



Spatial Determinants of Entrepreneurship in India

**Ejaz Ghani
William R. Kerr
Stephen O'Connell**

Working Paper

12-027

October 17, 2011

Copyright © 2011 by Ejaz Ghani, William R. Kerr, Stephen O'Connell

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Spatial Determinants of Entrepreneurship in India

Ejaz Ghani, William R. Kerr and Stephen O'Connell

October 2011

Abstract: We analyze the spatial determinants of entrepreneurship in India in the manufacturing and services sectors. Among general district traits, quality of physical infrastructure and workforce education are the strongest predictors of entry, with labor laws and household banking quality also playing important roles. Looking at the district-industry level, we find extensive evidence of agglomeration economies among manufacturing industries. In particular, supportive incumbent industrial structures for input and output markets are strongly linked to higher establishment entry rates. We also find substantial evidence for the Chinitz effect where small local incumbent suppliers encourage entry. The importance of agglomeration economies for entry hold when considering changes in India's incumbent industry structures from 1989, determined before large-scale deregulation began, to 2005.

Keywords: Entrepreneurship, agglomeration, development, India, South Asia.

JEL Classification: L10, L26, L60, L80, M13, O10, R00, R10, R12

Author institutions and contact details: Ghani: World Bank, Eghani@worldbank.org; Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu; O'Connell: World Bank and CUNY Graduate Center, soconnell@gc.cuny.edu.

Acknowledgments: We thank Ahmad Ahsan, Mehtabul Azam, Rajeev Dehejia, Lakshmi Iyer, Henry Jewell, Arvind Panagariya, and Hyoung Gun Wang for helpful comments on this work. We thank the World Bank's South Asia Labor Flagship team for providing the primary datasets used in this paper. We are particularly indebted to Shanthi Nataraj for sharing her wisdom regarding the industrial survey data. Funding for this project was provided by World Bank and Multi-Donor Trade Trust Fund. The views expressed here are those of the authors and not of any institution they may be associated with.

Introduction

Many policy makers want to encourage entrepreneurship in their local economies given its central role in economic growth and development. Entrepreneurship helps allocate resources efficiently, strengthens competition among firms, supports innovation and new product designs, and promotes trade growth through product variety. Perhaps most important for policy makers, high rates of local entrepreneurship are linked to stronger subsequent job growth for regions. Ghani et al. (2011) show this pattern for the manufacturing sector in India since 1990: Even after controlling for overall state and industry dynamics, places in India that had higher rates of entry at the start of the 1990s experienced stronger local job growth in the formal sector over the next two decades. Similar results are evident in the United States (e.g., Glaeser and Kerr 2011).

This importance of entrepreneurship leads to a natural, policy-relevant question: Which regional traits encourage local entrepreneurship in South Asia? Multiple studies have considered this question in advanced economies, especially for the manufacturing sector, but there is very little empirical evidence for developing countries like India. This lack of research hampers the effectiveness of policy efforts to promote job growth through entrepreneurship. The roles that education or infrastructure, for example, have for entry in an advanced economy like the United States may be quite different from a setting like India where illiteracy and lack of roads and sanitation continue to hamper development. Likewise, we have extensive evidence on the importance of agglomeration economies in advanced countries, but the relevance of these patterns in developing economies has not been consistently established.

We investigate these questions for manufacturing and services in India at the district level. Within these two industry groups, we also compare the formal and informal sectors. We quantify the factors and traits of districts and industries that systematically predict stronger entry levels in recent years. Much of the work on spatial determinants of entrepreneurship for advanced economies focuses on manufacturing, and so it is natural to begin there and compare the patterns for India with those in the United States. Our goal is to determine the general degree to which the economic geography of India can be explained with a parsimonious set of specifications, and to compare the specific factors identified to be important in the two contexts. This work most closely relates to Glaeser and Kerr (2009), Mukim (2011), and Jofre-Monseny et al. (2011).¹

Our second sector is services. Ghani (2010) describes the unique role of the services sector in India's development, in part allowing India to overcome its underdevelopment in

¹ Other closely related studies include Chinitz (1961), Acs and Armington (2006), Drucker and Feser (2007), Rosenthal and Strange (2010), and Delgado et al. (2010a,b). Mukim and Nunnenkamp (2011) consider location decisions of foreign firms in India.

manufacturing. Given the importance of services to current South Asian growth, we quantify it in an estimating framework similar to that of manufacturing. This helps us identify where similarities and differences exist across the two industry groups. With the exception of Mukim (2011) discussed below, we are not aware of prior work on services entry similar to what we undertake.

Several important themes emerge from our study:

1. *India's economic geography is still taking shape* – We use an apparatus for our estimations that is similar to that used by Glaeser and Kerr (2009). For the United States, existing city population levels, city-industry employment, and industry fixed effects can explain 80% of the spatial variation in entry rates. The comparable explanatory power for India is 29% for manufacturing and 33% for services. While this lower explanatory power could be due in part to dataset features and/or subtle, necessary shifts in the empirical estimations, it is clear that a large portion of this gap is due to India being at a much earlier stage of development, both generally and for these particular sectors. The industrial landscape is also adjusting after the deregulations of the 1980s and 1990s (e.g., Fernandes and Sharma 2011). At such an early point and with industrial structures not entrenched, local policies and traits can have profound and lasting impacts by shaping where industries plant their roots.
2. *Agglomeration economies are very important for India's entry patterns* – In a similar manner, we find extensive evidence that the incumbent compositions of local industries influence new entry rates at the district-industry level within manufacturing. This influence is through both traditional Marshallian economies like a suitable labor force and proximity to customers (e.g., Duranton and Overman 2005, Duranton and Puga 2004, Rosenthal and Strange 2004) and through the Chinitz (1961) effect that emphasizes small suppliers. These factors are especially pronounced in conditional estimations that control for both district and industry fixed effects, with magnitudes similar to or greater than advanced economies.
3. *Education and physical infrastructure matter greatly* – The two most consistent factors that predict overall entrepreneurship for a district are education and the quality of local physical infrastructure. These patterns are true for manufacturing and services, and the relationship is stronger than that found for the United States. Higher education in a local area increases the supply of entrepreneurs and increases the talent available to entrepreneurs for staffing their companies. Investment in people is an easy call for policy makers. Likewise, local areas must provide adequate electricity, roads, telecom, and water/sanitation facilities. Entrepreneurs are especially dependent upon these public goods.

Beyond these three key findings, we identify other district-level traits that influence entrepreneurship and/or confirm prior work. For example, several studies (e.g., Besley and

Burgess 2004, Aghion et al. 2008) link strict labor regulations in India to slower economic growth and development. We find this pattern too, especially for the organized manufacturing sector where these laws are most binding.

Our study thus makes several contributions to the literature. Most importantly, we are among the first studies to quantify the spatial determinants of entrepreneurship in India. Moreover, we move beyond manufacturing to consider services, which are very important for India's economic growth, and we compare the formal and informal sectors within each industry group. More broadly, we and Mukim (2011) are the first studies to apply the incumbent industrial structures frameworks of Glaeser and Kerr (2009) and Jofre-Monseny et al. (2011) to a developing economy, thereby providing insights into how agglomeration economies resemble and differ from each other. We also consider changes in industrial conditions from 1989, before India's large-scale deregulation began, to 2005 in order to provide greater empirical traction than in prior cross-sectional work on agglomeration economies. More research on agglomeration economies and entrepreneurship in developing countries is important for urban and development economics going forward.

In contemporaneous and independent work, Mukim (2011) also examines spatial entry patterns for India's informal sector. Encouraging, despite some differences in specifications choices and empirical approach, the two studies both point to special roles for input-output linkages across firms and infrastructure investments. Our work primarily differs from Mukim (2011) in its more direct comparisons of the formal and informal sectors, its more explicit focus on within-district variations across industries (i.e., the conditional estimations), and its use of long-term changes from 1989 for better traction in the formal sector. Mukim (2011) considers agglomeration economies within services at a deeper level and uses historical land revenue institutions to instrument for current industrialization.

Identifying these attributes and acting upon them is essential to foster economic growth. Figure 1 shows that entrepreneurship rates are lower in South Asia than what its stage of development would suggest. Given that entrepreneurship leads to job growth for India in the formal sector (Ghani et al. 2011), entrepreneurship will play a fundamental role in urban economics for South Asia in the decades ahead. Fernandes and Pakes (2010) observe that India's manufacturing sector is underdeveloped relative to economies of similar size, and stronger entrepreneurship will help close these gaps. Proper local conditions will also help move people out of subsistence entrepreneurship and into entrepreneurship in the formal sector. Khanna (2008) emphasizes entrepreneurship for India's future, and reallocation can help close India's productivity gap (e.g., Hsieh and Klenow 2009).

While focusing on entrepreneurship, we recognize that large firms also play an important role in regional development. The giant firms of South Asia are becoming globally powerful and

growing in efficiency, and they too will shape employment opportunities in the decades ahead. However, the history of regional development shows that big firms are not sufficient. An entrepreneurial foundation that provides for local growth and regeneration is essential for long-term success and prosperity. This is evident in the current struggles of Detroit, Michigan, and its car manufacturers. Jane Jacobs (1970) highlights how Manchester, England, and its giant textile mills in the 1800s were a model of short-term efficiency and power but also were ultimately insufficient for long-term regional growth. Likewise, the very dynamic times ahead for South Asia require that entrepreneurship be embraced and supported at the regional level.

The plan of this paper is as follows. The next section discusses our entrepreneurship data and spatial differences in entry across India. Section 3 reviews the determinants of spatial locations for firms and our metrics. Sections 4 and 5 provide our empirical results for unconditional and conditional estimations, respectively, and the last section concludes.²

Spatial Entrepreneurship Rates in India

This section discusses our two primary datasets and describes the spatial variation on entrepreneurship in India. We first define how we will measure entrepreneurship.

Definitions of Entrepreneurship

Defining entrepreneurship is hard and controversial, even in advanced countries.³ One approach, dating back to Cantillon (1730), is to describe entrepreneurship as the number of people leading independent businesses. Evans and Jovanovic (1989), Blanchflower and Oswald (1998), and many others have accordingly used self-employment rates as their metric of entrepreneurship. This choice is often made by default given that self-employment questions have typically been included in censuses of families or households. Thus, data on self-employment rates became available much earlier than any other potential measure.

Recent studies note, however, the challenges that come with using self-employment metrics to describe the entrepreneurship necessary for economic growth and job creation. Due to

² For brevity, we do not provide an extensive background on the entrepreneurship literature, instead highlighting relevant studies throughout the paper. Ghani et al. (2011) provide a short introduction, and Parker (2009) provides a complete review. Deichmann et al. (2008) survey prior work on firm locations in developing economies. Sections 2 and 3 draw on Ghani et al. (2011) and Glaeser and Kerr (2009) for describing our entrepreneurship data and metrics, respectively.

³ See, for example, the work of the OECD Entrepreneurship Indicators Programme in 2008 and 2009.

the much larger raw count of self-employed workers, self-employment indices accord much more weight to small-scale, independent businesses and hobby entrepreneurship than entrepreneurship that can lead to substantial job creation for others. The vast majority of self-employment businesses as captured on labor and census surveys will not generate employment opportunities for other workers, and they may even be the product of a lack of employment opportunities for the business owner (e.g., Astebro et al. 2010, Schoar 2009). This latter factor is particularly acute in South Asia given its earlier development and large, persistent informal sector.

These criticisms are evident in some simple examples. Silicon Valley is the poster child of US regional entrepreneurship and the world's largest venture capital market. Yet San Jose, CA, ranks near last among America's 300+ metropolitan areas in terms of self-employment, with West Palm Beach, FL, instead having the highest self-employment rate. This questionable pattern is also evident in country rankings. Southern European countries like Portugal and Greece rank much higher than Scandinavian countries in terms of self-employment within Europe, but much lower in terms of most growth-entrepreneurship indicators like venture capital markets (e.g., Bozkaya and Kerr 2011). Klapper et al. (2010) document the negative correlation between self-employment levels and economic development across a broad cross-section of countries, and Ghani et al. (2011) show that self-employment metrics for India perform poorly for describing job growth compared to better-specified alternatives.

Hence, most recent studies of entrepreneurship and growth instead focus exclusively on formal firms that employ paid workers. These thresholds—being incorporated, paying payroll taxes—are in some sense arbitrary, but they are natural given our fundamental interest in describing entry that leads to regional job creation and growth. Their relevance is even greater for South Asia given the need to pull members of the informal economy into the formal sector. Here too, however, there are substantial disagreements as to what should be measured. A vast, earlier literature develops estimates of the number or share of small businesses in a region. A similar metric is average firm size (e.g., Glaeser 2007).

These measures of small businesses overcome some of the problems of earlier self-employment measures, but they still weight very heavily small businesses that are not attempting to grow substantially (e.g., a small family-run business that has employed the same number of workers for many years). In other words, they do not capture the important dynamic aspects of entrepreneurship for economic growth. These metrics may also be additionally skewed in the Indian context given the well-known problem of the 'missing middle' in the firm-size distribution, which is often associated with rigid employment laws. A large count of small firms may not represent an entrepreneurial cluster very well in an environment where labor laws are particularly binding for firm size (e.g., Besley and Burgess 2004).

Understandably, many researchers are instead drawn to metrics that are more tightly connected to the dynamic nature of entrepreneurship. The more recent micro-data have both enabled these measurements and stressed their better performance relative to small business counts (e.g., Haltiwanger et al. 2010). One approach focuses on start-ups within a single industry so that finer characterizations and case studies can be made (e.g., Saxenian 1994, Feldman 2003). Others look at new product introductions (e.g., Audretsch and Feldman 1996), venture capital placement (e.g., Samila and Sorenson 2011), or the founding of new firms (e.g., Kerr and Nanda 2009, Rosenthal and Strange 2010). These dynamic measures of entrepreneurship are less available than self-employment rates or small business statistics, but are more tightly connected to entrepreneurship that can generate job growth.

We follow this emerging strand and principally define entrepreneurship as the presence of young establishments. Haltiwanger et al. (2010) emphasize how young firms are more tightly associated with employment growth than small firms (conditioning on age) in the US. While we would also like to employ measures of entering establishments in their first year (e.g., Glaeser and Kerr 2009), the young establishment definition—less than three years old—is the most refined measure feasible for India at this time. Incumbent establishments, which are used to model existing activity in the district-industry and Marshallian spillovers, are defined as firms that are three years old or more. We principally define entry measures through employment in these new establishments. We also test robustness using an entry share measure based upon establishment counts.

Most of the Indian surveys described next unfortunately do not record whether new establishments belong to larger, multi-plant firms or whether they are independent enterprises. While it would be advantageous to study single-unit starts separately from multi-unit expansions, we are unable to do so consistently. Encouragingly, we are able to separate establishment types for organized manufacturing, and we find very similar results to those presented below when modeling the single-unit start-up entry rate separate from multi-unit expansions, but we are unable to ascertain this stability across all of the sectors we study. The Indian context is one in which a major limitation for development is the growth and replication of successful initial businesses. Thus, from this perspective, many policy makers are as concerned about encouraging entry of new expansion establishments as they are initial start-ups.

Thus our study of entrepreneurship falls between two more common types of studies, thereby hitting on a key element of South Asia's future. On one hand, we purposely steer clear of self-employment measures. Even for the United States, self-employment is a second-best link for measuring entrepreneurship for job creation. These methodological challenges for low return or subsistence efforts are compounded in South Asia where the informal sector employs 90% of workers (e.g., Schoar 2009, Ardagna and Lusardi 2008). To realize long-term employment growth, it is necessary to distinguish transformative entrepreneurship from subsistence

entrepreneurship, and Ghani et al. (2011) show the strong job growth linkage to entry into the formal sector. We thus focus our efforts on describing the spatial distribution of these entrepreneurs.

On the other hand, we also do not study the development of very high-growth entrepreneurship or venture capital markets in South Asia (except to the extent that they are part of our official statistics). This is not because these entrepreneurs are not important for South Asia; quite the opposite. Many recent studies focus on software's emergence, returnee entrepreneurs, diaspora, and similar exciting developments (e.g., Arora and Gambardella 2005, Kerr 2008, Nanda and Khanna 2010, Agrawal et al. 2011). Yet these specialized sectors are still an extremely small part of the Indian economy, and we focus more on identifying measures of entrepreneurship like young businesses that apply to the entire manufacturing and services sectors.⁴

Indian Entrepreneurship Data in Manufacturing and Services

We employ cross-sectional establishment-level surveys of manufacturing and service enterprises carried out by the Government of India. Our primary manufacturing data are taken from surveys conducted in fiscal years 2005-06; the services sector data come from 2001-02. Even though these surveys were undertaken over two fiscal years, for simplicity we refer below to the initial year only. This section describes some key features of these data for our study, and our unpublished data appendix (available upon request) provides greater details. Nataraj (2009), Kathuria et al. (2010), Hasan and Jandoc (2010), Dehejia and Panagariya (2010), and Fernandes and Pakes (2010) provide detailed overviews of similarly constructed databases.

It is important to first define and characterize the distinction between the organized and unorganized sectors in the Indian economy, which our estimations consider separately. These distinctions in the Indian context relate to establishment size. In manufacturing, the organized sector is comprised of establishments with more than ten workers if the establishment uses electricity. If the establishment does not use electricity, the threshold is 20 workers or more. These establishments are required to register under the India Factories Act of 1948. The unorganized manufacturing sector is, by default, comprised of establishments which fall outside the scope of the Factories Act. Unorganized establishments do not pay taxes and are generally outside the purview of the state, thus approximating common definitions of the informal sector (e.g., Kanbur 2011).

⁴ Our definition of entrepreneurship places this study in a well-established literature examining the locational decisions of plants, primarily studied in developed countries. Examples include Due (1961), Oakland (1978), Schmenner (1982), Carlton (1983), Bartik (1995), and McConnell and Schwab (1990).

For services, there is no simple legal distinction as in manufacturing. Service establishments, regardless of size or other characteristics, are not required to register and thus are all officially unorganized. There are various existing methodologies to comparably differentiate small-scale, autonomous establishments from larger employers which constitute the organized sector, as generally defined. We assign establishments with less than five workers and/or listed as an “own-account enterprise” (OAE) to the unorganized sector. OAE enterprises are firms that do not employ any hired worker on a regular basis. The choice of five employees as the size cutoff recognizes that average establishment size in services is significantly smaller than in manufacturing. Using this demarcation, the organized sector makes up approximately 25% of employment in both manufacturing and services.

The organized manufacturing sector is surveyed by the Central Statistical Organisation every year through the Annual Survey of Industries (ASI), while unorganized manufacturing and services establishments are separately surveyed by the National Sample Survey Organisation (NSSO) at approximately five-year intervals. The survey years we employ are the most recent data by sector for which the young establishment identifiers are recorded. Establishments are surveyed with state and four-digit National Industry Classification (NIC) stratification. Districts are administrative subdivisions of Indian states or territories. We use the provided sample weights to construct population-level estimates of total establishments and employment by district and three-digit NIC industry. Employment is formally defined as “persons engaged” and includes working owners, family and casual labor, and salaried employees.

Currently there are approximately 630 districts spread across 35 states/union territories. In order to overcome empirical issues presented by small districts or those with an insufficient sample size, we exclude any districts with a population less than one million (based on 2001 census) or with fewer than 50 establishments sampled. For our main specifications we also exclude states with persistent conflict and political turmoil that make data quality questionable. After these adjustments, the resulting sample retains districts in 20 major states that include more than 94% of Indian employment in both manufacturing and services.

Table 1 provides some base descriptive statistics across the full sample, while Tables 2A-3B and Figures 2 and 3 show more detailed patterns by state. Table 2A shows that there are just over 14,000 young establishments in India’s organized manufacturing sector in 2005 for our sample. This reflects an entry rate of approximately 15%, using a weighted average across states, which varies spatially to a large degree. Among the larger states in terms of manufacturing employment, entry rates are highest in Uttar Pradesh and Karnataka at 18%-22%. Within Uttar Pradesh, the most entrepreneurial districts are Dehradun, Fatehpur, Faizabad, and Nagar Hardwar. The most entrepreneurial districts in Karnataka are Bangalore-Rural, Tumkur, Bangalore-Urban, and Dakshina Kannada. While possessing smaller manufacturing bases, entry rates are also high in Himachal Pradesh and Orissa.

The unorganized manufacturing sector in Table 2B has far more new establishments in any given year—almost 1.9 million in 2005 for our sample—although the entry rate is lower than the organized sector at 12%. Similar to the negative patterns discussed for self-employment versus formal entry in the United States, there is negative correlation of -0.2 between spatial entry rates for organized and unorganized sectors across states. High rates of unorganized entry are found in Delhi, Haryana, and Kerala, while Bihar, Karnataka, and Orissa have among the lowest rates. These contrasts are even starker when using self-employment measures: for the 15 districts with self-employment accounting for greater than 50% of total district employment, none have an organized sector entry rate above 5%.

In the organized services sector in Table 3A, there are about 120,000 young establishments in 2001, representing an entry rate of 20%. The highest rates are evident in Andhra Pradesh and Karnataka, with a number of other states closely following with entry rates of 20%-25%. Gujarat has the lowest entry rate. State-level entry rates for organized services have a 0.4 correlation to those in organized manufacturing. Table 3B shows that the largest entry levels in absolute terms occur in the unorganized services sector, with over 2.2 million establishments at a rate of 17%. Entry rates are particularly high in Kerala and Haryana. Unorganized and organized activities are more closely linked in services than in manufacturing with a spatial correlation across states of 0.3 for services.

Determinants of Entrepreneurship

We now describe the spatial and industrial factors that we use to predict the entrepreneurship patterns. We follow Glaeser and Kerr (2009), Jofre-Monseny et al. (2011), and Alcacer and Chung (2010) in our design of many these factors, and we refer interested readers to these papers for additional details on agglomeration metrics and their properties.

Population and Demographics

Our initial explanatory measures naturally focus on basic traits of the district: population, age profiles of the population (demographic dividend), and population density. Our population control comes from the 2001 population census, and models several effects. First, and most important, it provides a measure of the size of local markets in terms of consumers. For some industries, especially in services, these local markets constitute the firm's primary product market. For other industries that are traded at a distance, this output feature of the local market is less essential. The overall size of the local district nevertheless represents an important measure for the overall surrounding economic activity (e.g., general availability of workers).

The population control will also reflect to some degree the supply of potential entrepreneurs. Most entrepreneurs start their businesses in their current local area: for example, Dahl and Sorenson (2007) document that over 70% of new firms in Denmark are founded in commuting regions where entrepreneurs were previously living. While some entrepreneurs move to new cities to start their businesses, this is mostly confined to niche, high-growth industries like biotech where a few dominant clusters form. Looking across all industries and types of firms, several studies have instead shown an underlying pattern where entrepreneurs are disproportionately located in their home regions compared to wage employees. This pattern has been shown in the United States, Italy, and Portugal (e.g., Figueiredo et al. 2002, Michelacci and Silva 2007). The population control will pick up some of this supply side effect.

We next consider the age structure of the district, which is also referred to as the “demographic dividend” in the Indian context. The propensity to start firms changes over the lifetimes of individuals (e.g., Evans and Leighton 1989), and the age structure of a district can have an additional effect on the entry rate. For example, Bönte et al. (2009) document an inverted-U shape between regional age structures and entrepreneurship rates in Germany, and Glaeser and Kerr (2009) found high manufacturing entry rates in cities that had a disproportionate share of workers aged 20-40 years old. We construct the demographic dividend measure as the ratio of working age population to non-working age population using 2001 population census counts. The inverse of this measure is commonly known as a dependency ratio.

Third, we include a measure of population density. Higher population density again reflects some measure of local market size, but it also goes beyond to include the competition for local resources. While local sales are easier with a denser market, higher population density is also associated with higher wage levels and higher land rents. Density has also been linked to stronger knowledge flows (e.g., Carlino et al. 2007, Arzaghi and Henderson 2008). Thus the relationship between entry and population density is unclear, although many studies of advanced economies link higher population density to reduced manufacturing entry rates, especially for larger plants that are using established production techniques and seeking to minimize costs.

District-Level Conditions

Beyond these basic demographics, we consider five primary traits of districts: education of the local labor force, quality of local physical infrastructure, access or travel time to major Indian cities, stringency of labor laws, and household banking conditions. While these traits do not constitute an exhaustive list of local conditions, they are motivated by the literatures on entrepreneurship and India’s development. Unless otherwise noted, these traits are taken from the 2001 Population Census data for India.

We first measure the general education level of the local labor force. Doms et al. (2010) find that local skill levels correlate with higher rates of self-employment and better start-up performance in the United States, and Glaeser et al. (2010) associate higher education levels with higher entry rates using employment-based entry metrics. Many local policy makers stress developing the human capital of their workforces, and India is no different (Amin and Mattoo 2008). We measure the general education level of a district by the percentage of adults with a graduate (post-secondary) degree. All of our results below are robust to alternatively defining a district's education as the percentage of adults with higher secondary education.

Our second trait is the physical infrastructure level of the district. Basic services like electricity are essential for all businesses, but new entrants can be particularly dependent upon local infrastructure (e.g., established firms are better able to provision their own electricity if need be, which is quite common in India). Aghion et al. (2011) provide a theoretical model of this dependency. Many observers cite upgrading India's infrastructure as a critical step towards economic growth, and the Indian government has set aside substantial funds for this investment. The population census provides figures on the number of villages in a district which have telecommunications access, electricity access, paved roads, and access to safe drinking water. We calculate the percentage of villages that have infrastructure access within a district and sum across the four measures to create a continuous composite metric of infrastructure which ranges from zero (no infrastructure access) to four (full access to all four infrastructure components).⁵

India's economy is undergoing dramatic structural changes (Desmet et al. 2011). From a starting point in the 1980s when the government used licensing to promote industrial location in regions that were not developing as quickly, the economic geography of India has been in flux as firms and new entrants shift spatially (e.g., Chari 2008, Fernandes and Sharma 2011). One feature for a district that is important in this transformation is its link to major cities. We thus include a measure from Lall et al. (2011) of the driving time from the central node of a district to the nearest of India's ten largest cities⁶ as a measure of physical connectivity and across-district infrastructure. This is calculated based on data on India's road networks using GIS software.

We next model local labor regulations using state-level variation in policies. Several studies link labor regulations in Indian states to economic progress (Besley and Burgess 2004, Aghion et al. 2008), and strict labor regulations have been found to hurt entrepreneurship beyond

⁵ In six districts (major cities) which were not further subdivided into separate geographic units, these indicators were not reported in the Census data. In these cases we assign the infrastructure access components as 100%. Our results are fully robust to excluding these major cities from the analysis sample.

⁶ These are Ahmedabad, Bangalore, Bhubaneswar, Chennai, Delhi, Guwahati, Hyderabad, Kolkata, Mumbai, and Patna.

self-employment in advanced economies (e.g., Bozkaya and Kerr 2011). This effect may occur through reduced likelihood of wanting to start a new firm, or through reduced likelihood of opening new facilities to avoid regulations. There may also be reduced ‘push’ into entrepreneurship with more protected employment positions. Fallick et al. (2006) model how labor mobility affects industry structure, with implications for entrepreneurship rates. Our measure is taken from Ahsan and Pages (2007), who break down the labor regulations index proposed by Besley and Burgess (2004) into separate components affecting labor adjustment and labor disputes legislation. Using these separate measures, we create a composite labor regulations index by state.

Our final measure is the strength of household banking environments. A large body of work considers the link between financial constraints and entrepreneurship, as surveyed in Kerr and Nanda (2011). Our measure of local financing conditions is the percentage of households that have banking services as measured in the 2001 census. This measure is likely to be particularly reflective of financing environments for unorganized sector activity.⁷

Agglomeration Theories

The above factors are district-level phenomena. Some factors will matter more for certain industries than others—for example, industries that have high labor inputs will be more sensitive to general labor regulations than capital-intensive sectors—and we discuss these interactions below to test the robustness of the identified traits. But we also want to define metrics that quantify how suitable the local industrial environment is for a particular industry. For example, does the local industry mix employ the types of occupations needed by start-up companies for a given district-industry? We will consider these forces within the manufacturing sector.

We develop metrics that unite the incumbent industrial structures of cities with the extent to which industries interact through the traditional agglomeration rationales first defined by Marshall (1920). This conceptual approach has also been used to describe location choice decisions and city structures in several advanced economies: Glaeser and Kerr (2009) for the United States, Jofre-Monseny et al. (2011) for Spain, and Dauth (2011) for Germany. Alcacer and Chung (2010) use this framework to consider foreign direct investment decisions. This paper

⁷ We have also tested additional district-level indicators, including metrics of caste/religious population, female-to-male sex ratio, conflict, consumption and income levels, trade levels, and general development levels (leading/lagging designations at the state and district level). These additional controls do not substantively affect the results presented, and we maintain the more parsimonious specification to mirror other work from outside of India. Iyer et al. (2011) investigate the relationship of India’s caste system to entrepreneurship.

and Mukim (2011) are the first applications of this technique to an emerging economy of which we are aware.⁸

In all of our estimations, we first control for the size of the incumbent district-industry employment. This is important given that entrepreneurs often leave incumbent firms to start their companies. Klepper (2010) shows in detail the importance of this spawning process in the history of Detroit and Silicon Valley, and many econometric studies find the existing business landscape the most important factor for the spatial location of new entrants (e.g., Glaeser and Kerr 2009). We use incumbent establishments only for this measure and the Marshallian metrics discussed next.⁹

The first agglomeration rationale is that proximity to customers and suppliers reduces transportation costs and thereby increases productivity. This reduction in shipping costs is the core agglomerative force of the new economic geography (e.g., Fujita et al. 1999). Where customers and suppliers are geographically separate, firms must trade off distances. While transportation costs have declined dramatically over the past two centuries, the fundamentals remain important even in advanced economies. For example, much of the clean energy production in the United States based upon biomass is located near sources for the raw inputs.

To test the importance of this mechanism, we measure the extent to which districts contain potential customers and suppliers for a new entrepreneur. We begin with an input-output table for India developed by India's Central Statistical Organization. We define $Input_{i \leftarrow k}$ as the share of industry i 's inputs that come from industry k , and $Output_{i \rightarrow k}$ as the share of industry i 's outputs that go to industry k . These measures run from zero (no input or output purchasing relationship exists) to one (full dependency on the paired industry). These shares are calculated relative to all input-output flows and are not symmetrical by design ($Input_{i \leftarrow k} \neq Input_{k \rightarrow i}$, $Input_{i \leftarrow k} \neq Output_{k \rightarrow i}$).

Following Glaeser and Kerr (2009), we summarize the quality of a district d in terms of its input flows for an industry i as $Input_{di} = - \sum_{k=1, \dots, I} \text{abs}(Input_{i \leftarrow k} - E_{dk}/E_d)$, where I indexes industries. This measure simply aggregates absolute deviations between the proportions of industrial inputs required by industry i and district d 's actual industrial composition, with E

⁸ Several papers assess the relative importance of these determinants for industrial agglomeration, including Audretsch and Feldman (1996), Ellison and Glaeser (1997), Rosenthal and Strange (2001, 2003), Henderson (2003), Ellison et al. (2010), and Greenstone et al. (2010). These assessments are harder empirically than our current exercise due to the endogeneity of linkages that form between clustered firms. We take these linkages among incumbent firms to be exogenous to new entrants in this paper. Coagglomeration behavior is more broadly analyzed by Ellison et al. (2010) and Helsley and Strange (2010).

⁹ Mukim (2011) provides extensive tabulations of agglomeration levels across Indian industries.

representing employment. The measure is mostly orthogonal to district size, which we separately consider, and a negative value is taken so that the metric ranges between negative two (i.e., no inputs available in the local market) and zero (i.e., all inputs are available in the local market in precise proportions). The construction of Input_{di} assumes that firms have limited ability to substitute across material inputs in their production processes.¹⁰

To capture the relative strength of output relationships, we also define a consolidated metric $\text{Output}_{di} = \sum_{k=1, \dots, I} E_{dk}/E_d \cdot \text{Output}_{i \rightarrow k}$. This metric multiplies the national share of industry i 's output sales that go to industry k with the fraction of industry k 's employment in district d . By summing across industries, we take a weighted average of the strength of local industrial sales opportunities for industry i in the focal market d . This Output_{di} measure takes on higher values with greater sales opportunities. Unlike our input measure, this output metric pools across industries that normally purchase goods from industry i . By measuring the aggregate strength of industrial sales opportunities in district d , the metric assumes that selling to one large industrial market is the same as selling smaller amounts to multiple industries.

Related to these customer/supplier metrics is the Chinitz effect. Chinitz (1961) observed that entrepreneurs often find it difficult to work with large, vertically-integrated suppliers. The entrepreneur's order sizes are too small, and often the entrepreneur's needs are non-standard. Chinitz emphasized the role of small input suppliers in his account of entrepreneurial differences between New York City and Pittsburgh. Chinitz argued that the large, integrated steel firms of Pittsburgh depressed external supplier development; moreover, existing suppliers had limited interest in providing inputs to small businesses. By contrast, New York City's much smaller firms, organized around the decentralized garment industry that then dominated the city, were better suppliers to new firms. A number of empirical studies for the United States have emphasized the role of the Chinitz effect in local start-up conditions: Glaeser and Kerr (2009, 2011), Rosenthal and Strange (2010), Glaeser et al. (2010), and Drucker and Feser (2007).

We quantify the Chinitz hypothesis—as distinct from the high-quality, general-input conditions of Marshall (1920) captured in Input_{di} —through a metric that essentially calculates the average firm size in district d in industries that typically supply a given industry i : $\text{Chinitz}_{di} = \sum_{k=1, \dots, I} \text{Firms}_{dk}/E_d \cdot \text{Input}_{i \leftarrow k}$. Glaeser and Kerr (2009) provide more details on this metric's design. Higher values of the Chinitz_{di} metric indicate better supplier conditions for entrepreneurs in particular.

¹⁰ The input metric is not perfectly orthogonal to district size to the degree to which larger districts have more independent economic zones than smaller zones. Thus, even if the very localized input conditions within a small and large district are similar for a start-up, the measured quality of input conditions will be less in the larger district as the input metric will sum over more economic zones. We thank Juan Alcacer for pointing this out.

Moving from material inputs, labor is perhaps the most important input into any new firm, and entrepreneurship is quite likely to be driven by the availability of a suitable labor force (e.g., Combes and Duranton 2006). While a district's education and basic demographics are informative about the suitability of the local labor force, these aggregate traits can miss the very specialized nature of many occupations. As an extreme example, Zucker et al. (1998) describe the exceptional embodiment of human capital in specialized workers in the emergence of the US biotech industry. These specialized workers are often tightly clustered together (Rosenthal and Strange 2010).¹¹

Studies of the United States are able to model direct occupational flows between industries (e.g., Ellison et al. 2010, Glaeser and Kerr 2009). We unfortunately lack such data for India. We instead take a very simple approach. Greenstone et al. (2010) calculate from the Current Population Survey the rate at which workers move between industries in the United States. Using their measure of labor similarity for two industries, we define $\text{Labor}_{di} = \sum_{k=1, \dots, I} E_{dk}/E_d \cdot \text{Mobility}_{i \leftarrow k}$. This metric is a weighted average of the labor similarity of industries to the focal industry i , with the weights being each industry's share of employment in the local district. The metric is again by construction mostly orthogonal to city size.

These metrics condense large and diverse industrial structures for cities into manageable statistics of local industrial conditions. The metrics do have limitations, though. First, we do not capture potential interactions that exist beyond the local district, but factor and product markets can be wider than a district (e.g., Rosenthal and Strange 2001, Kerr and Kominers 2010). Second, the metrics do not consider final consumers. In unconditional estimates, we separately model city populations. Third, the metrics do not measure differences across districts in worker or input quality except to the extent that they are reflected in industry compositions or observable features like education levels. Finally, these metrics can suffer from omitted variable biases should another district-industry factor jointly determine both incumbent structures and entry rates. We will use changes in industrial conditions from 1989 to 2005 to partially address this concern.¹²

¹¹ The agglomeration of specialized workers and firms can occur through several channels. Marshall (1920) described how an agglomeration of workers and firms shields workers from firm-specific shocks. Workers can be more productive and better insured by moving from firms that are hit with negative shocks to better opportunities (e.g., Diamond and Simon 1990, Krugman 1991, Overman and Puga 2010). Larger labor pools further promote more efficient matches (e.g., Helsley and Strange 1990), and multiple firms protect workers against ex post appropriation of investments in human capital (e.g., Rotemberg and Saloner 2000). All of these mechanisms suggest that firms that employ similar types of workers will tend to locate near one another and that entrants will benefit from thick local markets for their specific labor needs, either through heightened availability or lower wages.

¹² There are several factors that we do not consider in this study: natural cost advantages (e.g., Ellison and Glaeser 1999), local industrial diversity (Jacobs 1970), and entrepreneurial culture (Hofstede 2001, Florida 2005).

Unconditional Estimations of Spatial Entrepreneurship

We characterize entry through a series of linear regressions with the above determinants as explanatory variables:

$$\ln(\text{Entry}_{di}) = \eta_i + \beta \cdot X_d + \gamma \cdot \ln(\text{Incumbent Employment}_{di}) + \gamma_I \cdot \text{Input}_{di} + \gamma_O \cdot \text{Output}_{di} \\ + \gamma_L \cdot \text{Labor}_{di} + \gamma_C \cdot \text{Chinitz}_{di} + \varepsilon_{di}.$$

We include in each estimation a vector of industry fixed effects η_i that control for fixed differences in industry sizes, entrepreneurship rates, competition, and so on. X_d is the vector of district traits like population and education levels. We further control for incumbent employment in the district-industry. Our main specifications are also robust to controlling for incumbent firm counts or value added rather than employment. Finally, we include our agglomeration metrics that vary by district-industry. We transform non-log variables to have unit standard deviation to aid interpretation, and we cluster standard errors by district to reflect the multiple mappings of some variables.

The dependent variable is the log measure of entry employment by district-industry. We recode less than one entering employee on average as one entering employee for these estimations. This maintains a consistent sample size, and we do not believe that the distinction between zero employees and one employee at the district-industry level is economically meaningful. Regardless, these cells can be excluded without impacting our results.

We weight estimations by an interaction of log industry size with log district population. We place more faith in weighted estimations than unweighted estimations since many district-industry observations experience very limited activity. We recognize, however, that weighted estimations may accentuate endogeneity concerns. We thus employ our interaction rather than observed district-industry size. The interaction minimizes the endogeneity spillover for very agglomerated industries, especially in conditional estimations with district and industry fixed effects. We find very similar effects without sample weights, indicating that these choices are not very material.

Perhaps most important, we also do not model Marshall's (1920) third agglomeration rationale of knowledge spillovers because the data used to typically measure knowledge spillovers (e.g., local patent citations) are not available consistently at the level of detail that we require in India. It should be noted that the customer/supplier and labor pooling rationales both overlap with knowledge flows to some extent. For example, Porter (1990) emphasizes that proximity to customers and suppliers can enhance innovation by increasing knowledge flows about which products are working and what new products are desired. Likewise, occupational sharing is often associated with the knowledge and skills of the workers. To some degree, knowledge flows will be included in these estimates.

Table 4A provides our basic spatial results for the organized manufacturing sector. The first column includes just district populations, district-industry employments, and industry fixed effects. Not surprisingly, existing district-industry employment strongly shapes the spatial location of entry: a 10% increase in incumbent employment raises entry employment by around 2%. In addition, a district's population increases entry rates with an elasticity of 0.5. Higher-order population terms are not found to be statistically significant or economically important. The adjusted R-Squared value for this estimation is quite modest at 0.13.

Glaeser and Kerr (2009) estimate a related specification for the United States that uses log long-term employment for a city-industry as the key explanatory variable. If we adjust our estimation to more closely match their technique by using log total employment as the explanatory variable, combining young establishments and incumbents, we obtain an elasticity of 0.8 that is very similar to the 0.7 elasticity measured by Glaeser and Kerr (2009). While this elasticity is comparable, the adjusted R-Squared value for this estimation remains quite modest at 0.29. This value is much lower than the adjusted R-Squared value of 0.80 for Glaeser and Kerr (2009).

There are likely several factors behind this lower explanatory power, including data differences, estimations at the district versus city level, and similar. But, it is also clear that many industries within India's manufacturing sector are at a much earlier development stage than in the United States, where the manufacturing sector is instead shrinking. Thus, while existing patterns of industrial activity explain the spatial distribution of entrepreneurship in India similar to the United States, India has much more variation in outcomes that we characterize further below. Fernandes and Sharma (2011) also study these variations with respect to policy deregulations.

Column 2 includes the district traits. Three factors stand out as discouraging entrepreneurship in organized manufacturing: high population density, strict labor regulations, and greater distance to one of India's ten biggest cities. The first pattern has been observed in many settings, and is closely studied by Desmet et al. (2011) in India. The traded nature of manufacturing products allows more rural settings for firms, and manufacturers often seek cheaper environments than the wages and rents associated with high density areas. The second pattern connects with the earlier studies of India that argue strict labor laws reduce economic growth. These policies are associated with reduced entry even after conditioning on district-industry size. The final factor highlights that while manufacturers avoid the high costs of urban areas, they also avoid the most remote areas of India in favor of settings that are relatively near large population centers, likely to access customers directly or to connect to shipping routes. On the other hand, the education of a district's workforce is strongly linked to higher entry rates. The elasticity is in fact stronger in economic magnitude, if not precision, than that evident in comparable studies of advanced economies.

The third column introduces the district-industry traits. The qualities of input and output markets are exceptionally strong with 0.4-0.5 elasticities. Labor market and the Chinitz measure have positive coefficients but are not precisely measured. We will return to the interpretation of these results after viewing the conditional estimations in Table 6 and thereafter. The decline in the main effect of incumbent employment suggests that these four new metrics capture the positive channels of agglomeration on entry. The last column shows quite similar results if we use the log count of entrant establishments rather than log entry employment, although the roles for education and labor market agglomeration economies are reduced. The Chinitz metric is more prominent when using entry counts.

Across these columns of Table 4, the adjusted R-Squared value increases from 0.13 to approximately 0.2-0.3. While perhaps modest in overall size, this growth in explanatory power is much stronger than the similar growth for entry levels in Glaeser and Kerr (2009) with the additional factors. This pattern again highlights the greater relative importance in India of existing district conditions relative to incumbent positioning for explaining entrepreneurship compared to patterns in the United States.

Table 4B repeats these estimations for the unorganized manufacturing sector. Several distinct differences exist. First, local population takes a much greater role with unit elasticity in Column 1's simplest estimation. This greater connection of entry to the overall size of local markets almost certainly reflects unorganized entry being proportionate to market size and servicing local needs. Evidence for this relationship is also evident in the independence of entry to local population density or travel time to a major city, the stronger relationship of entry to the age profile of the district, and the higher R-Squared value in Columns 1 and 2 of Table 4B compared to Table 4A. Unorganized manufacturing clearly conforms much more closely to the overall contours of India's economic geography than organized manufacturing.

The other two district traits that are associated with strong entry rates are the strength of local, within-district physical infrastructure and the strength of local household banking environments. This contrasts with organized manufacturing entry, where education stood out. An intuitive explanation, which will also be reflected in the services estimations, is that these patterns and their differences reflect the factors on which each sector depends most. Organized manufacturing establishments, for example, may have broader resources that reduce dependency on local infrastructure and household finance. Likewise, it is reasonable to believe that the unorganized sector depends less on educated workers than the organized sector. While intuitive, we are unable to rigorously confirm these district-level observations in this study, and these results should be viewed as partial correlations.

We again find evidence for agglomeration economies within the unorganized manufacturing sector. The framework is similar to Table 4A except that we do not consider the

Chinitz effect since by definition the unorganized sector is comprised of small firms. Partly as a consequence of this, the inputs metric is relatively stronger in these estimations. Mukim (2011) also finds an important role for input conditions in the informal sector. The final column shows that the unorganized sector results are very stable with the change in outcome variable. The initial gap in explanatory power between the organized and unorganized sectors that was evident in the first two columns is mostly gone by the complete estimations in Columns 3 and 4.

Table 5 presents comparable estimations for the services sector, with Columns 1-3 for the organized sector and Columns 4-6 for the unorganized sector. The patterns and their contrast to organized manufacturing are again quite intriguing. First, overall district population is as important as it was for unorganized manufacturing, with its elasticity greater than one. Similarly, the adjusted R-Squared value grows to 0.20 and 0.47 with just the parsimonious set of explanatory factors in Columns 1 and 4, respectively. The adjusted R-Squared value using the Glaeser and Kerr (2009) approach for organized services is 0.30. Similar also to unorganized manufacturing, population density and travel time to major cities are not important in the multivariate setting, while the district's age profile does contribute to higher entry levels.

Among district traits, education and infrastructure matter the most. Overall, education is found to be generally important, with particular relevance to the organized sectors of manufacturing and services (Amin and Mattoo 2008). Physical infrastructure is also generally important, with particular relevance to the unorganized sectors of the economy. The strength of the household banking sector is again also very important in the unorganized sectors of the economy. These channels provide three of the main ways that policy makers can influence the spatial distribution of entry.

The role of the existing incumbent employment by district-industry for services is weak in Table 5, likely suggesting that Marshallian economies are weaker in services (using our broad industry groups) than in manufacturing. Unreported estimations further attempted to model Marshallian interactions in the services sector similar to manufacturing. These results are also weak, at most suggesting a small role for labor market interactions. However, we hesitate to strongly interpret this difference as the weak results may be due to applying concepts and metrics originally designed for manufacturers to the service sector. We hope in future work to examine local spillovers in services among modern services firms, especially those involved in high-tech and ICT sectors, to identify if deeper agglomeration forces operate within these firms (e.g., Arzaghi and Henderson 2008).

Conditional Estimations of Spatial Entrepreneurship

We now turn to conditional estimations that focus just on district-industry variation. Table 6 next estimates a conditional specification of the form:

$$\ln(\text{Entry}_{di}) = \eta_i + \delta_d + \gamma \cdot \ln(\text{Incumbent Employment}_{di}) + \gamma_I \cdot \text{Input}_{di} + \gamma_O \cdot \text{Output}_{di} \\ + \gamma_L \cdot \text{Labor}_{di} + \gamma_C \cdot \text{Chinitz}_{di} + \varepsilon_{di}.$$

We now include a vector of district fixed effects δ_d that control for differences across districts that are common for all industries, for example Delhi's larger size. Specifications thus employ variation within districts and industries: How much of the unexplained district-industry variation in entrepreneurship can we explain through incumbent local conditions that are especially suitable for particular industries?

The first three columns are for the organized manufacturing sector, while the last two columns are for the unorganized manufacturing sector. We report robust standard errors reflecting the district-industry variation. The coefficient patterns in Table 4, compared to Tables 4A and 4B, are stronger and more precisely estimated. These results suggest that many of the agglomeration rationales that we discussed in Section 3 operate as strongly for entrants in India as they do in advanced economies, or perhaps even more strongly. We interpret the weaker performance in the earlier unconditional estimates, compared to the conditional estimates that fully control for district averages, as evidence that India's economic geography is not set to the degree that an advanced economy is.

In addition to these conditional tests of our agglomeration metrics, unreported estimations confirm some of the district traits analyzed through interaction regressions. After including the district effects, we can no longer estimate the direct impact of labor laws on entry rates, but we can estimate an interaction of labor laws with how important labor is as an input factor for an industry. We estimate the latter importance through the industry's wage bill divided by industry value added. This interaction is negative and statistically significant, indicating that entrepreneurship in labor-intensive sectors is disproportionately reduced by strict labor laws. We likewise find that the Chinitz effect and local input conditions matter more in materially intensive industries.¹³

Table 7 examines the entrant size distribution for the organized sector by separating our overall entry measures into establishment sizes of young firms. The entry of a ten-person

¹³ As another robustness check on our metric design, we find similar results when winsorizing our metrics at their 5% and 95% levels to weaken the influence of extreme values. The main adjustment is that the output metric grows in relative importance.

establishment is presumably a different phenomenon than the entry of a new firm with hundreds of employees. We care more about larger entrants in certain contexts, for example when worrying about the determinants of robust local labor demand. On the other hand, the entry of small establishments may be a purer reflection of entrepreneurship and hence more intrinsically interesting. More generally, empirical evidence exists that small and large establishments agglomerate differently (e.g., Holmes and Stevens 2002, Duranton and Overman 2008), and it is useful to extend this description to entering firms.¹⁴

Table 7 finds interesting distributional effects that also provide intuitive confirmation of the economic forces proposed. Most strikingly, the importance of the Chinitz effect is concentrated among small entrants, while the importance of overall output markets and labor spillovers grow with entrant size. For India, it appears that input cost factors are more influential in the location choices of small start-ups, while output conditions and labor markets are more important for large entrants.

Table 8 contains our final set of empirical results. Our work thus far has focused on the cross-sectional patterns of incumbent industrial structures and entry. By including district and industry fixed effects, we focus on within-district and within-industry variation for analysis. This approach thus guards against omitted factors that vary by district or by industry. Similarly, our focus on incumbent firms to explain new entrants mirrors Jofre-Monseny et al. (2011), taking the former to be pre-determined. As an alternative, Glaeser and Kerr (2009) use predicted spatial distributions of industries due to natural cost advantages to provide a measure of exogeneity.

Nonetheless, a concern persists that there are unique aspects of district-industries that may confound this relationship. To take a United States example, the automobile industry has been concentrated in Detroit for over a century. Over this span of time, localized entrepreneurship and incumbent industrial structures will have jointly influenced each other, and many other factors that we do not model may have arisen (e.g., special political connections and support by Detroit for the automobile industry). These latter factors that are particular to an industrial cluster would not be captured by city and industry fixed effects, and yet these instances of highly agglomerated activity are very important for identification in the above estimations. The long history of the Indian government's involvement in local industrial policy accentuates these econometric concerns for our estimates.

¹⁴ A second rationale exists for examining the entry size distribution. Better local conditions may foster a larger entry size for entrepreneurs due to factors like less uncertainty about local growth potential and faster assembly of key resources. As discussed in Kerr and Nanda (2009, 2010), however, metrics of average entry size confound this intensive margin adjustment with changes in the extensive margin of greater entry rates. Better local conditions may simultaneously foster greater entry by many small firms, which leads to an overall decline in average entry size. We feel it is more prudent to look at the distribution measure.

One approach to help address these concerns is to use time-varying conditions in localized agglomeration and entry by district-industry. By looking across two points in time, district-industry fixed effects can be included in the estimations. These fixed effects control for long-run levels of incumbent industrial structures and entry, focusing on changes within each district-industry. Such an approach does not fully overcome potential biases, as there could be time-varying factors within district-industries that continue to confound the analysis. The empirical bar, however, is set much higher.

A challenge to implementing this approach in many settings is that industrial structures can be very stable over time, providing little variation to exploit. India's organized manufacturing setting provides a unique opportunity in this regard. Prior to the large-scale deregulations, spatial location decisions for firms were set to a large degree by the government, with the goal to promote general equality across regions. In the two decades since these restrictions were lifted, India's manufacturing has seen large changes in spatial locations and agglomerations (e.g., Fernandes and Sharma 2011).

These changes provide much greater longitudinal variation than could typically be exploited. Micro-data for India's organized manufacturing sector extend back to 1989. We prepare our metrics for 1989 similar to those used in 2005. We restrict our sample to district-industry observations present in both periods. Table 8 estimates a panel specification of the form:

$$\ln(\text{Entry}_{\text{dit}}) = \eta_{\text{it}} + \delta_{\text{dt}} + \pi_{\text{di}} + \gamma \cdot \ln(\text{Incumbent Employment}_{\text{dit}}) + \gamma_{\text{I}} \cdot \text{Input}_{\text{dit}} + \gamma_{\text{O}} \cdot \text{Output}_{\text{dit}} + \gamma_{\text{L}} \cdot \text{Labor}_{\text{dit}} + \varepsilon_{\text{dit}}.$$

We now include a vector of district-industry fixed effects π_{di} that control for fixed differences across district-industries; we also extend our earlier fixed effects to be district-year and industry-year controls. These specifications thus employ panel variation: how much of the growth in district-industry entrepreneurship can we explain through changes in incumbent local conditions that are especially suitable for particular industries? By including district-year and industry-year fixed effects, we measure this effect after controlling for general district and industry development between 1989 and 2005.

Table 8 provides strong confirmation for our basic findings. In the first column, growth in general incumbent employment over the 16 years is linked to higher entrepreneurship. The elasticity is half the size estimated in the cross-section work. In the second column, we also find support for the Marshallian metrics related to input and output markets. The coefficients are slightly larger than in the cross-sectional estimations and precisely estimated. Interestingly, labor

conditions do not find support in the panel setting.¹⁵ Changes in the Chinitz metric yielded implausibly large coefficients values due to outliers, and we do not report them. Of our metrics, the Chinitz effect is the most sensitive due to how it embodies both the establishment size distribution and input-output exchanges. Its sensitivity is thus not very surprising.

Overall, the panel estimations strongly support our core evidence on the link between entrepreneurship and local industrial conditions in India. We think that India's industrial past, and the government-led spatial allocation of industrial activity that is rapidly becoming undone, provides a very interesting laboratory for testing many features of agglomeration and urban economics that are difficult to disentangle in advanced economies with more stable economic geographies.

Conclusions

Entrepreneurship is vital to economic growth. While India has historically had low entrepreneurship rates, this weakness is improving and will be an important stepping stone to further development. This paper explores the spatial determinants of local entrepreneurship in India for both manufacturing and services. At the district level, our strongest evidence points to the roles that local education levels and physical infrastructure quality play in promoting entry. We also find evidence that strict labor regulations discourage formal sector entry, and better household banking environments encourage entry in the unorganized sector. Policy makers wishing to encourage entrepreneurship in their local areas have several policy levers that can be exploited.

Looking more closely at district-industry activity, we find strong evidence of agglomeration economies in India's manufacturing sector. This work especially emphasizes the input-output relationships among firms. This evidence on localized agglomeration economies and entry is the first in a developing economy of which we are aware. This framework and its comparison to advanced economies also points out that the economic geography of India is still taking shape. India's economic geography is still adjusting from the government-imposed conditions that existed pre-deregulation, and much greater variation exists in spatial outcomes than is present in countries like the United States. This raises the importance of correct policy design for local areas, and it provides a nice testing ground for future work on agglomeration and urban economies.

¹⁵ One potential factor that may contribute to the greater sensitivity of the labor metric compared to input-output measures is that the labor metric builds upon worker flows in the United States which may differ substantially from India. The input-output metrics, by contrast, directly build upon India's input-output tables.

References

- Acs, Zoltan, and Catherine Armington, *Entrepreneurship, Geography and American Economic Growth* (New York, NY: Cambridge University Press, 2006).
- Aghion, Philippe, Robin Burgess, Stephen Redding, and Fabrizio Zilibotti, “The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India”, *American Economic Review* 98 (2008), 1397-1412.
- Aghion, Philippe, Ufuk Akcigit, Julia Cage, and William Kerr, “Taxation, Corruption, and Growth”, Working Paper (2011).
- Agrawal, Ajay, Devesh Kapur, John McHale, and Alex Oettl, “Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation”, *Journal of Urban Economics* 69:1 (2011), 43-55.
- Ahsan, Ahmed, and Carmen Pages, “Are All Labor Regulations Equal? Assessing The Effects Of Job Security, Labor Dispute And Contract Labor Laws In India”, World Bank Policy Research Working Paper 4259 (2007).
- Alcacer, Juan, and Wilburg Chung, “Location Strategies for Agglomeration Economies”, HBS Working Paper (2010).
- Amin, Mohammad, and Aaditya Mattoo, “Human Capital and the Changing Structure of the Indian Economy”, World Bank Policy Research Working Paper 4576 (2008).
- Ardagna, Silvia, and Annamaria Lusardi, “Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints”, Working Paper (2008).
- Arora, Ashish, and Alfonso Gambardella, eds., *From Underdogs to Tigers? The Rise and Growth of the Software Industry in Brazil, China, India, Ireland, and Israel* (Oxford, UK: Oxford University Press, 2005).
- Arzaghi, Mohammad, and J. Vernon Henderson, “Networking off Madison Avenue”, *Review of Economic Studies* 75:4 (2008), 1011-1038.
- Astebro, Thomas, Jing Chen, and Peter Thompson, “Stars and Misfits: Self-Employment and Labor Market Frictions”, Working Paper (2010).
- Audretsch, David, and Maryann Feldman, “R&D Spillovers and the Geography of Innovation and Production”, *American Economic Review* 86 (1996), 630-640.
- Bartik, Timothy, “Business Location Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States”, *Journal of Business & Economic Statistics* 3:1 (1985), 14-22.
- Besley, Timothy, and Robin Burgess, “Can Labor Regulation Hinder Economic Performance? Evidence from India”, *Quarterly Journal of Economics* 119 (2004), 91-134.
- Blanchflower, David, and Andrew Oswald, “What Makes an Entrepreneur?”, *Journal of Labor Economics* 16 (1998), 26-60.

- Bönte, Werner, Oliver Falck and Stephan Heblich, “The Impact of Regional Age Structure on Entrepreneurship”, *Economic Geography* 85 (2009), 269-287.
- Bozkaya, Ant, and William Kerr, "Labor Regulations and European Venture Capital", HBS Working Paper (2011).
- Cantillon, Richard, *Essai sur la Nature du Commerce en Général* (1730).
- Carlino, Gerald, Satyajit Chatterjee, and Robert Hunt, “Urban Density and the Rate of Invention”, *Journal of Urban Economics* 61 (2007), 389-419.
- Carlton, Dennis, “The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables”, *Review of Economics and Statistics* 65:3 (1983), 440-449.
- Chari, A.V., “The Aggregate Productivity Effects of Entry and Output Restrictions: An Analysis of License Reform in India”, Working Paper (2008).
- Chinitz, Benjamin, “Contrasts in Agglomeration: New York and Pittsburgh”, *American Economic Review* 51:2 (1961), 279-289.
- Combes, Pierre-Philippe, and Gilles Duranton, “Labour Pooling, Labour Poaching, and Spatial Clustering”, *Regional Science and Urban Economics* 36 (2006), 1-28.
- Dahl, Michael, and Olav Sorenson, “Home Sweet Home: Social Capital and Location Choice”, Working Paper (2007).
- Dauth, Wolfgang, “The Mysteries of the Trade: Interindustry Spillovers in Cities”, Working Paper (2011).
- Dehejia, Rajeev, and Arvind Panagariya, “Services Growth in India: A Look Inside the Black Box”, Working Paper (2010).
- Deichmann, Uwe, Somik Lall, Stephen Redding, and Anthony Venables, “Industrial Location in Developing Countries”, *World Bank Research Observer* 23:2 (2008), 219-246.
- Delgado, Mercedes, Michael Porter, and Scott Stern, “Clusters and Entrepreneurship”, *Journal of Economic Geography* 10:4 (2010a), 495-518.
- Delgado, Mercedes, Michael Porter, and Scott Stern, “Clusters, Convergence, and Economic Performance”, Working Paper (2010b).
- Desmet, Klaus, Ejaz Ghani, Stephen O’Connell, and Esteban Rossi-Hansberg, “The Spatial Development of India”, Working Paper (2011).
- Diamond, Charles, and Chris Simon, “Industrial Specialization and the Returns to Labor”, *Journal of Labor Economics* 8 (1990), 175-201.
- Doms, Mark, Ethan Lewis, and Alicia Robb, “Local Labor Force Education, New Business Characteristics, and Firm Performance”, *Journal of Urban Economics* 67:1 (2010).
- Drucker, Joshua, and Edward Feser, “Regional Industrial Dominance, Agglomeration Economies, and Manufacturing Plant Productivity”, CES Working Paper 07-31 (2007).

- Due, John, "Studies of State-Local Tax Influences on Location of Industry", *National Tax Journal* (1961).
- Duranton, Gilles, and Henry Overman, "Testing for Localization Using Micro-Geographic Data", *Review of Economic Studies* 72 (2005), 1077-1106.
- Duranton, Gilles, and Henry Overman, "Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data", *Journal of Regional Science* 48:1 (2008), 313-343.
- Duranton, Gilles, and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies", in Vernon Henderson and Jacques François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland, 2004), 2063–2117.
- Ellison, Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105 (1997), 889-927.
- Ellison, Glenn, and Edward Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?", *American Economic Review Papers and Proceedings* 89 (1999), 311-316.
- Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review* 100 (2010), 1195-1213.
- Evans, David, and Boyan Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints", *Journal of Political Economy* 97:4 (1989), 808-827.
- Evans, David, and Linda Leighton, "Some Empirical Aspects of Entrepreneurship", *American Economic Review* 79 (1989), 519-535.
- Fallick, Bruce, Charles Fleischman, and James Rebitzer, "Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster", *Review of Economics and Statistics* 88:3 (2006), 472-481.
- Feldman, Maryann, "The Locational Dynamics of the US Biotech Industry: Knowledge Externalities and the Anchor Hypothesis", *Industry and Innovation* 10:3 (2003), 311-328.
- Fernandes, Ana, and Ariel Pakes, "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Survey Data", *Working Paper* (2010).
- Fernandes, Ana, and Gunjan Sharma, "Together We Stand? Agglomeration in Indian Manufacturing", *Working Paper* (2011).
- Figueiredo, Octávio, Paulo Guimaraes, and Douglas Woodward, "Home-Field Advantage: Location Decisions of Portuguese Entrepreneurs", *Journal of Urban Economics* 52:2 (2002), 341-361.
- Florida, Richard, *Cities and the Creative Class* (New York, NY: Routledge, 2005).
- Fujita, Masahisa, Paul Krugman, and Anthony Venables, *The Spatial Economy: Cities, Regions and International Trade* (Cambridge, MA: MIT Press, 1999).
- Ghani, Ejaz (ed.) *The Services Revolution* (Oxford, UK: Oxford University Press, 2010).

- Ghani, Ejaz, William Kerr, and Stephen O'Connell, "Promoting Entrepreneurship, Growth, and Job Creation" in Ejaz Ghani (ed.) *Reshaping Tomorrow* (Oxford, UK: Oxford University Press, 2011).
- Glaeser, Edward, "Entrepreneurship and the City", NBER Working Paper 13551 (2007).
- Glaeser, Edward, and William Kerr, "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?", *Journal of Economics and Management Strategy* 18:3 (2009), 623-663.
- Glaeser, Edward, and William Kerr, "Entrepreneurship, Mines, and Urban Growth", Working Paper (2011).
- Glaeser, Edward, William Kerr, and Giacomo Ponzetto, "Clusters of Entrepreneurship", *Journal of Urban Economics* 67:1 (2010), 150-168.
- Government of India, Input Flow Matrix 2006-07 at Factor Cost, from IOTT 2006-07 – Central Statistical Organization (CSO), Ministry of Statistics & Programme Implementation, Government of India. (2007).
- Greenstone, Michael, Richard Hornbeck and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings", *Journal of Political Economy* 118 (2010), 536-598.
- Haltiwanger, John, Ron Jarmin, and Javier Miranda, "Who Creates Jobs? Small vs. Large vs. Young", NBER Working Paper 16300 (2010).
- Hasan, Rana, and Karl Jandoc, "The Distribution of Firm Size in India: What Can Survey Data Tell Us?", ADB Economics Working Paper 213 (2010).
- Henderson, J. Vernon, "Marshall's Scale Economies", *Journal of Urban Economics* 53 (2003), 1-28.
- Helsley, Robert, and William Strange, "Coagglomeration", Working Paper (2010).
- Helsley, Robert, and William Strange, "Matching and Agglomeration Economics in a System of Cities", *Regional Science and Urban Economics* 20 (1990), 189-212.
- Hofstede, Geert, *Culture and Organizations* (London, UK: Harper Collins, 2001).
- Holmes, Thomas, and John Stevens, "Geographic Concentration and Establishment Scale", *Review of Economics and Statistics* 84:4 (2002), 682-690.
- Hsieh, Chang-Tai, and Peter Klenow, "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics* 124 (2009), 1403-1448.
- Iyer, Lakshmi, Tarun Khanna, and Ashutosh Varshney, "Caste and Entrepreneurship in India", Working Paper (2011).
- Jacobs, Jane, *The Economy of Cities* (New York, NY: Vintage Books, 1970).

- Jofre-Monseny, Jordi, Raquel Marín-López, and Elisabet Viladecans-Marsal, “The Mechanisms of Agglomeration: Evidence from the Effect of Inter-Industry Relations on the Location of New Firms”, *Journal of Urban Economics*, forthcoming (2011).
- Kanbur, Ravi, “Avoid Informality Traps”, in Ejaz Ghani (ed.) *Reshaping Tomorrow* (Oxford, UK: Oxford University Press, 2011).
- Kathuria, Vinish, Seethamma Natarajan, Rajesh Raj, and Kunal Sen, “Organized versus Unorganized Manufacturing Performance in India in the Post-Reform Period”, MPRA Working Paper No. 20317 (2010).
- Kerr, William, “Ethnic Scientific Communities and International Technology Diffusion”, *Review of Economics and Statistics* 90 (2008), 518-537.
- Kerr, William, and Scott Kominers, “Agglomerative Forces and Cluster Shapes”, NBER Working Paper 16639 (2010).
- Kerr, William, and Ramana Nanda, “Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship”, *Journal of Financial Economics* 94 (2009), 124-149.
- Kerr, William, and Ramana Nanda, “Banking Deregulations, Financing Constraints and Firm Entry Size”, *Journal of the European Economic Association* 8:2-3 (2010), 582–593.
- Kerr, William, and Ramana Nanda, “Financing Constraints and Entrepreneurship,” in Audretsch, David, Oliver Falck, and Stephan Heblich, eds., *Handbook on Research on Innovation and Entrepreneurship* (London, UK: Edward Elgar, 2011).
- Khanna, Tarun, *Billions of Entrepreneurs: How China and India Are Reshaping Their Futures—and Yours* (Boston, MA: Harvard University Press, 2008).
- Klapper, Leora, Raphael Amit, and Mauro F. Guillén, “Entrepreneurship and Firm Formation across Countries”, in Lerner, Josh, and Antoinette Schoar, eds., *International Differences in Entrepreneurship* (Chicago, IL: University of Chicago Press, 2010).
- Klepper, Steven, “The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit”, *Journal of Urban Economics* 67 (2010), 15-32.
- Krugman, Paul, *Geography and Trade* (Cambridge, MA: MIT Press, 1991).
- Lall, Somik, Hyoung Wang, and Uwe Deichmann, “Infrastructure and City Competitiveness in India”, Working Paper (2011).
- Marshall, Alfred, *Principles of Economics* (London, UK: MacMillan and Co., 1920).
- McConnell, Virginia, and Robert Schwab, “The Impact of Environmental Regulation on Industry Location Decisions: The Motor Vehicle Industry”, *Land Economics* 66:1 (1990), 67-81.
- Michelacci, Claudio, and Olmo Silva, “Why So Many Local Entrepreneurs?”, *Review of Economics and Statistics* 89:4 (2007), 615-633.
- Mukim, Megha, “Industry and the Urge to Cluster: A Study of the Informal Sector in India”, Working Paper (2011).

- Mukim, Megha, and Peter Nunnenkamp, "The Location Choices of Foreign Investors: A District-Level Analysis in India", *World Economy* forthcoming (2011).
- Nanda, Ramana, and Tarun Khanna, "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry", *Journal of Economics and Management Strategy* 19:4 (2010), 991-1012.
- Nataraj, Shanthi, "The Impact of Trade Liberalization on Productivity and Firm Size: Evidence from India's Formal and Informal Manufacturing Sectors", Working Paper (2009).
- Oakland, W.H., "Local Taxes and Intra-urban Industrial Location: A Survey", in *Metropolitan Finance and Growth Management Policies* (Madison, WI: University of Wisconsin Press, 1978).
- Overman, Henry, and Diego Puga, "Labor Pooling as a Source of Agglomeration: An Empirical Investigation", in Glaeser, Edward (ed.) *Agglomeration Economics* (Chicago, IL: University of Chicago Press, 2010).
- Parker, Simon, *The Economics of Entrepreneurship* (Cambridge, UK: Cambridge University Press, 2009).
- Porter, Michael, *The Competitive Advantage of Nations* (New York, NY: The Free Press, 1990).
- Rosenthal, Stuart, and William Strange, "Evidence on the Nature and Sources of Agglomeration Economies", in Vernon Henderson and Jacques François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland, 2004), 2119-2171.
- Rosenthal, Stuart, and William Strange, "Geography, Industrial Organization, and Agglomeration", *Review of Economics and Statistics* 85 (2003), 377-393.
- Rosenthal, Stuart, and William Strange, "Small Establishments/Big Effects: Agglomeration, Industrial Organization and Entrepreneurship", in Glaeser, Edward (ed.) *Agglomeration Economics* (Chicago, IL: University of Chicago Press, 2010).
- Rosenthal, Stuart, and William Strange, "The Attenuation of Human Capital Spillovers", *Journal of Urban Economics* 64 (2008), 373-389.
- Rosenthal, Stuart, and William Strange, "The Determinants of Agglomeration", *Journal of Urban Economics* 50 (2001), 191-229.
- Rotemberg, Julio, and Garth Saloner, "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade", *Regional Science and Urban Economics* 30 (2000), 373-404.
- Samila, Sampsa and Olav Sorenson, "Venture Capital, Entrepreneurship, and Economic Growth", *Review of Economics and Statistics* 93 (2011), 338-349.
- Saxenian, AnnaLee, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Cambridge, MA: Harvard University Press, 1994).
- Schmenner, Roger, *Making Business Location Decisions* (Englewood Cliffs, N.J.: Prentice-Hall, 1982).

- Schoar, Antoinette, “The Divide Between Subsistence and Transformational Entrepreneurship”, in Lerner, Josh, and Scott Stern (eds.) *Innovation Policy and the Economy*, Number 10 (Chicago, IL: University of Chicago Press, 2009), 57-71.
- Zucker, Lynne, Michael Darby, and Marilyn Brewer, “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises”, *American Economic Review* 88 (1998), 290-306.

Fig. 1: Business Registration Density, 2008

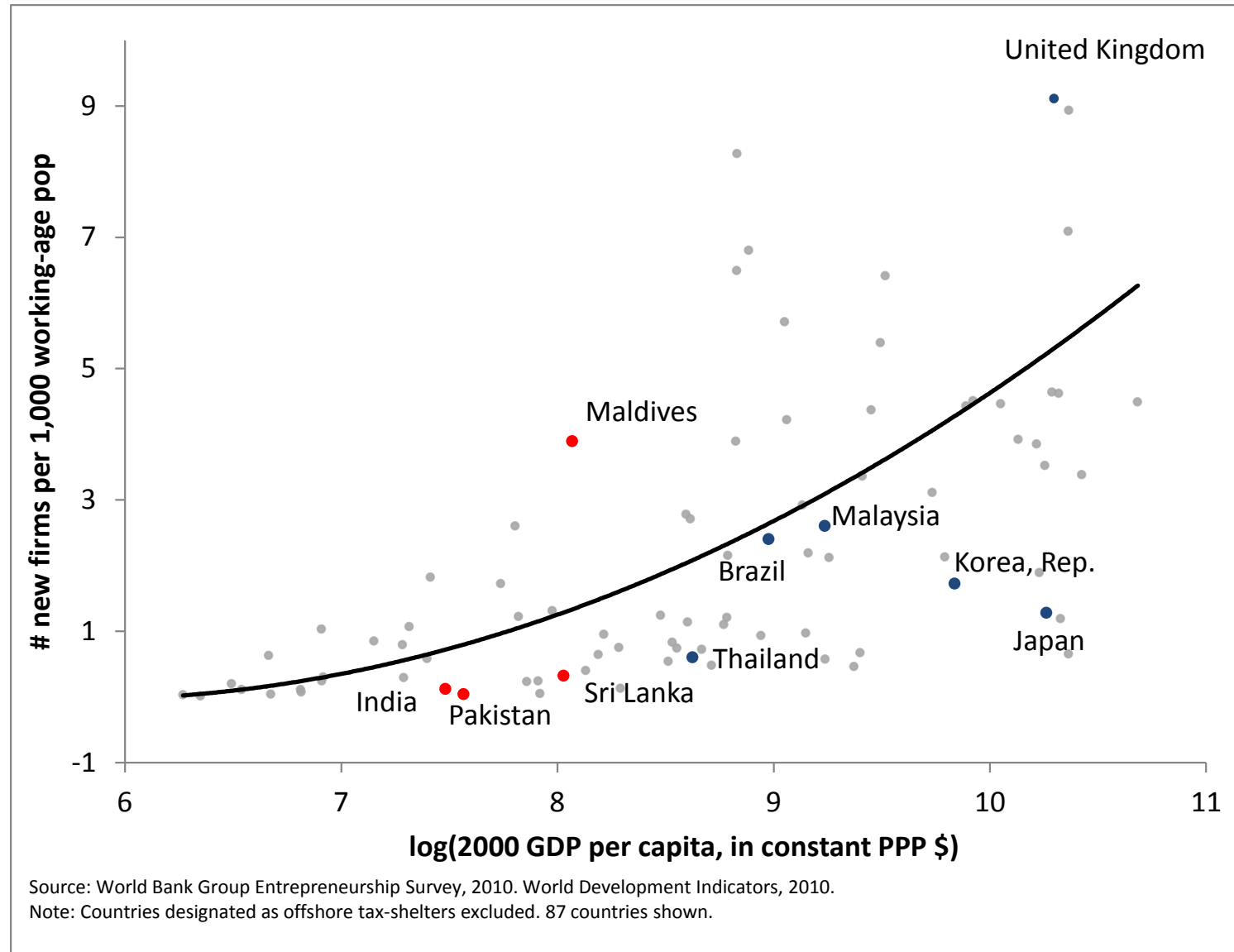


Fig. 2: Indian Manufacturing Entry Rates

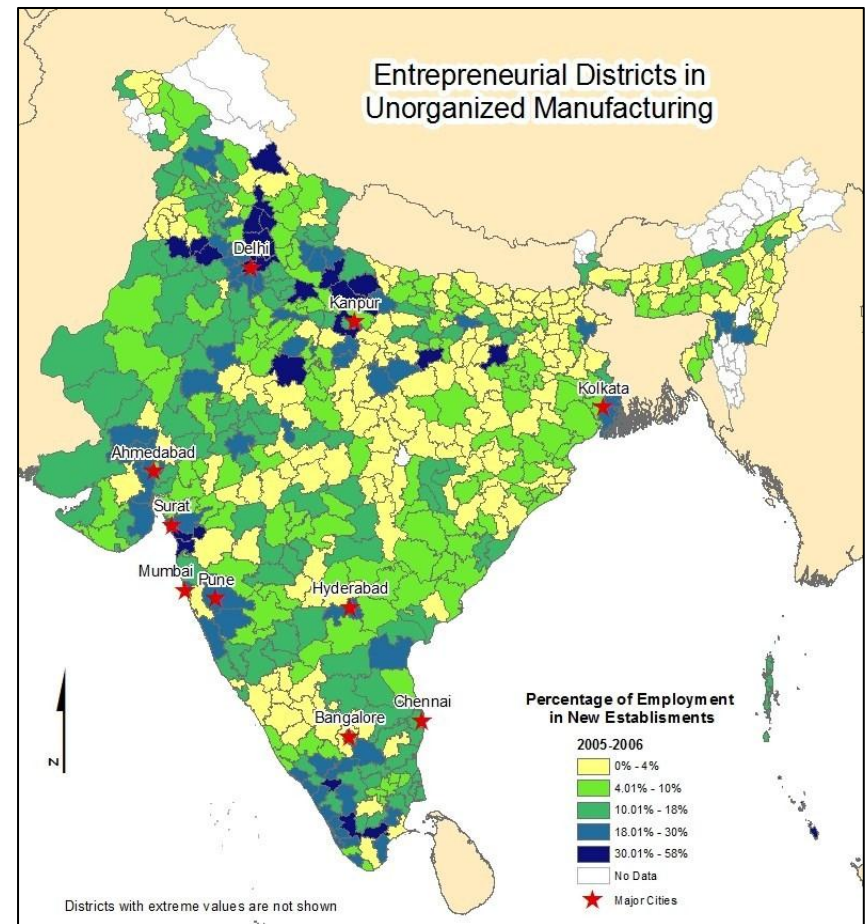
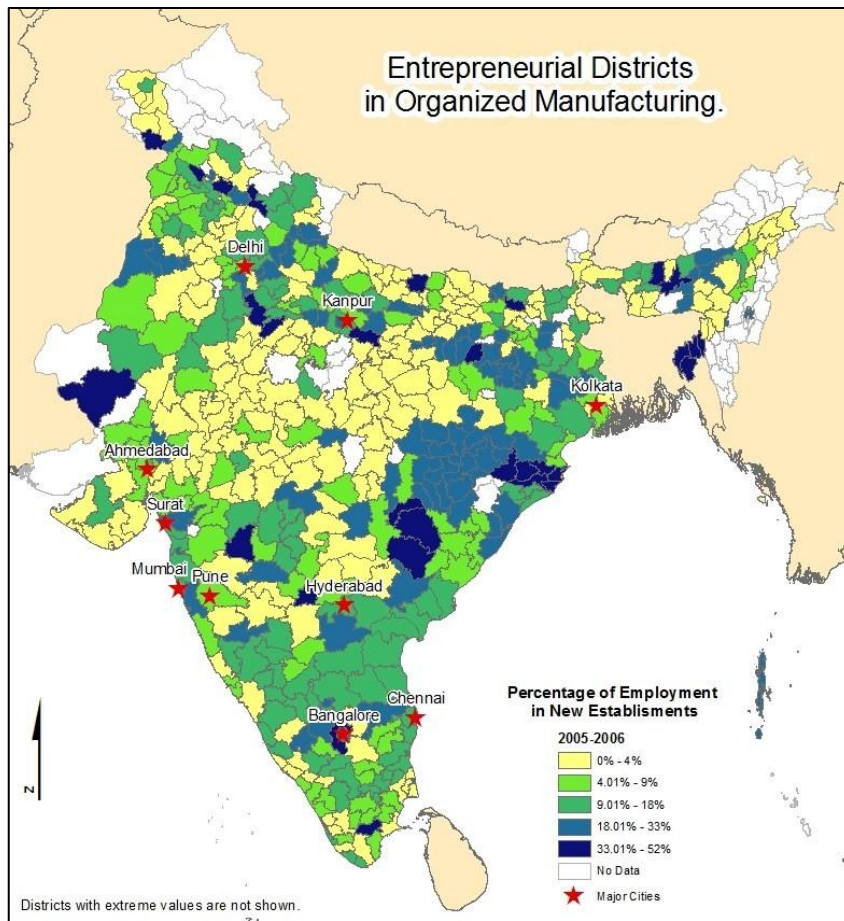


Fig. 3: Indian Services Entry Rates

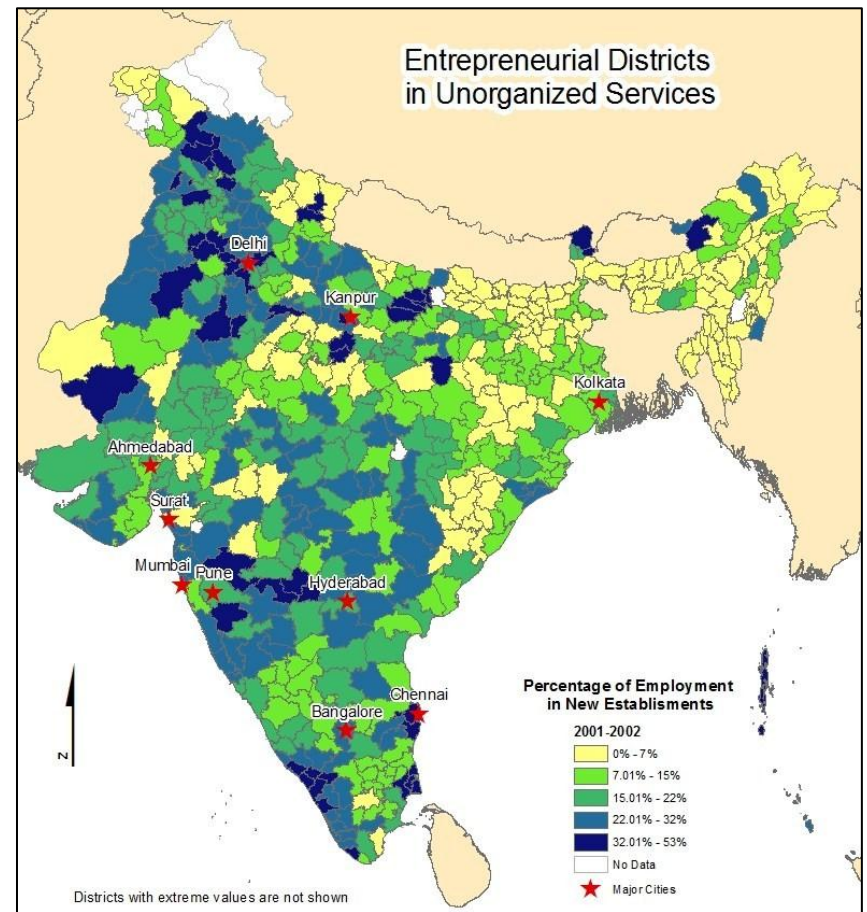
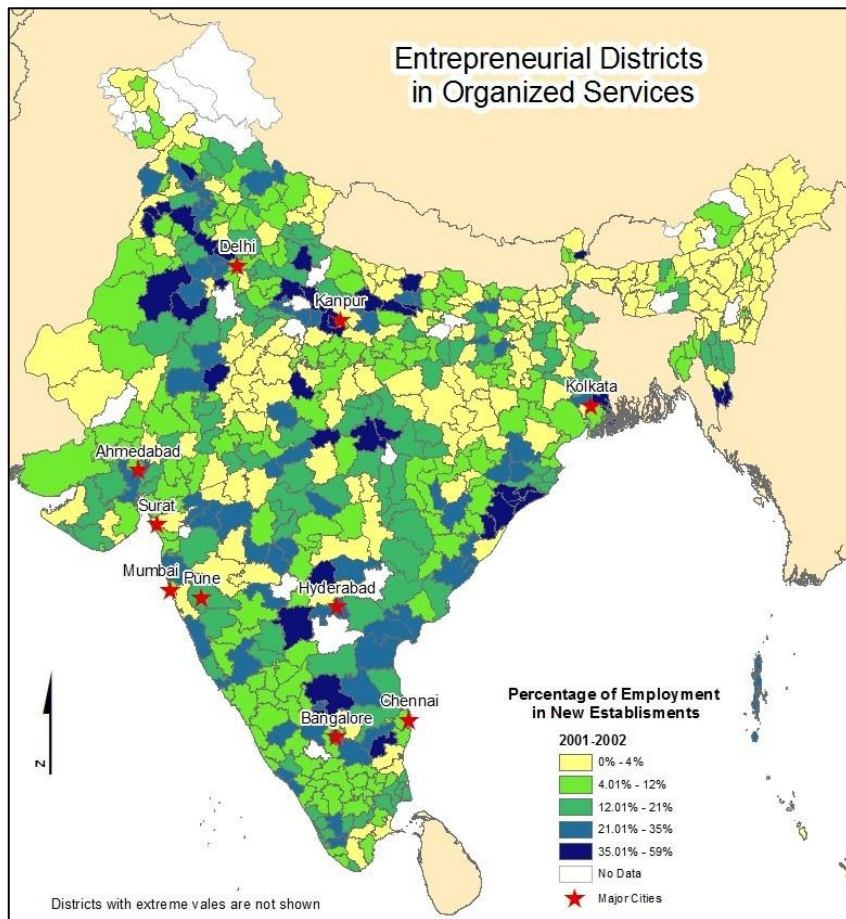


Table 1: Local industrial conditions for Indian entrepreneurship

	Organized Sector		Unorganized Sector	
	Mean	Standard deviation	Mean	Standard deviation
<u>District traits:</u>				
District population	2,972,828	1,731,997	2,963,367	1,721,814
District population density (persons per sq. km.)	810	2,477	806	2,447
Share of population with a graduate education	5.9%	2.7%	5.9%	2.7%
Demographic divided for a district (age profile)	2.09	0.30	1.32	0.26
Index of infrastructure quality for district	2.93	0.76	2.97	0.69
Stringency of labor laws: adjustments	0.69	0.84	0.68	0.84
Stringency of labor laws: disputes	-0.41	1.24	-0.41	1.23
Travel time to closest of India's ten largest cities (minutes driving)	446	240	444	239
<u>Industrial traits - Manufacturing:</u>				
Total employment in district-industry	1,383	5,020	4,517	15,389
Start-up employment in district-industry	151	788	553	2,938
Labor market strength	0.09	0.13	0.09	0.11
Input / supplier strength	-1.64	0.25	-1.71	0.24
Chinitz index of small suppliers	0.48	1.33		
Output / customer strength (x10 for presentation)	0.10	0.24	0.06	0.24
<u>Industrial traits - Services:</u>				
Total employment in district-industry	1,761	5,892	2,885	8,145
Start-up employment in district-industry	268	1,429	502	1,581

Notes: Descriptive statistics based on Annual Survey of Industries and National Sample Survey, various rounds.

Table 2A: Summary Statistics for Organized Manufacturing (2005-06) by State

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	9,835	2,032	813,062	65,373	21%	8%
Bihar	39	2,396	510	137,875	19,121	21%	14%
Chandigarh	1	242	12	8,747	327	5%	4%
Delhi	1	2,413	158	109,699	9,225	7%	8%
Goa	2	472	65	34,798	2,448	14%	7%
Gujarat	18	11,216	1,428	697,196	83,337	13%	12%
Haryana	16	3,772	279	354,229	35,713	7%	10%
Himachal Pradesh	11	691	261	52,120	17,391	38%	33%
Karnataka	17	5,401	981	459,946	80,582	18%	18%
Kerala	14	3,622	300	245,274	11,891	8%	5%
Madhya Pradesh	44	3,325	419	242,135	20,745	13%	9%
Maharashtra	30	14,058	1,749	1,035,176	101,400	12%	10%
Orissa	13	1,498	354	104,619	23,484	24%	22%
Pondicherry	4	480	108	37,516	5,007	22%	13%
Punjab	10	7,302	616	410,178	27,099	8%	7%
Rajasthan	26	4,865	731	257,648	28,570	15%	11%
Tamil Nadu	21	12,834	1,902	1,032,308	100,686	15%	10%
Uttar Pradesh	58	9,126	1,989	636,359	112,996	22%	18%
West Bengal	17	4,847	403	440,758	28,467	8%	6%

Notes: Descriptive statistics based on Annual Survey of Industries and National Sample Survey, various rounds. State-level traits are calculated over districts included in estimation samples.

Table 2B: Summary Statistics for Unorganized Manufacturing (2005-06) by State

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	1,429,916	150,789	2,582,039	233,198	11%	9%
Bihar	39	1,304,917	60,122	2,082,220	101,453	5%	5%
Chandigarh	1	1,318	175	2,427	189	13%	8%
Delhi	1	97,424	27,483	448,017	152,965	28%	34%
Goa	2	10,132	775	25,575	1,845	8%	7%
Gujarat	18	637,662	106,278	1,384,543	215,246	17%	16%
Haryana	16	224,098	46,957	481,043	87,316	21%	18%
Himachal Pradesh	11	102,675	16,056	135,317	22,944	16%	17%
Karnataka	17	937,626	57,736	1,798,046	114,712	6%	6%
Kerala	14	569,116	103,104	1,092,465	220,759	18%	20%
Madhya Pradesh	44	1,017,799	85,875	1,852,854	156,109	8%	8%
Maharashtra	30	1,105,928	156,352	2,505,209	376,932	14%	15%
Orissa	13	830,361	40,010	1,435,721	64,572	5%	4%
Pondicherry	4	13,328	3,451	40,641	12,906	26%	32%
Punjab	10	284,890	32,573	547,420	59,377	11%	11%
Rajasthan	26	605,522	68,804	1,029,463	119,245	11%	12%
Tamil Nadu	21	1,454,360	192,973	3,052,908	429,085	13%	14%
Uttar Pradesh	58	2,250,792	393,130	4,418,741	738,044	17%	17%
West Bengal	17	2,689,853	353,045	4,742,212	533,644	13%	11%

Notes: See Table 2A.

Table 3A: Summary Statistics for Organized Services (2001-02) by State

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	50,203	16,649	472,339	115,668	33%	24%
Bihar	39	24,335	2,426	186,560	15,992	10%	9%
Chandigarh	1	1,951	233	18,528	2,315	12%	12%
Delhi	1	22,471	3,278	232,993	23,766	15%	10%
Goa	2	1,250	262	12,622	2,150	21%	17%
Gujarat	17	42,785	3,275	536,731	24,907	8%	5%
Haryana	16	13,366	2,614	149,652	19,345	20%	13%
Himachal Pradesh	12	2,623	563	25,317	3,731	21%	15%
Karnataka	17	42,364	11,963	386,475	72,577	28%	19%
Kerala	14	34,709	7,037	361,322	55,934	20%	15%
Madhya Pradesh	44	38,538	5,548	374,968	43,684	14%	12%
Maharashtra	30	90,819	19,160	869,617	160,806	21%	18%
Orissa	13	18,178	4,200	148,874	27,274	23%	18%
Pondicherry	4	2,762	669	29,876	7,527	24%	25%
Punjab	10	14,438	3,463	185,046	29,020	24%	16%
Rajasthan	26	26,970	6,491	244,282	48,074	24%	20%
Tamil Nadu	21	48,190	7,841	548,315	59,680	16%	11%
Uttar Pradesh	58	103,267	20,564	897,253	171,210	20%	19%
West Bengal	17	35,660	4,445	382,695	38,626	12%	10%

Notes: See Table 2A.

Table 3B: Summary Statistics for Unorganized Services (2001-02) by State

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	1,283,845	227,385	2,056,671	352,485	18%	17%
Bihar	39	1,433,440	69,281	2,010,992	98,044	5%	5%
Chandigarh	1	16,092	3,731	25,496	5,607	23%	22%
Delhi	1	224,651	57,135	372,375	93,097	25%	25%
Goa	2	12,705	3,222	22,494	5,223	25%	23%
Gujarat	17	556,165	85,688	779,596	122,316	15%	16%
Haryana	16	208,836	61,600	299,033	85,968	29%	29%
Himachal Pradesh	12	77,258	24,386	116,024	36,810	32%	32%
Karnataka	17	633,884	121,182	996,150	200,301	19%	20%
Kerala	14	560,170	179,377	833,808	272,035	32%	33%
Madhya Pradesh	44	567,032	87,800	872,027	143,214	15%	16%
Maharashtra	30	1,094,885	252,756	1,732,570	416,709	23%	24%
Orissa	13	593,974	66,535	997,465	100,347	11%	10%
Pondicherry	4	16,095	3,887	32,017	8,068	24%	25%
Punjab	10	343,779	81,993	474,572	117,711	24%	25%
Rajasthan	26	514,407	107,974	748,729	161,132	21%	22%
Tamil Nadu	21	962,148	176,200	1,575,196	301,720	18%	19%
Uttar Pradesh	58	2,472,139	446,279	3,448,423	663,916	18%	19%
West Bengal	17	1,524,382	166,715	1,939,358	216,087	11%	11%

Notes: See Table 2A.

Table 4A: District entrepreneurship estimations - Organized manufacturing

	Base estimation	District traits	Full estimation	Using log entry counts
	(1)	(2)	(3)	(4)
<i>DV is log entry employment by district-industry</i>				
Log of incumbent employment in district-industry	0.229+++ (0.043)	0.186+++ (0.040)	-0.028 (0.048)	0.032+ (0.018)
Log of district population	0.531+++ (0.179)	0.483+++ (0.155)	0.475+++ (0.156)	0.216+++ (0.056)
<u>District Traits:</u>				
Log of district population density		-0.569+++ (0.088)	-0.563+++ (0.080)	-0.197+++ (0.029)
Share of population with graduate education		0.211+ (0.110)	0.235++ (0.107)	0.078+ (0.042)
Demographic dividend for district (age profiles)		0.605 (0.458)	0.567 (0.446)	0.271 (0.177)
Index of infrastructure quality for district		0.018 (0.100)	0.096 (0.094)	0.015 (0.038)
Strength of household banking environment		0.143 (0.104)	0.095 (0.100)	0.027 (0.036)
Stringency of labor laws in district's state		-0.210+++ (0.070)	-0.161++ (0.064)	-0.095+++ (0.023)
Log travel time to closest large city		-0.275+++ (0.090)	-0.241+++ (0.083)	-0.091+++ (0.031)
<u>Local Industrial Conditions by Incumbent Firms:</u>				
Labor market strength for district-industry			0.161 (0.102)	0.026 (0.041)
Inputs / supplier strength for district-industry			0.485+++ (0.098)	0.154+++ (0.043)
Outputs / customer strength for district-industry			0.388+++ (0.140)	0.167+++ (0.057)
Chinitz small suppliers metric for district-industry			0.279 (0.213)	0.337+++ (0.129)
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843
Adjusted R-squared	0.128	0.166	0.218	0.279

Notes: Estimations quantify the relationship between district-industry employment in new establishments and local conditions. District-level traits are taken from the 2001 Census. Industrial conditions are calculated from 2005-06 using incumbent establishments in the district-industry. Labor regulations are a composite of adjustment and disputes laws. Estimations weight observations by an interaction of district size and industry size, include industry fixed effects, and cluster standard errors by district. Non-logarithm variables are transformed to have unit standard deviation for interpretation.

Table 4B: District entrepreneurship estimations - Unorganized manufacturing

	Base estimation	District traits	Full estimation	Using log entry counts
	(1)	(2)	(3)	(4)
<i>DV is log entry employment by district-industry</i>				
Log of incumbent employment in district-industry	0.163+++ (0.031)	0.123+++ (0.029)	-0.075++ (0.029)	-0.040 (0.026)
Log of district population	1.051+++ (0.161)	0.878+++ (0.157)	1.010+++ (0.160)	0.866+++ (0.138)
<u>District Traits:</u>				
Log of district population density		-0.019 (0.070)	-0.044 (0.068)	-0.044 (0.057)
Share of population with graduate education		-0.002 (0.080)	-0.026 (0.084)	-0.046 (0.074)
Demographic dividend for district (age profiles)		0.954+++ (0.326)	1.053+++ (0.330)	0.798+++ (0.285)
Index of infrastructure quality for district		0.386+++ (0.096)	0.365+++ (0.097)	0.325+++ (0.086)
Strength of household banking environment		0.222+++ (0.080)	0.211+++ (0.080)	0.193+++ (0.071)
Stringency of labor laws in district's state		-0.007 (0.069)	0.000 (0.069)	0.030 (0.062)
Log travel time to closest large city		-0.004 (0.069)	0.009 (0.074)	0.017 (0.065)
<u>Local Industrial Conditions by Incumbent Firms:</u>				
Labor market strength for district-industry			0.263+++ (0.075)	0.228+++ (0.067)
Inputs / supplier strength for district-industry			0.553+++ (0.107)	0.504+++ (0.096)
Outputs / customer strength for district-industry			0.291+++ (0.050)	0.246+++ (0.044)
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	6451	6451	6451	6451
Adjusted R-squared	0.195	0.233	0.264	0.294

Notes: See Table 4A.

Table 5: District entrepreneurship estimations - Services

	Organized Services			Unorganized Services		
	Base estimation	District traits	Using log entry counts	Base estimation	District traits	Using log entry counts
	(1)	(2)	(3)	(4)	(5)	(6)
<i>DV is log entry employment by district-industry</i>						
Log of incumbent employment in district-industry	-0.003 (0.038)	-0.104+++ (0.033)	-0.054++ (0.023)	0.094+++ (0.024)	0.037+ (0.021)	0.037+ (0.021)
Log of district population	1.278+++ (0.148)	1.023+++ (0.135)	0.711+++ (0.092)	1.213+++ (0.107)	1.113+++ (0.111)	1.113+++ (0.111)
Log of district population density		-0.014 (0.086)	-0.028 (0.056)		-0.097+ (0.057)	-0.097+ (0.057)
Share of population with graduate education		0.348+++ (0.085)	0.230+++ (0.059)		0.179+++ (0.068)	0.179+++ (0.068)
Demographic dividend for district (age profiles)		0.548+ (0.331)	0.329 (0.230)		0.574++ (0.229)	0.574++ (0.229)
Index of infrastructure quality for district		0.339+++ (0.096)	0.242+++ (0.067)		0.420+++ (0.068)	0.420+++ (0.068)
Strength of household banking environment		0.174++ (0.087)	0.108+ (0.060)		0.323+++ (0.068)	0.323+++ (0.068)
Stringency of labor laws in district's state		-0.117+ (0.067)	-0.076+ (0.046)		-0.154+++ (0.048)	-0.154+++ (0.048)
Log travel time to closest large city		-0.011 (0.054)	-0.021 (0.037)		0.048 (0.051)	0.048 (0.051)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3340	3340	3340	6552	6552	6552
Adjusted R-squared	0.201	0.252	0.252	0.471	0.536	0.536

Notes: See Table 4A.

Table 6: Conditional entrepreneurship estimations - Manufacturing

	Organized Manufacturing			Unorg. Manufacturing	
	Base	Including	Including	Base	Including
	estimation	Marshallian	Chinitz	estimation	Marshallian
	(1)	(2)	(3)	(4)	(5)
<i>DV is log entry employment by district-industry</i>					
Log of incumbent employment in district-industry	0.155+++ (0.025)	-0.064++ (0.031)	-0.063++ (0.030)	0.109+++ (0.026)	-0.100+++ (0.026)
Labor market strength for district-industry		0.515+++ (0.089)	0.496+++ (0.089)		0.370+++ (0.066)
Inputs / supplier strength for district-industry		0.344+++ (0.080)	0.263+++ (0.088)		0.448+++ (0.108)
Outputs / customer strength for district-industry		0.291+++ (0.113)	0.304+++ (0.111)		0.264+++ (0.045)
Chinitz small suppliers metric for district-industry			0.597+++ (0.203)		
District and industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	6451	6451
Adjusted R-squared	0.259	0.306	0.309	0.308	0.340

Notes: See Table 4A. Estimations consider entry patterns conditional on industry and district fixed effects.

Table 7: Entry size distribution estimations - Manufacturing

	Entering establishment employment of:			
	10-19	20-39	40-99	100+
	(1)	(2)	(3)	(4)
<i>DV is log entry employment by district-industry</i>				
Log of incumbent employment in district-industry	0.038++ (0.016)	0.062+++ (0.019)	0.048++ (0.022)	0.024 (0.027)
Labor market strength for district-industry	-0.002 (0.055)	0.134++ (0.066)	0.218+++ (0.071)	0.447+++ (0.087)
Inputs / supplier strength for district-industry	0.106+ (0.055)	0.127++ (0.063)	0.115+ (0.069)	0.046 (0.080)
Outputs / customer strength for district-industry	0.142+++ (0.051)	0.160++ (0.064)	0.226+++ (0.079)	0.277+++ (0.091)
Chinitz small suppliers metric for district-industry	0.598+++ (0.167)	0.384++ (0.153)	0.108 (0.138)	-0.028 (0.138)
District and industry fixed effects	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843
Adjusted R-squared	0.190	0.188	0.194	0.219

Notes: See Tables 4A and 6. Estimations disaggregate entry into an entry size distribution based upon initial employment in the establishment.

Table 8: Entrepreneurship estimations using changes from 1989 to 2005

	Base estimation	Including Marshallian
	(1)	(2)
<i>DV is log entry empl.</i>		
Log of incumbent employment in district-industry	0.068++ (0.033)	-0.083++ (0.036)
Labor market strength for district-industry		-0.085 (0.127)
Inputs / supplier strength for district-industry		0.428+++ (0.124)
Outputs / customer strength for district-industry		0.495+++ (0.087)
District-industry, district-year, and industry-year fixed effects	Yes	Yes
Observations	5848	5848
Adjusted R-squared	0.269	0.292

Notes: See Tables 4A and 6. Estimations consider changes in incumbent industrial conditions and entrepreneurship from 1989 to 2005 for the organized manufacturing sector.