility by including destination fixed effects in the estimating equation. The estimating equation in Table 4, Columns 5-8 includes sector fixed effects and location fixed effects together with the birth county characteristics. This equation is estimated separately in each time period and so the fixed effects capture the changing fortunes of sectors and locations over time.[26] Although our analysis focuses on county-born businessmen, we place no restriction on the location of their firms; there are 3,235 counties or urban districts where firms locate in our data. Birth county population density continues to have a positive and significant effect on entry in each time period in Table 4, Columns 5-8. A comparison of the results obtained with the benchmark specification in Columns 1-4 and the augmented specification in Columns 5-8 indicates that the inclusion of the destination effects actually increases the point estimates. This tells us that entrepreneurs born in high population density counties are selecting sectors and locations that are less advantageous (receiving fewer entrants overall).

4.2 Evidence on Concentration

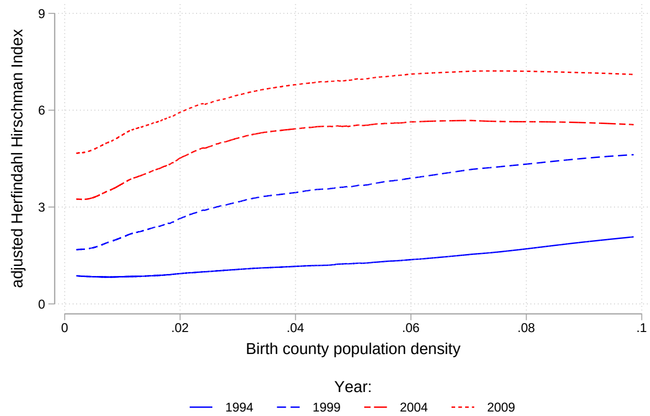
The model predicts that the concentration of the stock of firms, measured by the Herfindahl Hirschman Index (HHI) across destinations, is (i) increasing in birth county population density at each point in time, (ii) increasing over time, and (iii) increasing more steeply in population density over time. Destinations are defined by sectors or by locations within sectors. Figure 6a 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 intervals from 1994 to 2009. The HHI is based on the stock of existing firms (net of exits) and 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 derived in Appendix C. The adjusted HHI is evidently increasing in population density at each point in time and increasing over time, although it is difficult to visually assess whether the slope of the relationship gets steeper over time. Figure 6b reports nonparametric estimates of the relationship between spatial concentration, within one-digit sectors, and birth county population density in five-year intervals.[27] Although the model assumes that all destinations are symmetric, one obvious asymmetry in practice is that transportation costs are lower when the entrepreneur chooses to stay back home. We avoid this asymmetry by including all locations in the analysis of spatial concentration, measured at the county or urban district level, except for the birth county. As with the analysis of sectoral concentration, the spatial concentration within each sector for a given birth county is based on the stock of firms (net of exits) and is adjusted for the number of firms and the number of external destinations, which would generate variation in the measured HHI just by chance. Matching the predictions of the model, the spatial HHI is evidently (i) increasing in birth county population density in each time period, (ii) increasing over time, and (iii) increasing more steeply over time.

Table 5 reports parametric estimates corresponding to Figure 6a and Figure 6b. The usual birth county characteristics are included in the estimating equation. Sector fixed effects are also included in the estimating

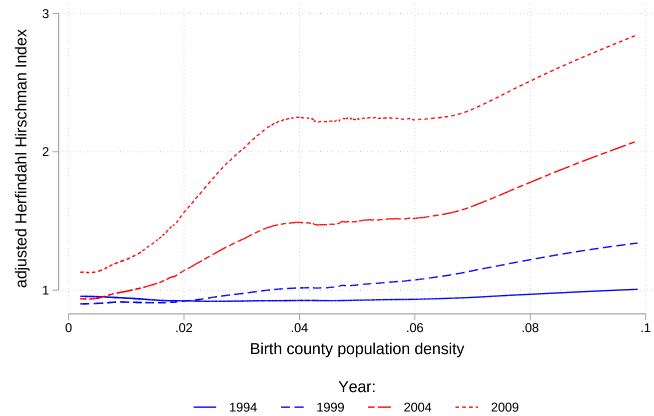
[26] Entry in Table 4, Columns 5-8 is measured at the birth county-sector-location level in each time period. The number of entrants is thus multiplied by the county-specific product of the number of sectors and the number of locations so that the dependent variable reflects entry at the level of the county.

[27] We measure spatial concentration within one-digit rather than two-digit sectors to allow for a sufficient flow of firms across locations. 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 there are multiple entrants from each birth county in each time period.

Figure 6. Concentration and Population Density



(a) Sectoral Concentration



(b) Spatial Concentration

Source: SAIC registration database and 1982 population census.
Sectoral concentration is 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 stock of firms and the number of sectors at each point in time.
Spatial concentration, within one-digit sectors, is measured by the Herfindahl Hirschman Index (HHI) across destination locations (outside the birth county) divided by the expected HHI that would be obtained by random assignment, given the stock of firms and number of destination locations at each point in time.

equation with spatial concentration (within sectors) as the dependent variable to allow for the possibility that concentration varies independently across sectors (possibly due to the nature of the production technology and the associated gains from agglomeration). Population density in the birth county has a positive and significant effect on (adjusted) sectoral and spatial concentration at each point in time. The mean of the dependent variable and the population density coefficient are increasing over time, in line with predictions (ii) and (iii), which we test formally below.

Table 5. Sectoral and Spatial Concentration and Population Density

Dependent variable:	adjusted HHI across sectors				adjusted HHI across locations			
Year:	1994	1999	2004	2009	1994	1999	2004	2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth county population density	0.106***	0.417***	0.444***	0.574***	0.014*	0.041***	0.047*	0.072*
	(0.023)	(0.057)	(0.061)	(0.074)	(0.008)	(0.014)	(0.027)	(0.040)
Mean of dependent variable	1.039	2.839	4.622	6.065	0.936	1.010	1.295	1.777
Observations	1,622	1,624	1,624	1,624	5,450	15,076	23,727	26,769

Note: sectoral concentration measured across two-digit sectors and spatial concentration, within one-digit sectors, is measured across destination locations (outside the birth county). Concentration statistics are adjusted for expected concentration due to random assignment.
Population density is measured in units of 10,000 people per square km, and then converted to Z-score.
Sector fixed effects included in the regression with spatial HHI as the dependent variable.
Control variables include population, education and occupation distribution in the birth county.
Number of firms is obtained from the SAIC registration database and population density, population, education and the occupation distribution are derived from the 1982 population census.
Robust standard errors are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Table 6 formally tests the model's predictions for changes in firm entry, sectoral concentration, and spatial concentration over time and across birth county population density 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. Since the cross-sectional relationship with population density in each time period has been previously reported with each outcome, we only report the coefficient on the time period variable and the interaction coefficient. 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 cities in Table 6, Columns 4-6.[28] There is no association between localized trust and population density in cities and thus we do not expect to find support for the model's predictions with this set of entrepreneurs. The time period coefficient and the interaction coefficient are both positive and significant with entry as the dependent variable but, as discussed, many alternative models can generate this result without a role for community networks. The model's predictions for concentration are less easy to explain away. Reassuringly, the interaction coefficient for the city-born entrepreneurs is *negative* and significant with sectoral concentration as the dependent variable and statistically indistinguishable from zero (at conventional levels) with spatial concentration as the dependent variable, contrary to the predictions of our model.

Table 6. Entry, Concentration, and Population Density (time and interaction effects)

Birth place:	county			city		
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***	1.686***	0.506***	0.661***	0.299***	2.054***
	(0.016)	(0.020)	(0.021)	(0.026)	(0.008)	(0.069)
Birth place population density × time period	0.353***	0.165***	0.134***	0.353***	-0.026***	0.030
	(0.029)	(0.019)	(0.012)	(0.041)	(0.004)	(0.026)
Observations	6,496	6,494	71,022	3,284	3,283	21,046

Note: the estimating equation includes, in addition, birthplace population density and a constant term.
Number of entering firms from each birth place 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 destination locations (outside the birth county) divided by the expected HHI that would be obtained by random assignment. Sector fixed effects are included in this specification.
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%.

4.3 Evidence on Firm Size

The model predicts that the ability and the initial capital of the marginal entrant is (i) decreasing in birth county population density at each point in time, (ii) decreasing over time, and (iii) decreasing more steeply in population density 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

[28] We measure urban population density at the city rather than the dissaggregated urban district level because urban districts were created after 1982. However, firm location is always measured at the urban district level.

apply to average initial capital as well. However, only the positive network productivity effects are relevant for post-entry growth rates of firm size.

To test the model, we measure marginal ability and initial capital at the level of the birth county-sector or birth county-sector-location in each time period and then estimate the (average) effect of birth county population density and its interaction with time on these variables. We begin in Figure 7a by nonparametrically estimating the relationship between a measure of ability, based on education, of the marginal entrepreneur in each birth county-sector-time period and population density in the birth county. It is standard practice to proxy for ability with education, and recent evidence from the U.S. indicates that education is also a good measure of entrepreneurial ability (Levine and Rubinstein, 2017). In a developing economy, however, 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. Our measure of ability is thus the entrepreneur's percentile rank in his birth county-birth cohort education distribution.[29] The marginal entrant is the entrepreneur who is placed at the bottom one percentile of the ability distribution among entering entrepreneurs in each birth county-sector-time period. We see in Figure 7a that the marginal entrant's measured ability declines over time; from around the 70^{th} percentile of his birth county-birth cohort education distribution in the 1990-1994 period to just around the 40^{th} percentile in the 2005-2009 period. The relationship between the marginal entrant's ability and population density is also negative in each time period and grows steeper over time.[30] Notice, however, that there is a bottoming out by the last, 2005-2009, period. Our model is only designed to capture firm dynamics up to this point, which is why the empirical analysis does not extend beyond 2009. For the dynamic analysis of negative selection that follows, and for the structural estimation, the analysis period will be restricted even further to the 1990-2004 period.

Figure 7b reports complementary 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.[31] As predicted by the model, marginal initial capital is decreasing over time and decreasing in birth county population density in each time period.[32]

Notice from Figure 7b that the decline in initial capital with birth county population density does not grow steeper over time (as implied by the model). One reason why this might be the case is because marginal initial capital within birth county-sector-time periods is effectively averaged across sectors in the figure. Although this is not a feature of our model, the capital requirement will vary across sectors, and this must be accounted for

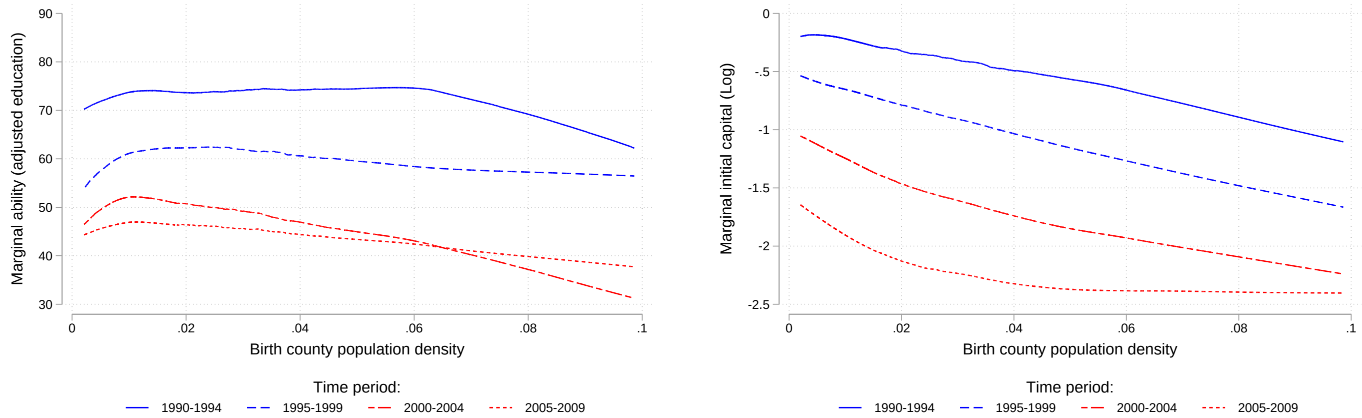
[29]The education distribution is constructed in each county for birth cohorts from 1920 to 1989 in five-year intervals, based on data from the 2000 population census. Each entrepreneur is assigned to a birth cohort interval based on his birth year, which is available from the registration database, 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 has no bearing on the complementary analysis of firm size, which includes all registered firms.

[30]Appendix Table E.1 reports parametric estimates corresponding to Figure 7a, separately in each time period. These estimates indicate that birth county population density has a negative and significant effect on marginal ability among entering entrepreneurs at each point in time.

[31]The initial capital for a firm is determined by its initial registered capital, which can be recovered from the SAIC registration database. The initial registered capital represents the total amount paid up by the shareholders. This amount is deposited with the SAIC 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.

[32]Appendix Table E.1 reports parametric estimates corresponding to Figure 7b, separately by time period. The population density coefficient is negative and significant in each time period.

Figure 7. Marginal Ability, Marginal Initial Capital and Population Density



(a) Marginal Ability **(b)** Marginal Initial Capital

Source: SAIC registration database and 1982 population census.
The entrepreneur's ability is measured by his percentile rank in his birth county- birth cohort education distribution (obtained from the 2000 population census). The marginal entrepreneur is defined as the individual at the bottom one percentile of the ability distribution among entering entrepreneurs in a given birth county-sector-time period.
Marginal initial capital is defined as the bottom one percentile of the initial capital distribution at the birth county-sector-time period level.

in the empirical analysis. Table 7 allows for this by studying the change in the ability of entering entrepreneurs and their capital investments over time, within birth county-sectors. The analysis is restricted to the 1990-2004 period because our measure of marginal ability and initial capital both bottom out (and flatten out) in the 2005-2009 period in Figures 7a and 7b. We see in Table 7, 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 in our measure of ability is significantly steeper for entrants from higher population density counties, as predicted by the model. Table 7, Columns 2-3 use the distribution of initial capital (in logs) in each entering cohort of firms, in five-year windows over the 1990-2004 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. Although 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 location choices, and the possibility that variation in these choices across birth counties could be driving the results. Table 7, Columns 4-5 thus includes location fixed effects, in addition to birth county-sector fixed effects in the estimating equation. Initial capital is now measured at the birth county-sector-location level in each time period.[33] 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

[33]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-location-time period. The sample in Columns 4-5 is restricted to birth county-sector-locations with entrants in the initial period. Similarly, the sample in Columns 2-3 is restricted to birth county-sectors with entrants in the initial period.

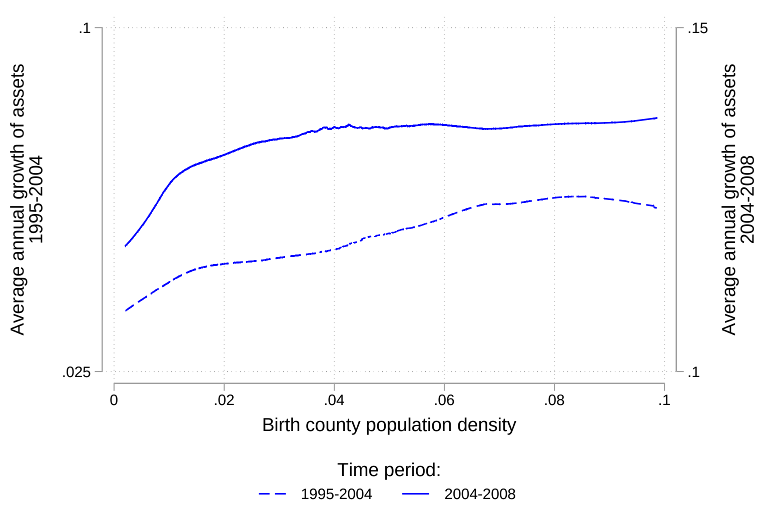
variables, as predicted by the model. As with the analysis of firm entry, accounting for location effects only strengthens our results.

Table 7. Evidence on Negative Selection

Dependent variable:	marginal ability	marginal initial capital	average initial capital	marginal initial capital	average initial capital
	(1)	(2)	(3)	(4)	(5)
Time period	-17.908***	-0.868***	-0.116***	-0.609***	-0.095***
	(0.496)	(0.012)	(0.009)	(0.010)	(0.008)
Birth county population density × Time period	-0.926***	-0.026**	0.002	-0.061***	-0.020***
	(0.351)	(0.011)	(0.007)	(0.009)	(0.006)
Mean of dependent variable	49.36	-1.744	-0.401	-1.223	-0.374
Origin-sector fixed effects	Yes	Yes	Yes	Yes	Yes
Location fixed effects	No	No	No	Yes	Yes
Observations	21,028	43,578	43,578	46,417	46,417

Note: The entrepreneur's ability is measured by his percentile rank in his birth county- birth cohort education distribution (obtained from the 2000 population census). The marginal entrepreneur is defined as the individual at the bottom one percentile of the ability distribution among entering entrepreneurs in a given birth county-sector-time period.
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 or the birth county-sector-location-time period level (when location fixed effects are included). Average initial capital is the mean of the distribution.
Time period is an ordinal variable taking value from 1 to 3 corresponding to successive five-year time windows over the 1990-2004 period. Education and initial capital are obtained from the SAIC registration database and birth county 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%.

Figure 8. Asset Growth and Population Density



Source: Industrial census (1995, 2004, 2008) and 1982 population census.
Firm-level average annual growth of assets is averaged up to the birth county-sector level in each time period.

Conditional on having entered, the model predicts that firms from higher population density counties will grow faster. Although the registration database is well suited to examine entry, concentration, and initial capital investments, it is less suitable for analyses of capital growth. Registered capital does change, but given that these changes are self-reported and involve substantial administrative costs, it 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 and which has reasonable coverage from 2004 onwards. We thus compute the average annual growth rate over the 1995-2004 and 2004-2008 periods with the industrial census and, to be consistent, over the 2004-2008 period with the inspection data.[34] Figure 8 reports asset growth, separately in the 1995-2004 period and the 2004-2008 period, based on the industrial census. The average annual growth of assets is increasing in population density in each time period and increasing over time, as predicted by the model, in contrast with the patterns that we observe in the data for initial firm size.

Table 8. Growth of Assets and Population Density

Dependent variable:	average annual growth of assets					
Source:	industrial census		industrial census		inspection data	
Time period:	1995-2004		2004-2008		2004-2008	
	(1)	(2)	(3)	(4)	(5)	(6)
Birth county population density	0.006***	0.007*	0.004**	0.003**	0.004***	0.002*
	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)
Initial capital	–	0.002***	–	0.001***	–	0.001***
		(0.000)		(0.000)		(0.000)
Mean of dependent variable	0.0528	0.0557	0.133	0.136	0.106	0.110
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Location fixed effect	No	Yes	No	Yes	No	Yes
Observations	5,517	5,664	31,234	64,258	18,701	43,470

Note: firm-level average annual growth of assets is averaged up to the birth county-sector level in specifications with sector fixed effects and to the birth county-sector-location level in specifications with sector fixed effects and location fixed effects.
Initial capital (in million Yuan) obtained from the SAIC registration database and birth county 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 8 reports parametric estimates corresponding to Figure 8. Columns 1-4 report results with industrial census data and Columns 5-6 repeat the analysis with SAIC inspection data, which include all sectors (not just manufacturing, as in the industrial census).[35] Since growth rates can only be computed at two points in time with the industrial census data and the inspection data cover a relatively short period of time, we focus on the cross-sectional predictions of the model. To test the model's predictions, we measure firm growth at the birth county-sector level in Table 8, Columns 1, 3 and 5 and at the birth county-sector-location level in Table 8, Columns 2, 4 and 6 and then estimate the (average) effect of birth county population density on these growth measures. The estimating equation in Columns 1, 3 and 5 includes sector fixed effects, while the estimating equation in Columns 2, 4 and 6 includes sector fixed effects, location fixed effects, and the firm's initial capital. The fixed effects account for exogenous variation in firm growth across sectors and locations, which is not a feature of our model. The firm's initial capital is included to allow for convergence; recall that one alternative explanation for why firms from high population density birth counties start small and then

[34]The average annual growth between period t and t' is computed as the difference in log assets in t' and t divided by $t' - t$. Although there are no exits in the model, this is a feature of the data. In practice, firms with low profit levels – the young and the less able – are more likely to exit. This selective exit, based on the profit level, does not bias our estimates because growth rates in the model are determined entirely by network quality and CTFP, which apply equally to all active firms from a given birth county at a given point in time.

[35]Data coverage for seven provinces is poor with the inspection data and these provinces are thus dropped from the analysis.

grow faster is mechanical convergence (with initial size being accidentally determined). What we observe, instead, is that firms that are larger to begin with, subsequently grow faster. Moreover, the consistent finding across specifications is that population density in the birth county has a positive and significant effect on firm growth.[36] Firms from high population density birth counties start small but subsequently grow faster, after accounting for sector and location effects. As discussed, this result is especially useful in distinguishing our model from alternative non-network explanations.

4.4 The Mechanism

The key relationship underlying the model is that greater network presence in a particular destination is associated with higher entry into that destination. This is on account of both the increased firm productivity at that destination and the referral effect. While the results on entry and concentration reported above are consistent with this relationship, we now proceed to test it directly (being mindful of potential omitted variable bias as discussed below) by estimating the effect of initial entry on subsequent entry within birth county-sector-locations. Our model implies that initial entry will affect entry in the next period and, via a compounding network multiplier effect, entry in all subsequent periods. Initial entry is measured in the 1990-1994 period, when private firms were first starting to emerge in China, and subsequent entry is measured separately in 2000-2004 and 2005-2009. The benchmark specification in Table 9, Columns 1-2 includes, in addition, birth county-sector fixed effects and the total number of initial entrants in the sector-location from all origins to capture generalized location-based agglomeration.[37] We see that initial entry from the birth county has a positive and significant effect on subsequent entry, with this effect growing stronger over time (although not significantly), which is consistent with a network multiplier effect.

Conditional on the number of initial entrants from the birth county, the total number of initial entrants in a given sector-location has no effect on subsequent entry from that birth county in that sector-location. This result provides empirical support for the assumption in the model that the birth county is the domain within business networks are organized in China, rather than destination sectors and locations, and that these networks operate independently. The absence of a negative cross-community effect also provides support for the assumption that individual networks cannot influence the price of the product. If outputs produced by any single network had an impact on the market price for the product, we would expect to see a negative cross-network effect (since greater entry from any network would depress the market price, and thus entry from other networks).[38] While pricing may be non-competitive in China (see, for example, Brooks *et al.* (2016)), the origin-based networks do not appear to be directly associated with these distortions.

Although we assume that network spillovers increase the productivity of their members, an alternative explanation for clustering is that the networks capture political rents. Bai *et al.* (2019), for example, describe how favored firms have superior access to capital allocated by local governments. If local government officials

[36] A pooled regression (not reported) which combines industrial census data over both time periods indicates, in addition, that firm growth is increasing significantly over time. While our model can explain this result, it must be augmented, perhaps by introducing credit constraints, to explain why initial capital has a positive and significant effect on firm growth.

[37] We are interested in estimating the marginal effect of initial entry on subsequent entry and, hence, birth county-sector-locations with zero initial entry can also be included in the analysis. All locations which had a positive number of entrants by 2000-2004 and 2005-2009, respectively, for a given birth county-sector are consequently included in the estimation sample.

[38] Appendix Table E.2 replicates Table 9 restricting attention to locations outside the birth county. Although the coefficient on the total number of initial entrants is now statistically significant, it remains positive and an order of magnitude smaller than the coefficient on the number of entrants from the birth county.

Table 9. The Effect of Initial Entry on Subsequent Entry

Dependent variable:	subsequent entrants from the birth county			
Time period:	2000-2004	2005-2009	2000-2004	2005-2009
	(1)	(2)	(3)	(4)
Initial entrants from the birth county	7.120***	8.935***	5.198***	5.723***
	(0.686)	(0.956)	(0.972)	(1.281)
All initial entrants at the destination	0.054	-0.020	–	–
	(0.048)	(0.056)		
Initial entrants from the birth county × birth county population density	–	–	1.356**	2.277**
			(0.564)	(0.937)
Distance to the birth county	–	–	-2.225***	-2.962***
			(0.104)	(0.106)
Mean of dependent variable	3.156	3.170	3.156	3.170
Origin-sector fixed effects	Yes	Yes	Yes	Yes
Location fixed effects	No	No	Yes	Yes
Observations	384,031	778,897	384,031	778,897

Note: number of entrants is measured at the birth county- sector-destination level.
Initial entry is computed over the 1990-1994 period and sectors are defined at the two-digit level.
Number of entrants is obtained from the SAIC registration database and birth county population density is computed from the 1982 population census.
Standard errors clustered at birth county-sector level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

favor entrepreneurs from their hometown, then this would explain why firms from a given birth county cluster at the same location.[39] However, while pre-existing social ties might confer a temporary advantage resulting in additional entry, it will not persist because government officials are frequently rotated precisely to avoid such corruption. Based on data from city yearbooks and the online CV's of government officials, we estimate that the median (mean) term-length for a city mayor in China is 4 (3.7) years. Government favoritism based on pre-existing social ties, thus cannot explain the long-term effect of initial entry on subsequent entry that we observe in Table 9.

A related mechanism that might explain the inter-temporal correlation is based on the idea that stronger networks are more effective at lobbying the government official who is in place, regardless of his origin. A testable implication of network competition for scarce government credit, and resources more generally, is that greater entry from other origins into a particular location will have a negative effect on entry from a given origin into that location. As already noted, there is no evidence of such negative cross-community spillover effects.[40] While particular entrepreneurs may have favored access to government credit, the birth county networks do not appear to be facilitating this process. Consistent with this finding, recently available data from the Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC) which uses the registration database as the sampling frame for a subset of firms indicates that initial registered capital is obtained from the following sources: self finance (76%), the owners' social network (15%), and bank loans (9%). In contrast, and in line with the productivity enhancing role of the network that is assumed in the model, 55% (62%) of

[39] Fisman *et al.* (2018) document the same type of favoritism, based on hometown ties, in the Chinese Academy of Sciences.

[40] We have assumed thus far that the destination is defined by the two-digit sector and the (rural) county or urban district. It is possible that government patronage and associated network spillovers extend to the prefecture level. Appendix Table E.3 replicates Table 9, with destinations defined by the two-digit sector and prefecture. There continues to be no evidence of negative cross-community spillover effects; instead, the effects are positive and significant.

the firms report that their largest stable supplier (buyer) either belongs to their social network or was referred by a member of the network.[41]

In contrast with the tests of Propositions 1-3, our test of the network mechanism assumes that initial entry in the 1990-1994 period is exogenously determined. While there is very likely an accidental element to initial entry patterns, it is possible that initial entrants selected into persistently favorable or proximate destinations. A positive correlation between initial entry and subsequent entry could then be obtained without requiring networks to be active. We account for this possibility in Table 9, Columns 3-4 by adding location fixed effects and the distance between the birth county and the location under consideration in the estimating equation. Location fixed effects can be included because entrepreneurs from multiple origins will establish their firms in the same place, but the variable measuring initial entry from all origins will now be subsumed in the fixed effects. The additional controls will account for a wide class of omitted variables; what remains is the possibility that some factor other than distance ties birth counties to specific sector-locations. We address this possibility with a stronger test of the network mechanism, which is that the initial entry effect should be increasing in birth place population density. Table 9, Columns 3-4 reports estimation results with the augmented specification, where we see that the interaction coefficient is positive and significant, and increasing from 2000-2004 to 2005-2009, in line with the results for entry and concentration reported above.[42] This result could also be generated without networks if entrepreneurs from higher population density counties have preferred access to faster growing destinations, but as noted, this does not appear to be the case.

5 Structural Estimation

Having validated the model, we next proceed to estimate its structural parameters. This will allow us to quantify the contribution of the community networks to the growth in the number of firms and the capital stock at the aggregate level. For the network to have any effect on entry and initial capital size in the model, there must be positive lagged entry. The structural estimation and the quantification of network effects that follows is thus restricted to destination-time periods with positive lagged entry. To increase the fraction of firms that are included in these destination-time periods and, hence, in the structural analysis, we now define the destination by the one-digit sector and prefecture. The implicit assumption is that birth county networks operate at this level. Our results do not appear to be sensitive to the precise choice of the network domain in any case; based on the tests of the network mechanism that the estimated initial entry effect is very similar in magnitude when the destination location is defined by either county-urban district in Table 9 or by prefecture in Appendix Table E.3.[43]

Based on the entry and initial capital equations in the model, (6) and (10), and retaining its notation, we estimate the following structural equations simultaneously:

$$e_{ci,t} = G(\alpha, \sigma, r, A_0)k_c s_{ci,t-1} + \frac{\theta}{(1-\sigma)(1-\alpha)}k_c s_{ci,t-1} \cdot pn_{ci,t-1} + u_{ci,t} \tag{17}$$

[41] The social network is defined to include individuals with pre-existing social ties to the entrepreneur: relatives, *laoxiang* (homeplace fellows), classmates, friends, army comrades, colleagues, and past business partners. These categories are not exclusive; in particular, relatives, classmates, and many individuals from the other categories will be *laoxiang*.

[42] The same result is obtained with the alternative specifications reported in Appendix Tables E.2 and E.3.

[43] If the total number of initial entrants has a linear effect on subsequent entry at a given point in time and the domain of the network is defined by the prefecture, then a marginal increase in the number of entrants at the prefecture level will have the same effect as a marginal increase in the number of entrants in any county or urban district within the prefecture.

$$logK^a_{ci,t} = J_t(\alpha, \sigma, r, A_0, f_t) + \frac{\theta(1-2\sigma)}{2(1-\sigma)(1-\alpha)} pn_{ci,t-1} + v_{ci,t} \tag{18}$$

The functional forms for $G(\alpha, \sigma, r, A_0)$ and $J_t(\alpha, \sigma, r, A_0, f_t)$ are obtained directly from equations (6) and (10), with the addition of the separable f_t term (described below) in the J_t function. $e_{ci,t}$ measures the number of entrants and $logK^a_{ci,t}$ measures average initial capital (in logs) for birth county c and destination (sector-location) i in time period t. We parameterize the $\theta(p)$ function to be increasing linearly in p: $\theta(p) = \theta p$, with the restriction that $\theta(0) = 0$. This restriction is based on the idea that as the population density goes to zero in a hypothetical county, there can be no social interactions and, hence, no enforceable trust. The network effect is thus represented by a single parameter, θ. $n_{ci,t-1}$, which is the stock of firms from birth county c that are already established in destination i at the beginning of time period t, is treated as exogenously determined in the structural estimation. $s_{ci,t-1}$, which denotes the share of destination i in the stock of firms originating from county c at the beginning of period t is also treated as exogenous. 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 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 represents the price of capital goods in period t.

The theoretical model assumes that the size of each cohort and the fraction of potential entrepreneurs in the cohort, k, are the same in all counties. In practice, cohort size will depend on the county population and the fraction of potential entrepreneurs will depend on the level of education in the county. k_c in equation (17) thus refers to the number of potential entrepreneurs in county c. This number is calculated from the 1990 population census, based on the characteristics of actual entrepreneurs when they established their firms. Most entrepreneurs in the SAIC database have at least high school education and relatively few were younger than 25 when their firm was established. Assuming, as in the model, that individuals must make a one-time occupational choice at the start of their careers, k_c for each (five-year) cohort of entering entrepreneurs is thus specified to be the number of men born in county c, aged 25-29, with at least high school education, as reported in the 1990 population census.

The residual terms, $u_{ci,t}$ and $v_{ci,t}$, include the effect of local government inputs, agglomeration, and sector-level spillovers on access to capital, firm productivity, and accompanying entry. The structural equations are linear in observed variables; (i) $k_c s_{ci,t-1}$ (ii) $k_c s_{ci,t-1} \cdot pn_{ci,t-1}$ (iii) $pn_{ci,t-1}$ and, hence, we will be able to control flexibly for these potentially confounding effects, which could bias the structural estimates, as shown below. There are four reduced-form coefficients in equations (17) and (18). One of these coefficients, J_t, cannot be used to identify the structural parameters because f_t is unobserved. This leaves three reduced-form coefficients and five structural parameters: $\alpha, \sigma, r, A_0, \theta$. The model is under-identified and we thus assign values to r and A_0 prior to estimation.

We noted in Section 3 that the productivity channel and the credit channel for the network effect cannot be separately identified. Although the model is parameterized to allow networks to increase productivity, we remain agnostic about the specific channel through which the networks operate. For the structural estimation, we specify that the network operates through the productivity channel as in the model, setting r to 0.2 (which is in line with estimates of the average interest rate faced by Chinese firms).[44] The productivity multiplier is set to one in all destinations; i.e. $A_0 = 1$. As in the model, variation in productivity across destinations (and origin counties) is generated entirely by the network effect; $exp(\theta p \cdot n_{ci,t-1})$. In addition to the factors included above in the residual terms, we thus also abstract from variation in product prices and labor productivity. The objective will be to assess how well our parsimonious model is able to match the observed dynamics of entry and firm size.

To accommodate differences in the capital requirement across sectors, we do, however, allow the α parameter, which measures the marginal returns to capital, to vary across six broad sector categories: high-tech services, wholesale and retail services, manufacturing and transportation, heavy industry (mining, electricity, and construction), non-financial services (hotels, catering, education), and finance. This increases the number of structural equations to 12, given that there are now two equations in each sector category, and the number of structural parameters to be estimated to eight; $\alpha_1, ..., \alpha_6, \sigma, \theta$. The structural parameters are estimated by matching on entry and average initial capital in each birth county-destination-time period.[45] The model is estimated over the 1995-2004 period; i.e. over two time periods, matching the reduced form analysis with initial capital as the outcome. f_t, which adjusts capital from physical to monetary units in each time period, and which appears additively in the J_t function, is thus estimated separately in 1995-1999 (period 1) and 2000-2004 (period 2).

To estimate the structural parameters, we search for the set of parameters that minimize the distance between the actual and the predicted entry and average initial capital; i.e. for which the sum of squared errors over all birth county-destination-time periods is minimized. Parameter estimates, with bootstrapped standard errors in parentheses, are reported in Table 10, Column 1.[46] 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 estimated θ parameter is positive and highly significant. The remaining structural parameters (α_1, ..., α_6) and the capital adjustment term, f_t, which is estimated separately in each time period are not reported to preserve space.

Table 10, Column 2 reports estimates from an augmented specification that includes lagged entry in the sector-location from all origins (including urban origins). The additional $n_{i,t-1}$ term is included additively to capture generalized agglomeration effects and to proxy for destination effects, such as government infrastructure, that would induce entry from all origins. This term, like $n_{ci,t-1}$, is treated as exogenous in the structural estimation, and we allow for separate $n_{i,t-1}$ coefficients in the entry and initial capital equations. The augmented specification in Column 2 also adds time period effects to the entry equation to accommodate secular changes in the economy over time. The estimated σ and θ parameters are hardly affected by these additions to

[44] In our model, 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 Chinese industrial census, Brandt *et al.* (2016) estimate the real interest rate to be 0.15 in 1995 and 2004 and 0.18 in 2008. We thus set r to 0.2.

[45] Although the number of reduced form coefficients now exceeds the number of structural parameters, the model places additional restrictions on the reduced form coefficients that must hold within and across sector categories. For example, the ratio of the coefficients on $k_c s_{ci,t-1} \cdot pn_{ci,t-1}$ and $pn_{ci,t-1}$ in (17) and (18), respectively, must be $2/(1 - 2\sigma)$ in each sector category. The identification of the structural parameters is now more difficult to assess analytically and, hence, we verified that the parameters continue to be (just) identified by estimating the model with different values of r. The point estimates of the structural parameters do change in response, but the predicted entry and average initial capital (for each value of p) remain unchanged.

[46] When matching on entry and initial capital, we weight the error term by the reciprocal of the (bootstrapped) standard deviation of the mean of each variable. The unweighted estimates are very similar to what we report in the table.

the estimating equation. Given the additive structure of the estimating equations, we can control even more flexibly for destination effects by including sector-location-time period effects. Parameter estimates with this specification are reported in Table 10, Column 3. The σ parameter estimate is entirely unchanged, and while the θ estimate does decline slightly, we cannot reject the hypothesis that the θ estimates are equal across all specifications in Table 10. Consistent with the reduced form evidence, entrepreneurs from higher population density birth counties do not appear to have preferred access to superior destinations (if they did, then the θ estimates would change when destination-time period effects are included in the estimating equation).

Table 10. Structural Estimates

Model:	benchmark	with agglomeration effects	with destination-time period effects
	(1)	(2)	(3)
σ	0.779	0.780	0.798
	(0.001)	(0.001)	(0.020)
θ	0.883	0.878	0.654
	(0.274)	(0.283)	(0.106)

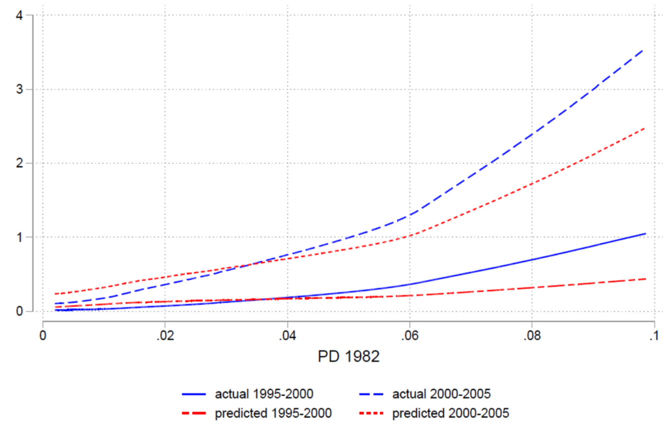
Note: all specifications include the following additional parameters (not reported); α_1, ..., α_6, and f_t (estimated in the initial capital equation, separately in time periods 1 and 2).
The augmented specification in column 2 includes the lagged stock of firms in each destination-time period to capture generalized agglomeration effects and adds time period effects to the entry equation.
The most flexible specification in column 3 includes destination-time period effects.
The destination is defined by the one-digit sector and prefecture.
Number of entrants and average initial capital are computed from the SAIC registration database and birth county population density is computed from the 1982 population census.
Bootstrapped standard errors in parentheses.

Figures 9a and 9b assesses the goodness of fit of the model by comparing actual and predicted values, separately for entry and initial capital, across the range of birth county population densities. The destination-time period effects in Table 10, Column 3 are not estimated and, hence, our most flexible specification cannot be used for prediction. The predicted values are thus based on the next most flexible specification, reported in Table 10, Column 2. Although there are just 14 parameters in this specification, it does a good job of predicting entry and initial capital across nearly 2,000 birth counties, in each time period, at a time when the Chinese economy was growing at an explosive rate.[47]

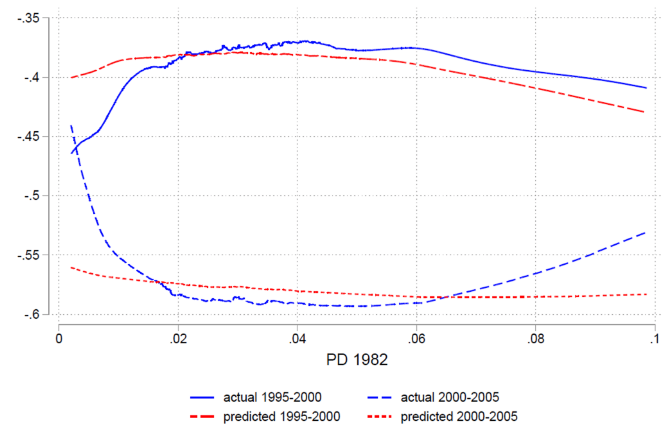
A major objective of our research is to quantify the role played by community networks in the growth of private enterprise in China. This is accomplished by setting the θ parameter to zero and then generating counter-factual entry and capital investment over the estimation period. This exercise incorporates the dynamic network multiplier effect that is a key feature of our model; exogenous initial entry in the birth county-destination induces subsequent entry, which, in turn, leads to further entry, and so on. A corresponding dynamic multiplier cannot be applied to overall entry in the destination, $n_{i,t-1}$, in the augmented specification reported in Table 10, Column 2 because it is outside the model; while this term can be included as a control in the estimation, it cannot be incorporated in the counter-factual simulation. The simulations that follow are thus based on the benchmark specification reported in Table 10, Column 1. The σ and θ estimates are almost

[47]The model under-predicts entry at the top of the population density distribution in Figure 9a, but this is on account of the fact that the distribution is skewed to the left (as documented in Appendix Figure D.3) with a mean of 0.03. The estimation thus puts more weight on matching at lower population densities. Overall, the prediction error in total entry is 9% in the 1995-1999 period and 2% in the 2000-2004 period.

Figure 9. Actual and Predicted, Entry and Initial Capital



(a) Entry



(b) Initial Capital

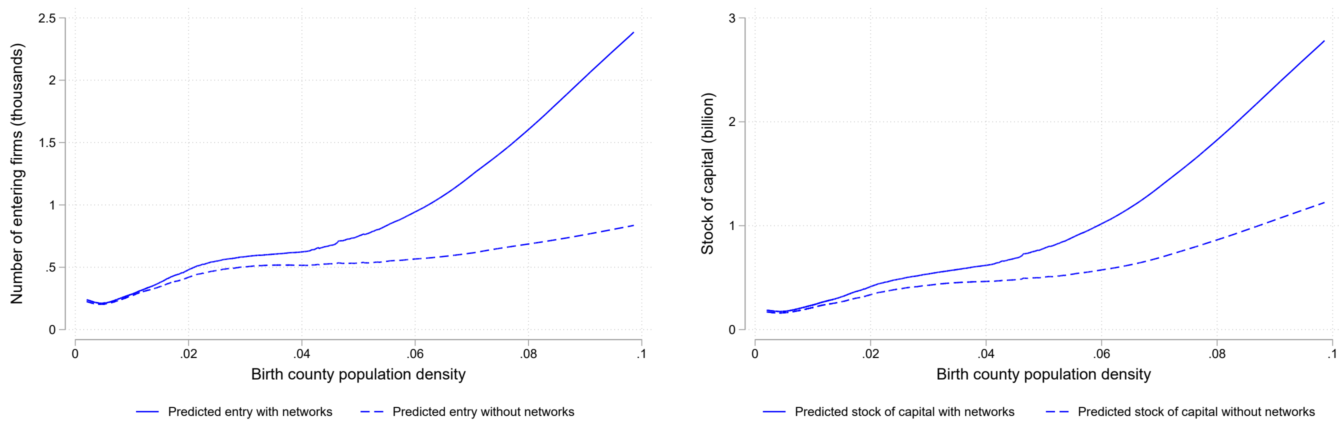
Source: SAIC registration database, model generated data, and 1982 population census.

identical for the two specifications in any case. The counter-factual simulation with entry as the outcome is reported in Figure 10a. It is evident from the figure that the number of entrants 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 predicted entrants would have declined by 21.7% over the 1995-2004 period if the networks had not been active. In a related counter-factual exercise, reported in Figure 10b, the predicted total stock of capital in 2004 (taking account of the number of firms that entered, their initial capital, and the subsequent growth in their capital) would have declined by 28.5% had the networks been absent. Adjusting for the fact that the counter-factual analysis is restricted to birth county-destination-time periods with positive lagged entry and to county-born entrepreneurs, for whom the hometown networks are relevant, this amounts to a 10.8% decline in the number of entrants and a 12.5% decline in the stock of capital for the economy as a whole.[48]

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. Song *et al.* (2011), 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 0.15; 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 11a, the total profit increase generated by the subsidy in 1995-1999 is less than the cost to the government in all birth counties. However, the spillover effect of the one-time subsidy on profits in the subsequent 2000-2004 period is substantial (and even larger than the direct effect on profits in high population density counties). This is because the credit

[48] Government infrastructure and prices remain fixed in the counter-factual simulation. If the network were shut down and the number of firms declined, then output (input) prices would increase (decrease). The resulting increase in profits would encourage some additional firms to enter. In contrast, if government infrastructure and the growth of the networks are complementary, then the removal of the networks would reduce the infrastructure level, generating a further decline in the number of firms in the counter-factual scenario.

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



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

Source: Model generated data and 1982 population census.

subsidy induces additional entry during 1995-1999, which through the compounding network effect generates large profit increases in the more socially connected counties in 2000-2004. With a discount factor of 0.8, the return on the subsidy, based on the additional (discounted) profits that were generated over the 1995-1999 and 2000-2004 periods minus the cost of the subsidy, would have been 5.7% for countries above the mean population density and -46.7% for counties below the mean.

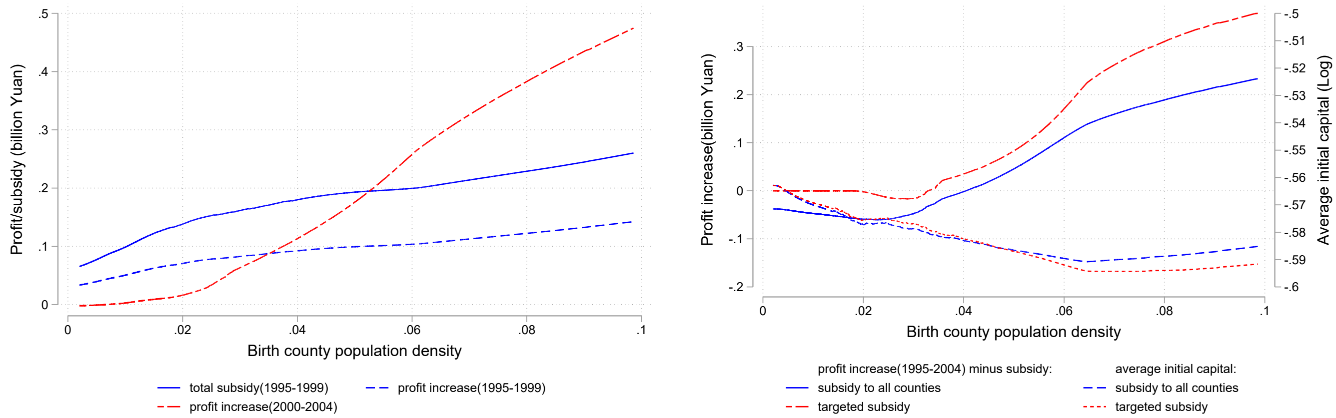
Figure 11b reports the impact of an alternative government program, which only gives the subsidy to those origin counties who would have increased their aggregate discounted profits over the 1995-2004 period by more than the amount of subsidy they received in the preceding counter-factual experiment; these counties, with a population density above 0.04, lie in the top quartile of the population density distribution. To keep the total amount of the subsidy constant, the interest rate for the targeted counties is reduced to 0.12. The increase in profits minus the subsidy received is reported across the population density distribution in the figure, both for the original subsidy scheme and for the targeted subsidy scheme. As can be seen, the targeted program does strictly better if the government's objective is to maximize total profit (less the subsidy cost). Notice also that average initial capital, which is declining with population density, declines even more steeply with the more efficient targeted program.[49] A distinguishing feature of our network-based mechanism is that efficiency-enhancing policies could actually result in even smaller firms and even greater dispersion in firm size in equilibrium (as observed in Figure 11b).

6 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 predictions for the dynamics of firm entry, sectoral and spatial concentration, and firm size across birth counties with different

[49]In our analysis, the marginal value product of capital does not vary across firms by construction. If we had used credit market imperfections to motivate network formation instead, then a efficiency-enhancing policy that exploited network spillovers would have increased the dispersion in marginal productivity within sectors (due to variation in interest rates across networks) as well. Haltiwanger *et al.* (2018) also make the point that an increase in efficiency could increase the dispersion in TFP, but with a different mechanism.

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



(a) Subsidy to all counties

(b) Targeted subsidy vs. subsidy to all counties

Source:Model generated data and 1982 population census.

levels of trust (measured by population density) when networks are active. We validate each of these predictions over a twenty year period 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 us to rule out alternative non-network based explanations. Additional results provide direct support for the network channel, indicating that spillovers occur within the birth-county-sector-location. Having validated the model, we estimate its structural parameters and conduct counter-factual simulations. The first simulation indicates that aggregate entry in the 1995-2004 period and total capital stock in 2004 would have been 11% and 12.5% lower in the economy as a whole in the absence of the rural hometown networks. While the contribution of these informal institutions to Chinese growth has thus been substantial, in line with the anecdotal evidence, this still leaves room for other factors that have been argued to have played an important role, such as government policies, infrastructure, and finance; high saving rates and foreign investment inflows; and the opening of the world market to Chinese exports.

The substantial inter-firm spillovers that we document are unlikely to be fully anticipated or internalized by individual entrepreneurs. This creates scope for industrial policies to stimulate private investment, and this is the subject of our second counter-factual simulation. This experiment, which simulates the effect of a one-time credit subsidy, shows that the optimal strategy to maximize total profits would be to target entrants from higher population density birth counties in order to take advantage of the larger resulting network externalities. There are, however, a number of caveats to such a policy prescription. 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, and this will depend on the underlying social structure. In particular, 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 (Zhang, 2016).[50]

[50]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, 2016), 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 concerns the normative consequences of such networks. Our analysis has primarily focused on their positive or descriptive consequences. With regard to efficiency, credit subsidies targeted at particular communities are justified if the resulting network spillovers increase firm productivity (and profits). The potential downside is that communities will start to lobby the government for access to cheap capital; in the extreme case, such rent seeking without accompanying productivity gains could worsen existing distortions in the economy. Moreover, there are important consequences for inequality that need to 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 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 inter-community inequality.

References

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

— and Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica*, **60** (2), 323–351.

Akcigit, U., Alp, H. and Peters, M. (2016). Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries. *NBER Working Paper Series*.

Alesina, A. and Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, **53** (4), 898–944.

Allen, F., Qian, J. and Qian, M. (2005). Law, Finance, and Economic Growth in China. *Journal of Financial Economics*, **77** (1), 57–116.

Asker, J., Collard-Wexler, A. and De Loecker, J. (2014). Dynamic Inputs and Resource (Mis) Allocation. *Journal of Political Economy*, **122** (5), 1013–1063.

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

Bai, C.-e., Hsieh, C.-T. and Song, Z. (2019). Special deals with chinese characteristics.

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

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

—, Van Biesebroeck, J. and Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, **97** (2), 339–351.

Brooks, W. J., Kaboski, J. P. and Li, Y. A. (2016). Growth Policy, Agglomeration, and (the Lack of) Competition. *NBER Working Paper Series*.

Caunedo, J. (2016). Industry Dynamics, Investment and Uncertainty. *Working Paper*.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, **86** (1), 54–70.

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

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica*, **80** (6), 2543–2594.

Ellison, G. (1994). Cooperation in the Prisoner's Dilemma with Anonymous Random Matching. *Review of Economic Studies*, **61** (3), 567–588.

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

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *Review of Economic Studies*, **81** (3), 1046–1070.

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

Fisman, R., Shi, J., Wang, Y. and Xu, R. (2018). Social Ties and Favoritism in Chinese Science. *Journal of Political Economy*, **126** (3), 1134–1171.

Fleisher, B., Hu, D., McGuire, W. and Zhang, X. (2010). The Evolution of an Industrial Cluster in China. *China Economic Review*, **21** (3), 456–469.

Goodman, B. (1995). *Native Place, City, and Nation: Regional Networks and Identities in Shanghai, 1853–1937*. University of California Press.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American economic review*, pp. 525–548.

— (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of political economy*, **102** (5), 912–950.

— and Tabellini, G. (2017). The Clan and the Corporation: Sustaining Cooperation in China and Europe. *Journal of Comparative Economics*, **45** (1), 1–35.

Haltiwanger, J., Kulick, R. and Syverson, C. (2018). *Misallocation measures: The distortion that ate the residual*. Tech. rep., National Bureau of Economic Research.

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

— (1996). *Regional Identity, Labor, and Ethnicity in Contemporary China*. Berkeley, California: Institute of East Asian Studies, University of California.

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

— and — (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.

Jackson, M. O., Rodriguez-Barraquer, T. and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, **102** (5), 1857–97.

— and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, **97** (3), 890–915.

Jones, C. I. (1995). R & d-Based Models of Economic Growth. *Journal of Political Economy*, **103** (4), 759–784.

Kandori, M. (1992). Social Norms and Community Enforcement. *Review of Economic Studies*, **59** (1), 63–80.

Levine, R. and Rubinstein, Y. (2017). Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More? *Quarterly Journal of Economics*, **132** (2), 963–1018.

Lin, J. Y. and Wang, Y. (2016). *Going Beyond Aid: Development Cooperation for Structural Transformation*. Cambridge University Press.

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

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

Morten, M. (2019). Temporary migration and endogenous risk sharing in village india. *Journal of Political Economy*, **127** (1), 000–000.

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

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

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

Nee, V. and Opper, S. (2012). *Capitalism from Below: Markets and Institutional Change in China*. Harvard University Press.

Peng, Y. (2004). Kinship Networks and Entrepreneurs in China's Transitional Economy. *American Journal of Sociology*, **109** (5), 1045–1074.

Peters, M. (2016). Heterogeneous Markups, Growth and Endogenous Misallocation. *Working Paper*.

Restuccia, D. and Rogerson, R. (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics*, **11** (4), 707–720.

Romer, P. M. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy*, **94** (5), 1002–1037.

— (1990). Endogenous Technological Change. *Journal of Political Economy*, **98** (5, Part 2), S71–S102.

Segerstrom, P. S. (1998). Endogenous Growth without Scale Effects. *American Economic Review*, pp. 1290–1310.

—, Anant, T. C. and Dinopoulos, E. (1990). A Schumpeterian Model of the Product Life Cycle. *American Economic Review*, pp. 1077–1091.

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

Summers, L. (2007). The Rise of Asia and the Global Economy. *Research Monitor*, (2).

Tsai, L. L. (2007). Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China. *The American Political Science Review*, **101** (2), 355–372.

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

Zhang, C. (2017). Culture and the Economy: Clan, Entrepreneurship, and Development of the Private Sector in China. *SSRN Scholarly Paper*, (ID 2865105).

— and Xie, Y. (2013). Place of Origin and Labour Market Outcomes Among Migrant Workers in Urban China. *Urban Studies*, **50** (14), 3011–3026.

Zhang, J. (2016). Chinese Foreign Assistance, Explained. Retrieved from https://www.brookings.edu/blog/order-from-chaos/2016/07/19/chinese-foreign-assistance-explained/.

Zhang, X. and Li, G. (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. 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 ξ 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 ability ω is

$$P_{it}(\omega) = \omega\xi A_0^{\frac{1}{1-\alpha}} \exp(\theta p n_{i,t-1}\frac{1}{1-\alpha})[1+\delta\exp(\theta(p)e_{it}\frac{1}{1-\alpha})] \tag{19}$$

while of staying in T is

$$N_{it}(\omega) = \omega^{\sigma}[1+\delta] \tag{20}$$

The agent will enter if

$$\log\omega > \frac{1}{1-\sigma}[-\log\xi - \frac{1}{1-\alpha}\log A_0 + \log(1+\delta) - \frac{1}{1-\alpha}\theta(p)n_{i,t-1} - \log\{1+\delta\exp(\theta(p)e_{it}\frac{1}{1-\alpha})\}] \tag{21}$$

Define the function

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

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

$$g(e|s_{i,t-1}, n_{i,t-1}, A_{i0}) = e \tag{22}$$

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

$$g'(e|s_{i,t-1}, n_{i,t-1}, A_{i0}) = s_{i,t-1}\frac{\delta\exp(\frac{\theta(p)e}{1-\alpha})}{1+\delta\exp(\frac{\theta(p)e}{1-\alpha})}\frac{k\theta(p)}{(1-\alpha)(1-\sigma)} \tag{23}$$

Hence if

$$\frac{k\theta\bar{p}}{(1-\alpha)(1-\sigma)} < 1 \tag{24}$$

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: Proofs

Proof of Proposition 1: We first prove that $s_{it} > s_{i-1,t}$ for all t. Suppose this is true at $t-1$: then $s_{i,t-1}$ is rising in i. Denote the growth rate of destination i share: $g_{it} \equiv \frac{s_{it}-s_{i,t-1}}{s_{i,t-1}} = \frac{N_{t-1}}{N_t} + [L+\kappa(p)N_{t-1}]s_{i,t-1} - 1$, upon using (6). Hence g_{it} is rising in $s_{i,t-1}$ and therefore in i, implying $s_{it} > s_{i-1,t}$. So shares are ordered for all cohorts exactly as they are in cohort 0. Also note that all destinations have positive shares in all cohorts, and growth rates cannot be zero at any t for all destinations.

Since $H_t \equiv \sum_i s_{it}^2 = \sum_i s_{i,t-1}^2(1+g_{it})^2 = \sum_i s_{i,t-1}^2 + \sum_i s_{i,t-1}^2 g_{it}^2 + 2\sum_i s_{i,t-1}^2 g_{it}$, it follows that

$$\begin{aligned} H_t - H_{t-1} &= \sum_i s_{i,t-1}^2 g_{it}^2 + 2\sum_i s_{i,t-1}^2 g_{it} \\ &> 2\sum_i s_{i,t-1}^2 g_{it} \\ &> 0 \end{aligned}$$

where the first inequality follows from the fact that all sector shares are positive and growth rates are not all zero. The second inequality follows from observing that: (i) if we define $x_{it} \equiv s_{i,t-1}g_{it} = s_{it} - s_{i,t-1}$ then $\sum_i x_{it} = 0$; (ii) $\sum_i s_{i,t-1} = 1$, and (iii) x_{it} and $s_{i,t-1}$ are both increasing in i, as explained above. Hence by a standard argument[51] $\sum_i s_{i,t-1}^2 g_{it} = \sum_i s_{i,t-1}x_{it} > 0$, which proves that concentration is rising in t, and hence (using (7)) the same is true for E_t.

Proof of Proposition 2: The increase in E_t, N_t, H_t with t follows from Proposition 1. So consider how a higher p alters the dynamics, given initial conditions. We claim that it raises aggregate entry E_t (and hence N_t) as well as H_t at every date t. This follows from an inductive argument. Observe first that it must be true for E_t (and N_t) at $t=1$, given the initial conditions N_0, H_0, upon applying equation (7) at $t=1$. Next observe that the right-hand-side of (8) is rising in p, given any N_{t-1} and $s_{1,t-1} > \frac{1}{2}$. Hence s_{11} must be rising in p, given the initial conditions. So the result holds for H_t at $t=1$. Next suppose it holds until some date $t-1$, i.e., N_{t-1} and H_{t-1} are rising in p. Equation (7) then implies E_t (and N_t) is rising in p. Moreover, observe that the right-hand-side of (8) is rising in N_{t-1} and in $s_{1,t-1}$, given p and $s_{1,t-1} > \frac{1}{2}$. The share s_{1t} will then be increasing in p because it is increasing in $s_{1,t-1}$, N_{t-1} and $\kappa(p)$ respectively. Induction now ensures this will be true at every t. This establishes part (a) of Proposition 2.

Turn now to part (b). Taking first differences of (7)

$$E_{t+1} - E_t = \kappa(p)[N_t H_t - N_{t-1}H_{t-1}] = \kappa(p)[E_t H_t + N_{t-1}(H_t - H_{t-1})] \tag{25}$$

Since $\kappa, E_t, H_t, N_{t-1}$ are all rising in p, the 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 this to hold is that it is true for s_{1t}: 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).

[51] The distribution across destinations first order stochastically dominates the uniform distribution, in which $s_{i,t-1}$ is the same for all i, and the expected value of x under the uniform distribution equals zero. Hence the expected value of x must be positive.

Now observe that (8) can be rewritten as

$$s_{1t} - s_{1,t-1} = \kappa(p) N_{t-1} \frac{(2s_{1,t-1} - 1)(1 - s_{1,t-1}) s_{1,t-1}}{(L + N_{t-1})(2 - s_{1,t-1}) + \kappa(p) N_{t-1}(s_{1,t-1}^2 + 1 - s_{1,t-1})} \tag{26}$$

$\kappa(p) < 1$ implies that the denominator of the right-hand-side of (26) 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. Part (b) then follows from the fact that $s_{1,t-1}, N_{t-1}$ are rising in p.

Proof of Proposition 3: To verify (a), observe that averaging (9) across destinations (and noting that $\sum_i s_{i,t-1} = 1$), initial capital of the marginal entrant is decreasing in t, p, and $p * t$ because $\theta(p)$ is increasing in p and N_{t-1} is increasing in t and p (more steeply over time).[52] A similar argument operates for ability and size of the average entrant from (10) and taking the average across destinations. Part (b) follows from averaging across destinations in (12), and applying Propositions 1 and 2.

Appendix C. 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 from a given birth county, (ii) k is the total number of destinations that they are allocated to, and (iii) $p_1, \cdots, p_k$ are the probabilities that a firm allocated randomly would end up in each of those destinations. We assume that there is an equal probability of choosing any destination; $p_j = \frac{1}{k}, \forall j$.

The expected HHI when firms make decisions 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,

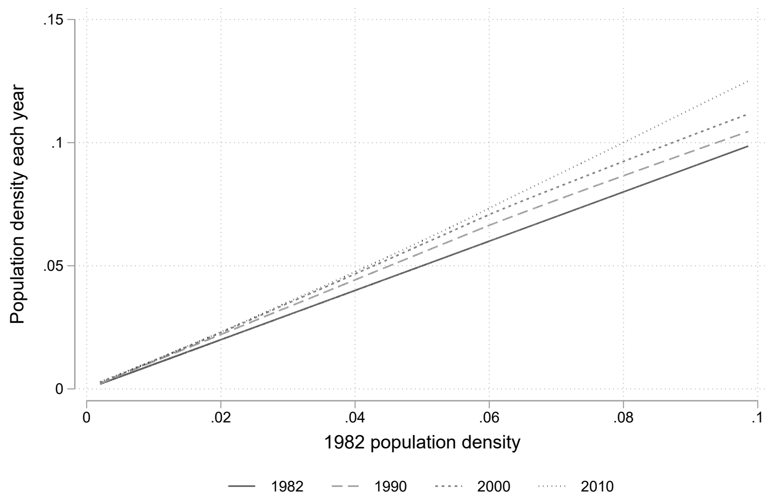
$$E(HHI) = \frac{1}{n^2}\left(k\left(\frac{n}{k}\right)^2 + k\left[n\frac{1}{k}\left(1 - \frac{1}{k}\right)\right]\right) = \frac{1}{k} + \frac{1}{n}\frac{k-1}{k}.$$

For large 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 destinations, divided by $E(HHI)$. If firms are allocated randomly, then the adjusted HHI will be close to one, providing a useful benchmark for this statistic.

[52] N_t is increasing in t (for any given p) and in p (for any given t) from Propositions 1 and 2. $N_t - N_{t-1} \equiv E_t$, which is increasing in p from Proposition 2, hence, the cross-partial derivative of N_t with respect to p and t is positive.

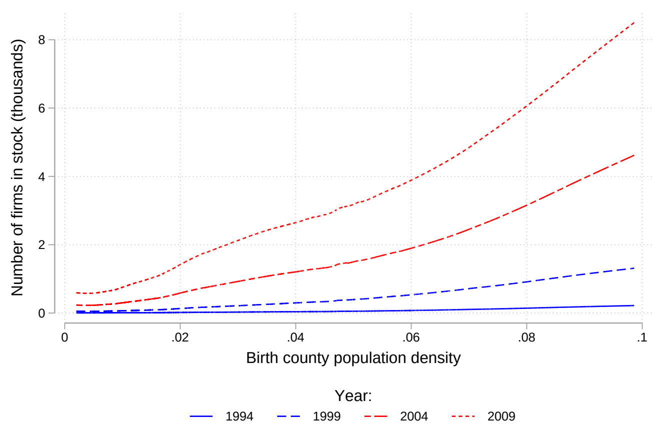
Appendix D. Figures

Figure D.1. Population Density over Time

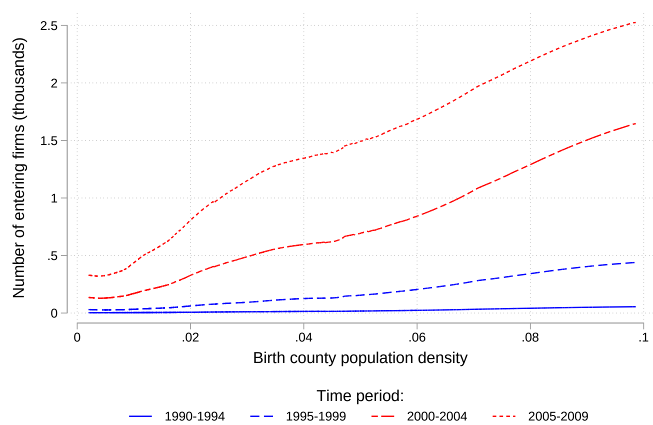


Source: Registration Database and 1982 population census.

Figure D.2. Firm Entry and Population Density



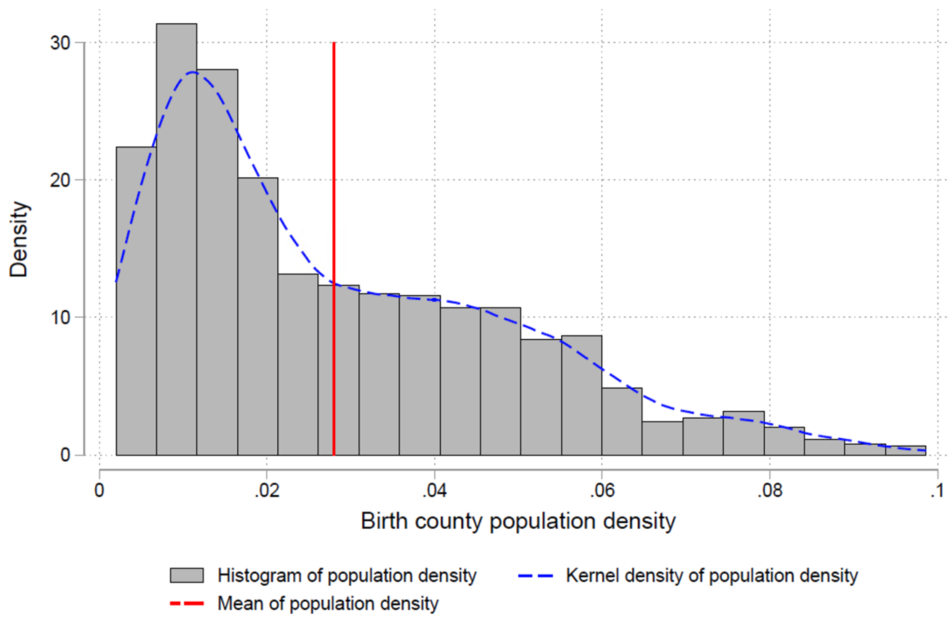
(a) Stock of Firms



(b) Firms Located Outside the Birth County

Source: Registration Database and 1982 population census.

Figure D.3. Distribution of Population Density



Source: 1982 population census.

Appendix E. Tables

Table E.1. Marginal Ability, Marginal Initial Capital and Population Density

Dependent variable:	marginal ability				marginal 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	-1.829	-2.685***	-3.570***	-1.435***	-0.100***	-0.118***	-0.134***	-0.041***
	(1.369)	(1.103)	(0.807)	(0.755)	(0.030)	(0.026)	(0.017)	(0.010)
Mean of dependent variable	66.07	52,47	40.79	40.15	-0.803	-1.093	-1.669	-2.225
Observations	4,079	6,595	10,354	11,137	15,601	46,877	83,276	99,877

Note: the entrepreneur's ability is measured by his percentile rank in his birth county-birth cohort education distribution.
The marginal entrepreneur is defined as the individual at the bottom one percentile of the ability distribution among entering entrepreneurs in a given birth county-sector-time period. Marginal initial capital defined as the bottom one percentile of the initial capital distribution at the birth county-sector-time period level.
Control variables include population, education and occupation distribution in the birth county.
Standard errors clustered at birth county level are reported in parentheses. * significant at 10%, ** at 5%, *** at 1%.

Table E.2. The Effect of Initial Entry on Subsequent Entry (for firms located outside the birth county)

Dependent variable:	subsequent entrants from the birth county			
Time period:	2000-2004	2005-2009	2000-2004	2005-2009
	(1)	(2)	(3)	(4)
Initial entrants from the birth county	5.481***	8.273***	3.303***	5.883**
	(1.044)	(1.948)	(1.197)	(2.470)
All initial entrants at the destination	0.294***	0.442***	–	–
	(0.036)	(0.047)		
Initial entrants from the birth county × birth county population density	–	–	1.820***	1.880*
			(0.507)	(0.966)
Distance to the birth county	–	–	-1.458***	-1.723***
			(0.077)	(0.069)
Mean of dependent variable	1.964	2.124	1.964	2.124
Origin-sector fixed effects	Yes	Yes	Yes	Yes
Location fixed effects	No	No	Yes	Yes
Observations	334,843	713,935	334,843	713,935

Note: number of entrants outside the birth county is measured at the (two-digit) sector-destination level.
Initial entry is derived over the 1990-1994 period.
Number of entrants is obtained from the SAIC registration database and birth county 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%.

Table E.3. The Effect of Initial Entry on Subsequent Entry (with destination defined by sector-prefecture)

Dependent variable:	subsequent entrants from the birth county			
Time period:	2000-2004	2005-2009	2000-2004	2005-2009
	(1)	(2)	(3)	(4)
Initial entrants from the birth county	8.693***	11.241***	6.298***	7.410***
	(0.817)	(1.186)	(1.134)	(1.551)
All initial entrants at the destination	0.594***	0.681***	–	–
	(0.081)	(0.078)		
Initial entrants from the birth county × birth county population density	–	–	1.642***	2.618***
			(0.633)	(1.105)
Distance to the birth county	–	–	-5.299***	-7.117***
			(0.280)	(0.261)
Mean of dependent variable	5.067	5.354	5.067	5.354
Origin-sector fixed effects	Yes	Yes	Yes	Yes
Location fixed effects	No	No	Yes	Yes
Observations	230,032	453,386	230,032	453,386

Note: number of entrants is measured at the (two-digit) sector-prefecture level.
Initial entry is derived over the 1990-1994 period.
Number of entrants is obtained from the SAIC registration database and birth county 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%.