

Local industrial structures and female entrepreneurship in India

*Ejaz Ghani**, *William R. Kerr**†* and *Stephen D. O’Connell****

*World Bank

**Harvard University, Bank of Finland, and NBER

***City University of New York, Graduate Center

†Corresponding author: William R. Kerr, Rock Center 212, Harvard Business School, Boston, MA 02163.
email <wkerr@hbs.edu>

Abstract

We analyze the spatial determinants of female entrepreneurship in India in the manufacturing and services sectors. We focus on the presence of incumbent female-owned businesses and their role in promoting higher subsequent female entrepreneurship relative to male entrepreneurship. We find evidence of agglomeration economies in both sectors, where higher female ownership among incumbent businesses within a district-industry predicts that a greater share of subsequent entrepreneurs will be female. Moreover, higher female ownership of local businesses in related industries (e.g. those sharing similar labor needs and industries related via input–output markets) predict greater relative female entry rates even after controlling for the focal district-industry’s conditions. The core patterns hold when using local industrial conditions in 1994 to instrument for incumbent conditions in 2000 and 2005. The results highlight that the traits of business owners in incumbent industrial structures influence the types of entrepreneurs supported.

Key words: Female, gender, entrepreneurship, agglomeration, cluster, business networks, development, informal sector, India, South Asia

JEL classifications: J16, L10, L26, L60, L80, M13, O10, R00, R10, R12

Date submitted: 5 March 2012 **Date accepted:** 18 February 2013

1. Introduction

A central driver of economic growth over the past century is the increased role of women. This growth in the role of women comes in many forms: increased female labor force participation generally, reduced discrimination and wage differentials that encourage greater effort, improved advancement practices that promote talented women into leadership and managerial roles and similar. Simply put, empowering half of your potential workforce has significant economic benefits beyond promoting gender equality (Duflo, 2005, 2011; World Bank 2012). This growth stimulus is particularly true for female entrepreneurship and the economic dynamics that entrepreneurship promotes. Yet, as Klapper and Parker (2011) review, we are still just beginning to identify the local business and social factors that unlock female entrepreneurship.

This article uses detailed micro-data on the unorganized manufacturing and services sectors of India in 2000–2001 and 2005–2006 to explore the spatial factors that promote female entrepreneurship and business ownership. The micro-data are a representative

sample of the Indian economy, and the establishment records identify the gender of the owner for proprietary establishments (which account for greater than 95% of establishments). For most surveys, we can also identify establishments that are new entrants. From these micro-data, we develop relative rates of female entrepreneurship and business ownership at the district-industry-year level.

The central focus of our article is on identifying and quantifying the importance of existing female business networks and economic geography for promoting subsequent entrepreneurship among women. We primarily evaluate these local conditions using conditional estimations that include industry-year and district-year effects. These estimations isolate variations in incumbent industrial conditions by district-industry-year. We prepare measures of the overall incumbent female business ownership in the district-industry and indices of how favorable the district's broader industrial structure is to new entrants in terms of typical agglomeration factors: the suitability of local labor markets and the strength of input-output markets for buying and selling goods. These indices link measures of how related two industries are on these two dimensions with the relative presence of industries in each district.

Our estimations emphasize that favorable incumbent industrial conditions increase the subsequent relative rate of female entrepreneurship and business ownership in a district-industry. We further calculate the indices separately using male-owned incumbents and female-owned incumbents. This separation strongly emphasizes that higher relative rates of female entrepreneurship closely follow upon industrial structures with existing female-owned businesses. Although we do not observe direct economic exchanges among businesses with our data, these patterns are consistent with hypotheses of the importance of existing business networks for promoting female entrepreneurs. We further confirm these results when instrumenting for current incumbent structures using lagged 1994 incumbent structures.

Our empirical methodology draws from the work on agglomeration and the spatial determinants of entrepreneurship in advanced economies, most often related to the manufacturing sector, that are reviewed later in this article. In related work, Ghani et al. (2011b) and Mukim (2011) evaluate spatial factors that explain a high rate of entrepreneurship in a district-industry generally and irrespective of the owner's gender. These earlier studies also contrast entry in the organized sector with that in the unorganized sector. Our current article focuses instead on factors that explain higher female entrepreneurship shares independent of aggregate levels of entry. We also specifically focus here on the local industrial conditions of female-owned businesses. In discussing these results, we identify central findings from Ghani et al. (2011b) and Mukim (2011) where appropriate.

In addition to the conditional estimations that provide the tightest econometric framework for our study, we also consider unconditional estimations that include industry-year effects only to identify general district-level attributes that correlate with high rates of female entrepreneurship. These estimations emphasize several factors. First, relative rates of female entrepreneurship do not appear strongly linked to the district's population or to the district-industry's overall employment levels, although relative female entrepreneurship rates do decline with population density. More strikingly and relevant for policy makers, better local infrastructure strongly connects with higher relative female entry in both manufacturing and services. Further, local education and the female literacy rate matter for services entry. Stricter labor regulations are also associated with higher rates of female entry.

These findings are important for Indian policy makers and business leaders. Despite significant economic advancement since liberalization began, the role of women in the Indian economy still lags well behind that of advanced economies (e.g. Dunlop and Velkoff, 1999; Mammen and Paxson, 2000; World Bank, 2011). Cross-country data from the World Bank Entrepreneurship Snapshots find that India's rate of entrepreneurship is lower than its stage of development would suggest; similar comparisons also highlight that India's gender ratio among entrepreneurs is lower than its peers. This dual under-performance has cultural and economic antecedents, but it is starting to change. Women are making economic gains in the Indian economy, and further progress represents a tremendous growth opportunity for the country.

This study contributes most directly to three literature strands. First, our work builds upon prior studies of gender differences in entrepreneurship (e.g. Estrin and Mickiewicz, 2011). Parker (2009) and Klapper and Parker (2011) offer a comprehensive review of this literature and appropriate references. Our article is among the first to study how spatial differences in gender ratios in entrepreneurship relate to local incumbent business structures. Existing research mostly employs cross-country studies of gender ratios in entrepreneurship (e.g. Minniti, 2010; Minniti and Naudé, 2010), whereas we analyze local area determinants. Second, our article contributes to a growing set of spatial studies on entrepreneurship in India (e.g. Khanna, 2008; Ghani et al., 2011b; Mukim, 2011) and a broader set of work on India's industrial organization and economic performance.¹ These contributions constitute an important input to the growing body of work on entrepreneurship and economic advancement in developing countries (e.g. Ardagna and Lusardi, 2008; Schoar, 2009; Klapper et al., 2010). Third, as discussed later in the article and in the conclusions, our work is an important input to studies of economic geography that quantify the causes and impacts of local industrial structures.

Beyond the few papers on the spatial determinants of entrepreneurship in India, the closest paper to our work is Rosenthal and Strange (2012). Rosenthal and Strange (2012) document within-metropolitan area sorting for female entrepreneurs in the USA. They show that a spatial mismatch exists for female entrepreneurs similar to that found earlier on racial lines. Our work has several key differences. First, Rosenthal and Strange (2012) focus on spatial differences within metropolitan areas (e.g. using one-mile rings around a business). Our Indian data only identify districts for firms, which are on average about twice the size of U.S. counties. Thus, we do not study sorting at the same spatial level. Second, and more generally, the contextual differences of the two countries are very stark. Rosenthal and Strange (2012), for example, consider automobile commuting patterns for U.S. entrepreneurs, whereas our focus will be on whether basic infrastructure such as paved roads and electricity exist.

Nonetheless, our two studies are closely connected in that they both describe a link of past female business ownership in a spatial area with subsequent female entry. To establish this connection, we employ below a modified form of the Rosenthal and Strange (2012) differencing methodology to show comparability between our studies. We also show that the poor access of female entrepreneurs to banking services that Rosenthal and Strange (2012) identify for the USA is present in India as well.

1 For example, Lall et al. (2004), Lall and Mengistae (2005), Nataraj (2011), Kathuria et al. (2010), Hasan and Jandoc (2010), Dehejia and Panagariya (2010), Fernandes and Pakes (2010), Fernandes and Sharma (2011) and Mukim and Nunnenkamp (2011).

We hope that further research continues to refine our understanding of this interface between incumbent conditions and gender ratios of new entrants. Studies from psychology suggest that gender is among the strongest predictors of an individual's peer group. These factors shape the identity that individuals develop (e.g. Akerlof and Kranton, 2000²) and the networks they form (Munshi and Rosenzweig, 2006). Moreover, this factor is present in developing and advanced economies (e.g. 'the old boy network') and industry contexts. A message from this article and Rosenthal and Strange (2012) is that studies of economic geography can make important progress by considering further how identities like gender associate with agglomeration economies and local business structure. We study there the interaction between female business owners and (potential) female entrepreneurs, and further work should consider other dimensions (e.g. worker flows between firms).

The plan of this article is as follows. Section 2 discusses our entrepreneurship data and spatial differences in female business ownership across India. Section 3 reviews the spatial determinants of entrepreneurship and our metrics. Section 4 quantifies the spatial determinants of female entrepreneurship in manufacturing and services. Section 5 presents our instrumental variable specifications that use lagged incumbent conditions. The final section concludes.

2. Female entrepreneurship rates in India

We employ cross-sectional establishment-level surveys of manufacturing and services enterprises carried out by the Government of India. Our manufacturing data are taken from surveys conducted in fiscal years 1994, 2000 and 2005. The services sector has only been surveyed more recently, in fiscal years 2001 and 2006. In all cases, the survey was undertaken over two fiscal years (e.g. the 1994 survey was conducted during 1994–1995), but we will only refer to the initial survey year for simplicity. This section describes some key features of these data for our study.

Our work considers portions of the Indian economy surveyed by the National Sample Survey Organisation (NSSO). The Government of India conducts the NSSO surveys for the collection of data on economic and operational characteristics of enterprises in the unorganized sector. These surveys are the foundation for many published reports on the state of Indian businesses and government agency monitoring of the Indian economy. The typical survey collects data from over 150,000 Indian establishments. In this respect, the surveys are comparable to the Annual Survey of Manufacturing conducted in the USA, with the Indian sampling frame being about three times larger.³

-
- 2 Akerlof and Kranton (2000) propose a model of behavior which takes into account differences in the identity of agents. Identity underlies a type of externality in which a person's actions have meaning for and evoke responses in others, and identity reveals a way that preferences can be changed as notions of identity evolve (or are manipulated) within a society. Our study investigates the extent to which gender forms an identity that influences the scope of economic interactions in local areas and industries, especially with respect to entrepreneurs that are very dependent upon the local economy for resources and sales opportunities.
 - 3 Nataraj (2011), Kathuria et al. (2010) and Hasan and Jandoc (2010) provide additional detail on the manufacturing survey data. Dehejia and Panagariya (2010) provide a detailed overview of the services data and its important characteristics. Published reports by the Government of India with the NSSO data are available at: http://mospi.nic.in/stat_act_t14.htm.

The NSSO survey collects information on the gender of establishment owners that are a key focus of our study. Due to the nature of India's data collection, the NSSO collects information on a representative sample of all services establishments and on a representative sample of the unorganized sector of manufacturing. A manufacturing business is considered as part of the unorganized sector if it has fewer than 10 employees and uses electricity. If the establishment does not use electricity, the threshold is 20 workers. Organized manufacturing establishments are not surveyed by the NSSO because they are instead surveyed by the Annual Survey of Industries (ASIs). The ASI unfortunately does not collect the gender of the business owner. The unorganized sector accounts for over 99% of Indian manufacturing establishments.

For most of our estimations of entrepreneurship in the services sector, we mimic the manufacturing sector's split into organized and unorganized components to provide better comparability. We do this by classifying services establishments with fewer than five workers and those listed as an 'own-account enterprise' (OAE) as the unorganized sector. OAE firms are those 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 unorganized sector comprises approximately 70–80% of total employment in both manufacturing and services, providing comparable baselines across the two sectors. Table A1 provides by survey year the relative sizes of these groups. Our results are not sensitive to this specific cutoff choice, and we show below estimations for the services sector that use the full sample of services firms.⁴

Establishments are surveyed by the NSSO with state and four-digit National Industry Classification (NIC) stratification. We use the provided sample weights to construct population-level estimates. Much of our analysis employs district-industry variation. Districts are administrative subdivisions of Indian states or territories that provide meaningful local economic conditions. The average district size is around 5500 km²—roughly twice the size of a U.S. county—and there is substantial variability in district size (standard deviation of ~5500). Based on these spatial dimensions and India's low geographic mobility for unorganized sector workers, Indian districts can be effectively considered as self-contained labor markets.

Across our state sample, there are 514 districts. Due to changes in district definitions (e.g. bifurcations and combinations), we build a concordance of district definitions that is consistent over time. This concordance reduces the number of unique districts to 368. This number of unique areas is then reduced slightly to 320 when we exclude districts that have a small number of plants (less than 50). The male or female business ownership questions that we use are an outcome of the survey and not a factor in the stratification design. We have confirmed that our results hold at the state-industry level, for which the stratification was originally designed. All analysis (excepting the summary statistics presented below) is performed at the three-digit NIC level.

The NSSO surveys the ownership type of each establishment. Establishments can be listed as male proprietary, female proprietary, other owned, cooperative, household

4 Reviewing documentation on the NSSO survey design and stratification procedures, we believe that the data remain a viable representative sample with the size cutoff imposed. Our core empirical focus is on conditional estimations that remove district and industry fixed effects, thereby minimizing remaining issues like the possibility that larger establishments are located in bigger cities. Either way, we show below very similar findings when using the full sample of services firms.

partnership, multi-household partnership, private LLC and unknown. We focus primarily on the establishments listed as either male proprietary or female proprietary. These two groups constitute 98% of establishments in the informal manufacturing sector in 2000 and 2005 and 98% and 91% of establishments in services in 2001 and 2006, respectively. We use incumbent establishments defined as female proprietary to determine female business-ownership shares by district-industry. When developing local indices of broader local business conditions, described in the next section, we use all incumbent establishments.⁵

Our analysis primarily considers female entrepreneurship. Female entrepreneurs are a subset of female business owners. We identify entrepreneurs by using establishments that are classified as being 3 years old or less. Ghani et al. (2011a) describe this choice of how to measure entrepreneurship and our entry metrics in further detail. This young establishment distinction is unfortunately not collected for services establishments in 2006. As a consequence, we are limited to a cross-sectional analysis of female entrepreneurship rates for services in 2001, while we have greater longitudinal flexibility in manufacturing. In addition, we define incumbent establishments to be those older than 3 years. We use this incumbent group, which is mutually exclusive from the entrants, in defining all of our incumbent industrial structures. The data unfortunately do not distinguish spin-off businesses or exiting firms. In 2000, 14% of manufacturing establishments in our sample are young entrants, and the comparable figure from 2001 for services is 20%.

Table 1 provides descriptive statistics on our sample. Table 2 lists by state the gender traits of business owners in manufacturing and services establishments.⁶ Figure 1 provides a visual presentation. Using pre-2001 definitions, our study considers 17 states out of the initial 32 states/union territories available. The 15 exclusions were due to either the state not being sampled across all surveys or to data quality concerns stemming from small sample sizes and/or persistent conflict and political turmoil existing in the region. In 2001, three new states were created: Jharkhand (from Bihar), Chhattisgarh (from Madhya Pradesh) and Uttaranchal (from Uttar Pradesh). Our descriptive tables and district-level estimations account for these bifurcations with the three states considered as part of their parent states.

There are on average about 20 (consistently defined) districts per state. Larger states such as Uttar Pradesh have more than 50 districts, whereas smaller states such as Himachal Pradesh have only 12 districts. The explicit criteria with respect to the minimum sizes of districts are that the district has a population of at least 1 million in the 2001 census and has 50 or more establishments sampled. The exclusions are minor in terms of economic activity, and the resulting sample accounts for over 90% of employment in both sectors throughout the period of study.

Table 2 presents relative rates of female business ownership defined as the count of female-owned businesses divided by the total count of male- and female-owned businesses. These shares use sampling weights to yield population-level estimates. Relative female business-ownership rates in unorganized manufacturing are particularly

5 Beyond ownership, the surveys do not collect whether an establishment's management and daily business operations are controlled by one or more women. To a large degree, this issue is minimized by our focus on the unorganized sector where business sizes are small.

6 Comparable entrepreneurship tables are available upon request. We document female business owner ratios in Tables 2 and 3 so that the longitudinal pattern in services can be discussed.

Table 1. Descriptive statistics

Variable	Mean	Standard deviation
District characteristics (2001 census)		
District population	2,955,445	1,727,678
District population density (persons per square kilometer)	808	2458
Age profile (working age population/non-working age population)	1.32	0.26
Share of population with a graduate education	5.9%	2.7%
Index of infrastructure quality for district (0–4 scale)	2.98	0.68
Travel time to closest of 10 largest cities in driving minutes	446	240
Stringency of labor adjustment laws for district's state	0.69	0.84
Stringency of labor disputes laws for district's state	–0.41	1.24
Share of households in district with a banking relationship	0.35	0.12
Female literacy rate	51.7	15.1
Total fertility rate based on age-specific district fertility rates	3.3	1.1
Sex ratio measured as females per male	0.939	0.057
Establishment characteristics, unorganized manufacturing		
Total female-owned establishments in district-industry	866	6237
Total employment in female-owned establishments in district-industry	1068	7939
Total male-owned establishments in district-industry	1455	4276
Total employment in male-owned establishments in district-industry	3234	10,944
Female business-ownership share	0.13	0.28
Female entrepreneurship share	0.21	0.37
Index of labor market strength	0.09	0.11
Index of input market strength	–1.71	0.24
Index of output market strength	0.04	0.06
Index of labor market strength, female-owned businesses	0.08	0.17
Index of input market strength, female-owned businesses	–1.79	0.29
Index of output market strength, female-owned businesses	0.02	0.03
Establishment characteristics, unorganized services		
Total female-owned establishments in district-industry	152	801
Total employment in female-owned establishments in district-industry	216	1041
Total male-owned establishments in district-industry	1773	5651
Total employment in male-owned establishments in district-industry	2569	7628
Female business-ownership share	0.09	0.21
Female entrepreneurship share	0.11	0.24
Index of labor market strength	0.10	0.11
Index of labor market strength, female-owned businesses	0.10	0.15

Notes: Authors' calculations based on National Sample Survey data (various rounds). Output index values are multiplied by 10 for presentation.

high in Karnataka, Kerala and Tamil Nadu. These states have an average female establishment ownership rate exceeding 45%. In contrast, low female ownership shares are evident in Delhi, Bihar, Haryana and Gujarat. The average female business-ownership share increased from 26% in 2000 to 37% in 2005. On an employment-weighted basis, the rate increased from 17% to 25%. The female ownership rates across major cities have a distribution that is mostly similar to the distribution across states.

Table 2. State traits for unorganized manufacturing and unorganized services^a

State	Male establishment counts		Female establishment counts		Female establishment count share		Female establishment share, employee wtd.	
	2000	2005	2000	2005	2000	2005	2000	2005
State traits for unorganized manufacturing								
Andhra Pradesh	1,032,699	692,239	416,866	718,217	0.29	0.51	0.19	0.31
Bihar	1,144,596	993,016	98,100	302,687	0.08	0.23	0.06	0.17
Chandigarh	3899	866	1516	385	0.28	0.31	0.11	0.20
Delhi	195,648	78,588	14,647	10,417	0.07	0.12	0.05	0.06
Gujarat	450,651	469,895	63,448	138,227	0.12	0.23	0.07	0.13
Haryana	154,817	165,126	23,065	41,991	0.13	0.20	0.08	0.11
Himachal Pradesh	77,657	74,085	15,023	27,581	0.16	0.27	0.15	0.21
Karnataka	499,626	393,172	523,535	536,165	0.51	0.58	0.32	0.37
Kerala	234,113	268,328	222,497	275,774	0.49	0.51	0.35	0.33
Madhya Pradesh	766,962	808,605	185,485	199,779	0.19	0.20	0.15	0.15
Maharashtra	922,538	728,672	262,371	347,123	0.22	0.32	0.13	0.17
Orissa	753,698	596,057	197,368	206,140	0.21	0.26	0.15	0.18
Punjab	215,899	168,229	98,765	108,097	0.31	0.39	0.19	0.24
Rajasthan	460,644	444,819	128,113	150,295	0.22	0.25	0.15	0.18
Tamil Nadu	843,035	659,086	578,512	752,856	0.41	0.53	0.25	0.33
Uttar Pradesh	1,867,552	1,666,111	400,623	536,389	0.18	0.24	0.14	0.18
West Bengal	1,880,428	1,369,120	825,903	1,287,765	0.31	0.48	0.20	0.37
Totals and weighted averages	11,504,463	9,576,011	4,055,836	5,639,889	0.26	0.37	0.17	0.25
Unweighted averages	676,733	563,295	238,579	331,758	0.25	0.33	0.16	0.22
State	2001	2006	2001	2006	2001	2006	2001	2006
State traits for unorganized services								
Andhra Pradesh	1,112,761	1,176,882	150,065	161,327	0.12	0.12	0.12	0.12
Bihar	1,380,710	1,162,661	42,212	88,591	0.03	0.07	0.03	0.06
Chandigarh	14,157	14,717	1553	1726	0.10	0.10	0.12	0.12
Delhi	195,528	118,091	23,161	13,458	0.11	0.10	0.11	0.10
Gujarat	504,218	537,730	38,907	46,232	0.07	0.08	0.07	0.07
Haryana	188,152	274,203	14,254	22,262	0.07	0.08	0.06	0.08
Himachal Pradesh	68,401	96,043	6533	6097	0.09	0.06	0.09	0.06
Karnataka	576,517	572,006	45,110	56,639	0.07	0.09	0.07	0.09
Kerala	465,601	598,452	68,090	95,891	0.13	0.14	0.12	0.13
Madhya Pradesh	519,084	491,348	36,893	35,169	0.07	0.07	0.06	0.07
Maharashtra	956,493	1,108,430	111,740	132,973	0.10	0.11	0.11	0.11
Orissa	550,710	451,867	37,094	28,274	0.06	0.06	0.05	0.05
Punjab	307,254	384,706	25,913	35,799	0.08	0.09	0.08	0.08
Rajasthan	474,826	511,490	21,760	26,608	0.04	0.05	0.04	0.05
Tamil Nadu	828,011	789,713	112,915	114,120	0.12	0.13	0.12	0.12
Uttar Pradesh	2,279,441	2,117,998	163,330	118,687	0.07	0.05	0.06	0.06
West Bengal	1,390,771	1,734,408	113,206	127,953	0.08	0.07	0.07	0.07
Totals and weighted averages	11,812,636	12,140,745	1,012,738	1,111,805	0.08	0.08	0.08	0.08
Unweighted averages	694,861	714,161	59,573	65,400	0.08	0.09	0.08	0.08

^aAuthors' calculations based on National Sample Survey data (various rounds).

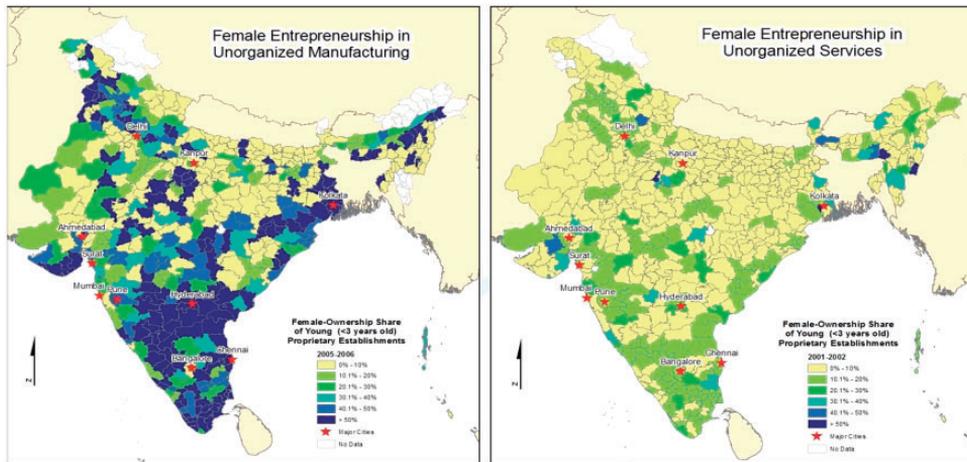


Figure 1. Female entrant shares in manufacturing and services.

In Table 2's services analysis, states with the highest female ownership rates are Kerala, Tamil Nadu and Andhra Pradesh, with average female ownership shares exceeding 12%. The lowest female ownership rates are in Rajasthan, Bihar, Orissa and Uttar Pradesh, each with 6% or less. The average female business-ownership share, with and without employment weights, was between 8% and 9% for 2001 and 2006. Female ownership rates in major cities tend to be higher than overall state averages in services. The correlation of state gender ratios between manufacturing and services is about 0.5.

Table 3 provides similar female business-ownership shares by two-digit NIC industry in manufacturing and services. Within manufacturing, female shares are highest and typically exceed 50% in industries related to chemicals and chemical products, tobacco products and paper and paper products. At the opposite end, female shares of 2% or less are evident in industries related to computers, motor vehicles, fabricated metal products and machinery and equipment. Among services industries, female ownership shares exceed 30% in industries related to sanitation and education. Industries related to land transport, machinery renting or other business activities have the lowest rates. Female-owned businesses are smaller in terms of employment than male-owned businesses in manufacturing, while they are larger in services.⁷

3. Spatial determinants of female entrepreneurship

We now describe the spatial factors that we relate to female entrepreneurship patterns. We first identify general district traits that can influence the gender ratio of business owners. We then construct indices of local industrial conditions for a specific district-industry.

7 The NSSO excludes from its service sector sampling frame a number of industries, notably retail trade, wholesale trade, financial intermediation and air transport. The research and development industry is not collected in 2006.

3.1. District-level conditions

Our initial explanatory measures naturally focus on basic traits of the district: population level (and its square), population density (persons per square kilometer), age profile and average education levels. Given our interest in gender balance, we also include the district's female literacy rate, total fertility rate and sex ratio. These traits are important as entrepreneurs tend to start their businesses in their current local area and are even disproportionately found in their region of birth (e.g. Figueiredo et al., 2002; Dahl and Sorenson, 2007; Michelacci and Silva, 2007). These controls will pick up supply-side effects associated with these factors. The measures are developed using India's 2001 population census, and Table 1 provides descriptive statistics.

Our population control captures the size of the local consumer market, which can be especially important for services businesses, and the overall level of surrounding economic activity (e.g. general availability of workers). Ghani et al. (2011b) find higher entry levels partially correlate with greater population, but we do not have a theoretical reason to suspect population influences the gender balance after controlling for other district attributes. Population density again reflects some measure of local market size, but it also goes beyond to consider the competition for local resources such as higher land rents (e.g. Drennan and Kelly, 2011). Ghani et al. (2011b) find population density reduces organized manufacturing entry, as has been shown in other contexts, but that it is not systematically related to entrepreneurship in the unorganized sectors that we study here. With respect to gender ratios for entrepreneurs, the work of Rosenthal and Strange (2012) on spatial sorting in the USA suggests that the female entrepreneurial share may be declining in density.

We model the district's age structure as the ratio of working age population to non-working age population. This ratio is the inverse of the dependency ratio, and it is sometimes referred to as the demographic dividend in the context of developing countries. Although some work finds age structure matters for entry rates (e.g. Evans and Leighton, 1989; Bönnte et al., 2009; Glaeser and Kerr, 2009), this fact has not been established for India. The general education of the workforce has been linked to higher entry in India (e.g. Ghani et al., 2011b) and the USA (e.g. Doms et al., 2010; Glaeser et al., 2010), and Amin and Mattoo (2008) further study human capital in the Indian context. It is not clear, however, whether general education should influence the gender balance of entrepreneurs. We measure the general education level of a district by the percentage of adults with a graduate (post-secondary) degree. All results are robust to alternatively defining a district's education as the percentage of adults with higher secondary education.

We next consider three traits specific to female advancement that are emphasized by prior studies.⁸ The first is the female literacy rate. Given the general link established between education and entrepreneurship, we anticipate a higher literacy rate will correlate with higher relative female entrepreneurship. The second is the total fertility rate measured as a composite of age-specific fertility rates in the district. This trait does not have clear prediction. The third factor is the sex ratio measured as the number of females per male in the district. We anticipate this to have a positive effect for raising female entry rates relative to male entry rates.

8 Klapper and Parker (2011) provide a review, and our working paper contains a set of references.

One notable omission from this list is the district's profile in terms of scheduled tribe and scheduled caste populations. Iyer et al. (2011b) show that castes and tribes vary in their overall rates of entrepreneurship. Castes and tribes also differ in their social norms on the appropriate role of women. Unreported tests analyze whether including these profiles influenced the results presented below. These inclusions are not important in our context when also controlling for intermediate variables such as the female literacy rate,⁹ and so we adopt the more parsimonious specification that is also more easily comparable in the future to other contexts outside of India. We believe, nonetheless, that the role of these social norms is important for future research with respect to female participation in the Indian economy. We later discuss scheduled tribe and scheduled caste populations when describing our instruments.

Beyond these basic demographics, we consider four additional local traits that may affect female entrepreneurship: quality of local physical infrastructure, travel time to one of India's 10 biggest cities, the stringency of a state's labor laws and the strength of local banking. Although these traits do not constitute an exhaustive list of local conditions, they are motivated by the literatures on entrepreneurship and India's development.

Basic infrastructure services such as electricity are essential for all businesses, but new entrants and the informal sector can be particularly dependent upon local infrastructure (e.g. established firms are better able to provision their own electricity if need be). Ghani et al. (2011b) and Mukim (2011) find that infrastructure strongly links to overall entrepreneurship levels in India. Lall (2007) and Datta (2011) link infrastructure investments to regional development in India. 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 that ranges from zero (no infrastructure access) to four (full access to all four infrastructure components).¹⁰

India's economy is undergoing dramatic structural changes (Desmet et al., 2012). 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 10 largest cities¹¹ as a measure of physical connectivity and across-district infrastructure. This is calculated based on data on India's road networks using GIS software.

Third, we model local labor regulations using state-level variation in policies. Several studies link labor regulations in Indian states to economic progress (Besley and Burgess,

9 The univariate correlation associates lower relative female entry rates with a high population share from scheduled castes and tribes; a univariate correlation is not observed for manufacturing.

10 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 robust to excluding these major cities from the analysis sample.

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

2004; Aghion et al., 2008), and Ghani et al. (2011b) find labor regulations suppress Indian entrepreneurship generally. van Stel et al. (2007) document similar patterns across multiple countries. 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.

Finally, Rosenthal and Strange (2012) identify that, compared with male entrepreneurs, female entrepreneurs in the USA tend to be in locations with weaker banking sectors. We develop a measure of local banking development through the share of households that report a banking relationship. Given our greater spatial scale than the Rosenthal and Strange (2012) analysis, we do not have a clear prediction for this measure. On the one hand, the U.S. experience would suggest a negative relationship. On the other hand, financial access may be important for empowering female entrepreneurs in developing countries. Given that we are looking at districts as a whole, rather than sorting within districts, we may capture more of this latter effect.

3.2. Agglomeration theories

The above factors are district-level phenomena, and our first exercise is to relate the female entrepreneurship ratio to these factors. The focus of our study, however, is on the local industrial conditions of district-industries and how they shape female entrepreneurship. 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). Duranton and Puga (2004) and Rosenthal and Strange (2004) review the subsequent literature in detail. Prior work emphasizes the importance of these conditions for explaining overall entry rates,¹² but this work does not explore the impact of favorable local conditions on the gender balance of entrepreneurs.

We also explore the importance of whether the gender profile of the incumbent industrial structure affects the gender profile of new entrants. The role of business networks among women in developing countries is frequently mentioned by development economists, and gender can be a core determinant of identities and peer groups. We are not aware of any prior work that systematically considers female business ownership across regions and industries in multiple economic sectors. For simplicity, we describe our metrics below in a generic language that discusses district-industry employment. Our empirical work, however, focuses on metrics that use female-owned incumbent businesses to determine resource access for new entrants.

In all of our estimations, we control in some format 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

12 This conceptual approach is used to describe location choice decisions and city structures by Glaeser and Kerr (2009) for the USA, Jofre-Monseny et al. (2011) for Spain, Dauth (2011) for Germany and Ghani et al. (2011b) and Mukim (2011) for India. Our working paper provides additional references for related literatures on industry linkages for agglomeration, FDI choice and location traits, social networks and entrepreneurship and network formation in developing economies.

the spatial location of new entrants (e.g. Figueiredo et al., 2009; Glaeser and Kerr, 2009). We model both the total employment in incumbent firms for the district-industry and the count of female-owned incumbent businesses specifically.

The first agglomeration rationale is that proximity to customers and suppliers reduces transportation costs and thereby increases productivity. To test the importance of this mechanism within the manufacturing sector, 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 $\text{Input}_{i \leftarrow k}$ as the share of industry i ’s inputs that come from industry k , and $\text{Output}_{i \rightarrow k}$ as the share of industry i ’s outputs that go to industry k . These measures run from 0 (no input or output purchasing relationship exists) to 1 (full dependency on the paired industry). These shares are calculated relative to all input–output flows and are not symmetrical by design ($\text{Input}_{i \leftarrow k} \neq \text{Input}_{k \leftarrow i}$, $\text{Input}_{i \leftarrow k} \neq \text{Output}_{k \rightarrow i}$).

We summarize the quality of a district d in terms of its input flows for an industry i as $\text{Input}_{di} = -\sum_{k=1, \dots, I} \text{abs}(\text{Input}_{i \leftarrow k} - E_{dk}/E_d)$, 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 representing employment among incumbent firms. The measure is mostly orthogonal to district size, which we separately consider, and a negative value is taken so that the metric ranges between -2 (i.e. no inputs available in the local market) and 0 (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.¹³

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.

For most of our estimations, we employ a combined input–output market strength measured as $\text{Input–Output}_{di} = \text{mean}(\text{Input}_{di}, \text{Output}_{di})$. We take the mean value after transforming both Input_{di} and Output_{di} to have zero mean and unit standard deviation. Our use of a consolidated metric is primarily motivated by the instrumental-variable

13 The input metric is not perfectly orthogonal to district size to the degree to which larger districts have more independent economic zones than smaller districts. 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.

Chinitz (1961) emphasizes that average size of local suppliers is an important factor for entrepreneurship beyond general supplier market suitability. A number of empirical studies referenced in our working paper find support in local start-up conditions. We do not examine the Chinitz effect in the unorganized sector context because the definition of an unorganized establishment in India is based on establishment size.

specifications that use the 1994 incumbent conditions. The consolidated metric reduces the number of endogenous regressors that we need to simultaneously instrument for. This need to consolidate the input and output metrics also accounts for why we do not build into our output measure the size of the underlying district (e.g. comparing Glaeser and Kerr (2009) metrics with Ellison et al. (2010) metrics). As our outcome variables are ratios that do not depend upon the size of the district-industry directly, we have greater flexibility in these choices. We instead control for any role that size may be playing through the fixed effects included in the estimations. We find similar results to those reported below when instead considering the max values over Input_{di} and Output_{di} .

Beyond 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). Although 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 U.S. biotech industry. These specialized workers are often tightly clustered together.¹⁴

We unfortunately lack Indian data to model direct occupational flows, so we instead take a simple approach. Greenstone et al. (2010) calculate from the Current Population Survey the rate at which workers move between industries in the USA at the two-digit industry level. 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. We calculate this connection for both manufacturing and services.

Using the rates calculated by Greenstone et al. (2010) requires that we map Indian industries to U.S. industries to apply this framework. This is straightforward at our level of aggregation. This process assumes that mobility flows among industries in India are similar to those evident in the USA. We cannot directly test this assumption and can only note that the industries rank similarly on dimensions such as education shares. To the extent that we mis-measure labor connections among Indian industries due to labor market differences across countries, this measurement error will downward bias our estimates toward finding no effect.

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 (e.g. Rosenthal and Strange, 2001; Duranton and Overman, 2005; Kerr and Kominsers, 2010). Second, the metrics do not consider final consumers. In unconditional estimates, we separately model city populations and gender ratios to capture some of the final consumer impetus. Third, the metrics do not measure quality differences across districts in inputs beyond basic traits such as education levels. Fourth, the metrics are built upon a survey of firms,

14 See Overman and Puga (2010) and Eriksson and Lindgren (2009) for recent theoretical and empirical evidence. Our working paper contains a more complete reference set of underlying theories.

rather than full census of businesses. This may introduce measurement error into their levels.¹⁵ 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 lagged industrial conditions as instruments to partially address this concern.¹⁶

4. Analysis of relative female entrepreneurship rates

We first characterize relative female entrepreneurship rates through a series of unconditional linear regressions with the above determinants as explanatory variables. Table 4 considers the 2000 and 2005 manufacturing surveys using a specification of the form:

$$\text{Female Entry}\%_{\text{dit}} = \eta_{it} + \beta \cdot X_d + \gamma \cdot Z_{\text{dit}} + \varepsilon_{\text{dit}}.$$

The outcome variable Female Entry%_{dit} is the ratio of female-owned young establishments in the district-industry-year to the sum of female- and male-owned young establishments. This measure utilizes sampling weights to yield population-level estimates. This outcome variable is the most intuitive, and we focus on a share outcome that is independent of the overall level of entrepreneurship in the district-industry. We find similar results when using variants such as the log female share or examining the log levels of female entrepreneurship directly. We exclude district-industries where we do not observe any entry, male or female, as the female fraction is undefined in these cases. We consider below a variety of robustness checks on this approach.¹⁷

While presenting results for female entrepreneurship ratios, we find very comparable patterns when examining female business-ownership ratios. We focus on the entrepreneurship rates since we can exclude entrepreneurs from the incumbent structures and directly circumvent some endogeneity concerns that would exist with overall business-ownership rates. We return to the endogeneity issue after reviewing our base results.

We include in each estimation a vector of industry-year fixed effects η_{it} that control for fixed differences in industry sizes, entrepreneurship rates, competition, and so on within each survey. These industry-year fixed effects also control for aggregate gender balances exhibited in Table 3. The vector X_d includes district traits such as population

15 This sampling issue is encountered in two ways. First, incomplete coverage in the surveys yields classical measurement error in our incumbent conditions. Second, non-classical measurement error is possible and may bias least squares results upward or downward. The issue descends from our measurement of entrants and incumbents from the same survey. Although these two groups are mutually exclusive, it is possible that clustered sampling artificially leads us to observe more of both entities in some areas. Population weights hopefully correct for this, but concern may persist. Note, however, that the bias could alternatively be downward due to limited sampling. In the limit, for example, if you only sample one entity from each district-industry, you cannot sample both an incumbent and an entrant. Our instrumental variables analysis will be particularly useful in this regard as it is a projection of recent entry levels on the industrial traits measured by a prior survey, removing this sampling issue.

16 There are several factors that we do not consider in this study: natural cost advantages, local industrial diversity, knowledge flows and entrepreneurial culture (Falck et al. 2011). Our working paper provides references on these factors. These exclusions are in part due to data constraints for India and in part due to our desire to maintain a consistent empirical framework between manufacturing and services.

17 We also exclude in the manufacturing estimates the district-industries that are not present in 1994. This is done to keep a consistent sample size between the least squares and instrumental variable specifications. We find very similar results when using the full sample of potential district-industries where entry occurs.

Table 3. Industry traits for unorganized manufacturing and unorganized services

Industry	Male establishment counts		Female establishment counts		Female establishment count share		Female establishment share, employee wid.	
	2000	2005	2000	2005	2000	2005	2000	2005
	Industry traits for unorganized manufacturing ^a							
15 Food products and beverages	2,465,014	1,943,115	307,284	380,881	0.11	0.16	0.09	0.11
16 Tobacco products	707,281	760,128	1,346,440	1,967,726	0.66	0.72	0.52	0.63
17 Textiles	1,319,150	1,190,366	647,578	831,089	0.33	0.41	0.19	0.23
18 Wearing apparel; dressing and dyeing of fur	1,812,817	1,632,443	887,051	1,427,978	0.33	0.47	0.22	0.35
19 Leather tanning; luggage, handbags, footwear	159,303	122,752	5931	11,731	0.04	0.09	0.04	0.06
20 Wood and wood products; straw and plating articles	2,160,422	1,396,976	432,969	347,713	0.17	0.20	0.13	0.19
21 Paper and paper products	46,768	34,381	36,140	124,301	0.44	0.78	0.27	0.62
22 Publishing, printing and media reproduction	120,983	94,498	8558	9217	0.07	0.09	0.07	0.08
23 Coke, refined petroleum and nuclear fuel	6458	2146	94	96	0.01	0.04	0.02	0.09
24 Chemicals and chemical products	43,856	64,077	154,997	316,288	0.78	0.83	0.49	0.63
25 Rubber and plastic products	59,118	47,792	19,631	12,659	0.25	0.21	0.13	0.14
26 Other non-metallic mineral products	717,221	532,770	32,448	30,443	0.04	0.05	0.04	0.04
27 Basic metals	33,516	28,749	496	1293	0.01	0.04	0.03	0.07
28 Fabricated metal products, except machinery	586,980	568,258	10,783	11,000	0.02	0.02	0.03	0.03
29 Machinery and equipment, not elsewhere classified (n.e.c.)	147,769	155,131	3222	3646	0.02	0.02	0.03	0.03
30 Office, accounting and computing machinery	127	785	0	0	0.00	0.00	0.00	0.00
31 Electrical machinery and apparatus, n.e.c.	55,978	95,954	2966	4257	0.05	0.04	0.04	0.06
32 Radio, television and communication equipment	5406	3973	262	493	0.05	0.11	0.05	0.26
33 Medical, precision and optical instruments, watches	7151	8597	147	406	0.02	0.05	0.02	0.06
34 Motor vehicles, trailers and semi-trailers	18,357	12,382	314	195	0.02	0.02	0.02	0.03
35 Other transport equipment	14,778	15,522	114	7918	0.01	0.34	0.01	0.49
36 Furniture, manufacturing, n.e.c.	1,016,010	865,217	158,410	150,557	0.13	0.15	0.10	0.09
Totals and weighted averages	11,504,463	9,576,012	4,055,836	5,639,889	0.26	0.37	0.17	0.25
Unweighted averages	522,930	435,273	184,356	256,359	0.16	0.22	0.12	0.19

(continued)

Table 3. Continued

Industry	Male establishment counts		Female establishment counts		Female establishment count share		Female establishment share, employee wtd.	
	2001	2006	2001	2006	2001	2006	2001	2006
Industry traits for unorganized services ^b								
55 Hotels and restaurants	1,666,165	1,582,458	197,073	168,755	0.11	0.10	0.09	0.09
60 Land transport (via pipelines)	4,194,370	3,811,346	18,531	25,541	0.00	0.01	0.01	0.01
61 Water transport	9065	5675	0	58	0.00	0.01	0.00	0.01
63 Supporting transport activities, travel agencies	74,949	92,709	3346	1863	0.04	0.02	0.05	0.02
64 Post and telecommunications	484,359	1,648,894	83,875	165,851	0.15	0.09	0.15	0.09
70 Real estate activities	72,580	162,973	2336	20,200	0.03	0.11	0.04	0.10
71 Renting of machinery, personal goods	470,773	422,415	12,773	11,678	0.03	0.03	0.02	0.02
72 Computer and related activities	12,364	34,556	2193	1359	0.15	0.04	0.16	0.04
73 Research and development	259	n.a.	0	n.a.	0.00	n.a.	0.00	n.a.
74 Other business activities	520,771	551,167	27,969	22,197	0.05	0.04	0.06	0.04
80 Education	622,552	464,531	265,978	258,146	0.30	0.36	0.27	0.33
85 Health and social work	1,090,858	809,959	119,451	121,294	0.10	0.13	0.10	0.13
90 Sewage and refuse disposal, sanitation, etc.	5510	26,764	37,607	20,374	0.87	0.43	0.76	0.41
91 Activities of membership organizations, n.e.c.	43,446	323,113	1089	1213	0.02	0.00	0.02	0.00
92 Recreational, cultural and sporting activities	305,105	188,217	4342	6869	0.01	0.04	0.01	0.03
93 Other service activities	2,239,510	2,015,968	236,174	286,406	0.10	0.12	0.09	0.12
Totals and weighted averages	11,812,636	12,140,746	1,012,738	1,111,805	0.08	0.08	0.08	0.08
Unweighted averages	738,290	809,383	63,296	74,120	0.12	0.10	0.11	0.10

^aAuthors' calculations based on National Sample Survey data (various rounds).

^bAuthors' calculations based on National Sample Survey data (various rounds). The research and development industry is not collected in 2006.

and education levels. These traits are measured in 2001 and do not vary over our sample period. Finally, Z_{dit} is a vector of incumbent employment levels and agglomeration metrics that vary by district-industry-year and are calculated using incumbent establishments only. We transform explanatory variables that do not have a logarithm scale to have unit standard deviation to aid interpretation, and we cluster standard errors by district to reflect the multiple mappings of some variables.

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 are very small and experience very limited entry. We recognize, however, that weighted estimations may accentuate endogeneity concerns. We thus employ our interaction of aggregate district and industry size rather than observed district-industry size. The interaction minimizes any endogeneity in highly agglomerated district-industries, especially in conditional estimations with district and industry fixed effects. We report below very similar effects without sample weights, indicating that these choices are not very material.

Table 4 provides our basic spatial results for manufacturing, before conditioning with district-year fixed effects. All regressions control for log total incumbent employment in the district-industry at the time of the survey and our basic district-level traits X_d . In terms of the agglomeration variables Z_{dit} , Column 3 further controls for log female incumbent business-ownership counts in the focal district-industry, Column 4 controls for log female incumbent business-ownership counts in the district as a whole and Column 5 controls for both of these measures. At the bottom of the table, the first estimation considers general Marshallian connections defined over all incumbent firms in the district, while Columns 2–5 use industrial conditions defined over female incumbent firms specifically. Column 6 repeats Column 5 without the estimation weights.

Beginning with Column 1, without controlling for female-specific agglomeration levels, a district-industry with more incumbent employment has a greater female entry share. Among general district traits, four factors are associated with a higher female entrepreneurship share in unorganized manufacturing: a higher female-to-male sex ratio, an age profile emphasizing working ages, better quality infrastructure and more stringent labor regulations. The relative entry rate, on the other hand, is lower in places with high population density and/or with more strongly developed banking sectors.¹⁸ Education levels and female literacy rates are not associated with greater entry shares in manufacturing. The regressions also include unreported controls for log district population and its square, total fertility rate and travel time to one of the 10 biggest cities in India. None of these variables are economically or statistically significant. Finally, strong input–output conditions in the district for the industry studied are linked to higher female entry ratios.

The predictive powers of these basic district traits do not vary much across the five columns as we adjust the agglomeration metrics. Among the basic district traits, the infrastructure correlation is the most prominent and policy relevant. Ghani et al. (2011b) and Mukim (2011) link higher quality infrastructure to greater entry rates in India. The finding that better infrastructure is associated with greater shares of female

18 This latter effect holds conditional on the other covariates. As a univariate correlation, the development of the local bank sector is not associated with material differences in the gender ratio of entrepreneurs.

Table 4. Unconditional estimations for manufacturing sector

Explanatory Variable	DV: Female-owned entrant share					
	(1)	(2)	(3)	(4)	(5)	(6)
Log total incumbent employment in district-industry	0.008++ (0.003)	0.002 (0.003)	-0.007++ (0.003)	-0.001 (0.003)	-0.008+++ (0.003)	-0.007+++ (0.002)
Log female-owned incumbent businesses in district-industry			0.027+++ (0.002)		0.025+++ (0.002)	0.025+++ (0.002)
Log female-owned incumbent businesses in district				0.029+++ (0.005)	0.020+++ (0.004)	0.019+++ (0.004)
District traits						
Female literacy rate	0.007 (0.011)	0.007 (0.011)	0.002 (0.010)	0.006 (0.010)	0.002 (0.009)	0.002 (0.009)
Sex ratio	0.025+++ (0.007)	0.025+++ (0.007)	0.020+++ (0.006)	0.019+++ (0.007)	0.016+++ (0.006)	0.015+++ (0.006)
Population density	-0.017+ (0.010)	-0.019++ (0.010)	-0.015+ (0.009)	-0.025+++ (0.009)	-0.019++ (0.008)	-0.019++ (0.008)
Education level	0.010 (0.009)	0.012 (0.009)	0.008 (0.008)	0.011 (0.008)	0.007 (0.007)	0.009 (0.007)
Age profile	0.027+ (0.014)	0.026+ (0.014)	0.020 (0.012)	0.018 (0.012)	0.015 (0.011)	0.015 (0.011)
Infrastructure level	0.032+++ (0.009)	0.037+++ (0.010)	0.033+++ (0.009)	0.031+++ (0.009)	0.029+++ (0.008)	0.028+++ (0.008)
Labor regulations stringency	0.020+++ (0.007)	0.020+++ (0.007)	0.016+++ (0.006)	0.015++ (0.006)	0.013++ (0.006)	0.011++ (0.005)
Share of households with banking	-0.176+++ (0.051)	-0.182+++ (0.052)	-0.138+++ (0.046)	-0.128++ (0.052)	-0.105++ (0.048)	-0.119+++ (0.046)
Local industrial conditions by incumbent firms						
Index of labor market strength for district-industry	-0.008 (0.008)					
Index of input-output strength for district-industry	0.020++ (0.010)					
Index of labor market strength, female-owned businesses		0.024+++ (0.007)	0.003 (0.006)	0.027+++ (0.007)	0.006 (0.006)	0.005 (0.006)
Index of input-output strength, female-owned businesses		0.032+++ (0.008)	0.011+ (0.006)	0.030+++ (0.007)	0.011+ (0.006)	0.010+ (0.006)
Sample weights	Yes	Yes	Yes	Yes	Yes	No
Observations	4336	4336	4336	4336	4336	4336
Adjusted R ²	0.309	0.321	0.347	0.329	0.351	0.336

Notes: Regressions consider relative rates of female entrepreneurship across manufacturing district-industries in India in 2000 and 2005. Regressions include industry-year fixed effects. Regressions include unreported controls for log district population and its square, total fertility rate, and travel time to one of the 10 biggest cities in India. None of these variables are economically or statistically significant. Regressions weight observations by an interaction of log district size and log industry size, excepting Column 6. Regressions cluster standard errors by district.

entry in particular is new. Lack of access to certain types of infrastructure services (transport, access to water and sanitation) seems to affect women more than men, perhaps because women typically bear a larger share of the time and responsibility for household maintenance and care activities. It is notable that while the within-district infrastructure quality is prominent, the strength of linkages across cities is not found to influence the gender balance.

Unreported specifications that disaggregate the infrastructure index stress that transport infrastructure and paved roads within villages are especially important. There may be several factors behind the within-district association. In terms of transport infrastructure, travel in India can be limited, dangerous and unpredictable, and women face greater constraints in geographic mobility imposed by safety concerns and/or social norms. Areas with better transport infrastructure may alleviate one of the major constraints to female entrepreneurs in accessing markets. In addition, better electricity and water access may reduce the burden of women in providing essential household inputs for their families and allow for more time to be directed toward entrepreneurial activities.

The positive association for stringent labor regulations is interesting as well. Several studies find that strict labor regulations suppress Indian entrepreneurship generally, especially in the formal sector. One channel through which these regulations could affect the gender balance of entrepreneurs is by shifting activity into industries that female entrepreneurs tend to be more involved in (Klapper and Parker, 2011). Although our industry fixed effects capture broad movements of this type, there may be subtler shifts at further levels of industry disaggregation that we do not observe. More intriguing, labor constraints in the organized sector may also shift the occupational decisions of men and women within the family. We do not investigate this further, given that the focus of our article is on the networks evident in local industrial structures, but the partial correlation is worthy of additional research.

Turning to the agglomeration metrics, the association of higher female entry shares to greater total incumbent employment in the first row disappears in Column 2 once we control for local conditions provided by incumbent female-owned businesses. These incumbent conditions suggest that female connections in labor markets or input–output markets both contribute to a higher entry share. These agglomeration indices are normalized to have unit standard deviation. Thus, a one-standard deviation increase in either of these incumbent conditions correlates with a 2–3% increase in the share of new entrants that are female. This compares to a base female entry ratio of 21%. There exists an important connection of past female business-ownership to future female entry rates.

In Columns 3–5, we add measures of the log count of incumbent female-owned businesses in the district and in the focal district-industry. These agglomeration levels again connect to higher relative entry rates for female entrepreneurs. A 10% increase in either measure, holding everything else constant, correlates with a 0.2–0.3% increase in the female entry share. Controlling for the overall size of incumbent female-owned businesses in the focal district-industry substantially accounts for the Marshallian linkages at the bottom of the table. On the other hand, controlling for the log count of the district's total incumbent female-owned businesses does not diminish the Marshallian metrics. If one drops the total incumbent employment regressor from the estimation in Column 5, the remaining variables have very similar values, with the Marshallian labor factor growing slightly in importance. Across the columns of Table 4, the adjusted R^2 value increases from 0.31 to 0.35.

Table 5 considers a similar set of estimations with the services sector, although we do not model Marshallian linkages for the services sector. In Table 7, we present conditional estimations for the services sector that includes labor market indices. These results show a positive labor spillover effect from incumbent female business ownership, but this benefit operates mostly through incumbent female businesses in the focal

Table 5. Unconditional estimations for services sector

Explanatory Variable	DV: Female-owned entrant share					
	(1)	(2)	(3)	(4)	(5)	(6)
Log total incumbent employment in district-industry	0.002 (0.002)	-0.003 (0.002)	0.001 (0.002)	-0.003 (0.002)	-0.009+++ (0.003)	-0.008+++ (0.002)
Log female-owned incumbent businesses in district-industry		0.013+++ (0.002)		0.012+++ (0.002)	0.012+++ (0.002)	0.012+++ (0.002)
Log female-owned incumbent businesses in district			0.011+++ (0.004)	0.006+ (0.004)	0.007+ (0.004)	0.009++ (0.004)
District traits						
Female literacy rate	0.010 (0.006)	0.008 (0.006)	0.010 (0.006)	0.008 (0.006)	0.010 (0.006)	0.010 (0.007)
Sex ratio	0.001 (0.004)	0.000 (0.004)	0.001 (0.004)	-0.000 (0.004)	-0.002 (0.004)	-0.003 (0.005)
Population density	-0.017+++ (0.006)	-0.015+++ (0.006)	-0.016+++ (0.005)	-0.015+++ (0.006)	-0.012++ (0.005)	-0.013++ (0.005)
Education level	0.010+ (0.005)	0.007 (0.005)	0.007 (0.005)	0.005 (0.005)	0.003 (0.005)	0.001 (0.005)
Age profile	0.018+ (0.010)	0.015 (0.010)	0.018+ (0.010)	0.015 (0.010)	0.017+ (0.010)	0.019+ (0.010)
Infrastructure level	0.018+++ (0.006)	0.015+++ (0.006)	0.014++ (0.006)	0.013++ (0.006)	0.010+ (0.006)	0.011+ (0.006)
Labor regulations stringency	0.010++ (0.005)	0.009+ (0.005)	0.010++ (0.005)	0.010++ (0.005)	0.010++ (0.005)	0.010++ (0.005)
Share of households with banking	0.029 (0.034)	0.032 (0.033)	0.036 (0.033)	0.036 (0.033)	0.046 (0.033)	0.060+ (0.033)
Sector	Informal	Informal	Informal	Informal	All	All
Sample weights	Yes	Yes	Yes	Yes	Yes	No
Observations	4292	4292	4292	4292	4458	4458
Adjusted R^2	0.224	0.232	0.225	0.233	0.232	0.220

Notes: Regressions consider relative rates of female entrepreneurship across services district-industries in India in 2001. Regressions include industry fixed effects. Regressions include unreported controls for log district population and its square, total fertility rate, and travel time to one of the 10 biggest cities in India. None of these variables are economically or statistically significant. Regressions weight observations by an interaction of log district size and log industry size, excepting Column 6. Regressions cluster standard errors by district.

industry. We thus present here the more parsimonious specification. We do not consider input–output conditions for the services sector given the difficulty of developing accurate input–output connections among services businesses.

There are some key similarities in Table 5 to Table 4: infrastructure quality, labor regulation stringency and age profiles again predict higher female entry shares, whereas population density is associated with a lower rate. Somewhat surprisingly, the association of a higher female entry ratio to a greater female sex ratio in the district that was present for manufacturing is not present for services. On the other hand, the female literacy rate and general education levels are marginally more predictive. This link may be due to the services sector being more skill intensive than the manufacturing sector in India (Ghani, 2010). Finally, the banking variable has a positive coefficient, compared with the negative relationship evident in manufacturing. Interestingly, this sector-level pattern mirrors the pattern observed in the USA by Rosenthal and Strange

Table 6. Conditional OLS estimations for manufacturing sector

Explanatory Variable	DV: Female-owned entrant share								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log total incumbent employment in district-industry	0.010+++ (0.003)	0.005 (0.003)	-0.000 (0.003)	0.003 (0.003)	-0.001 (0.003)	-0.006+ (0.003)	-0.006+ (0.003)	-0.007++ (0.003)	-0.006++ (0.003)
Log female-owned incumbent businesses in district-industry						0.027+++ (0.002)	0.027+++ (0.002)	0.023+++ (0.002)	0.023+++ (0.002)
Local industrial conditions by incumbent firms									
Index of labor market strength for district-industry		0.002 (0.008)		-0.026+++ (0.009)			-0.010 (0.008)		
Index of input-output strength for district-industry		0.023++ (0.011)		0.006 (0.012)			0.015 (0.011)		
Index of labor market strength, female-owned businesses			0.036+++ (0.006)	0.044+++ (0.007)	0.019+++ (0.007)			0.016++ (0.006)	0.015++ (0.006)
Index of input-output strength, female-owned businesses			0.027+++ (0.007)	0.025+++ (0.008)				0.010 (0.007)	0.010 (0.007)
Index of input market strength, female-owned businesses					0.137+++ (0.033)				
Index of output market strength, female-owned businesses					0.010+++ (0.004)				
Sample weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Observations	4336	4336	4336	4336	4336	4336	4336	4336	4336
Adjusted R ²	0.321	0.322	0.339	0.341	0.343	0.355	0.355	0.357	0.342

Notes: Regressions consider relative rates of female entrepreneurship across manufacturing district-industries in India in 2000 and 2005. Conditional estimations include district-year and industry-year fixed effects. Regressions weight observations by an interaction of log district size and log industry size, excepting Column 9. Regressions cluster standard errors by district-industry.

(2012), who also find female entrepreneurship in manufacturing is segmented from better banking markets, while services is less influenced.

The first four columns are for the unorganized sector, whereas Column 5 repeats the full specification with the complete services sample, finding similar results. Column 6 also shows similar results when dropping the estimation weights.

Examining the agglomeration metrics, the total size of the district-industry is not a factor for gender ratios in services. On the other hand, higher log counts of female-owned incumbent businesses in the district and district-industry again predict a greater female share. The coefficients are about half of their size in manufacturing, which we further confirm below in conditional estimations. As the overall female entrepreneurship share in services is lower than in manufacturing, at 11% versus 21% across the district-industry sample, this suggests that the effects are roughly similar in terms of proportions. The explanatory power of the estimations is slightly lower with an adjusted R^2 value of about 0.22.

Before proceeding further, it is worth pausing to emphasize again that Tables 4 and 5 represent partial correlations rather than a causal framework. This caveat is partly due to the timing of variables (e.g. using district traits from the 2001 Census to predict entry contemporaneously). But the issue extends deeper in that we do not have appropriate instruments for district-level traits such as education, and yet the scope for omitted factors is substantial. We thus view Tables 4 and 5 as important and informative correlations, but they are not causal parameters. We next focus on conditional estimations of industrial conditions where a causal assessment is feasible.

Tables 6 and 7 present conditional estimations of the form:

$$\text{Female Entry}\%_{\text{dit}} = \pi_{\text{dt}} + \eta_{\text{it}} + \gamma \cdot Z_{\text{dit}} + \varepsilon_{\text{dit}}.$$

We now include a vector of district-year fixed effects π_{dt} that controls for differences across districts that are common for all industries. Specifications thus employ within variation: how much of the unexplained district-industry variation in female entrepreneurship can we explain through local conditions that are especially suitable for particular industries? We no longer include the vector X_{d} of district-level traits as they are controlled for by the district fixed effects. To reflect the change in focus, we cluster standard errors by district-industry. Table 6 considers the manufacturing sample, and Table 7 considers services.

The conditional results in Table 6 confirm the earlier unconditional results in a more precise manner. We continue to find an important link between the log count of incumbent female businesses in a district-industry and the subsequent gender ratio for entrants, even after controlling for industry-year fixed effects, district-year fixed effects and the total size of the district-industry by year. Marshallian linkages across industries from incumbent female-owned businesses are also present, and they persist when controlling for total Marshallian linkages. These inter-industry links become less powerful once controlling for the female incumbent count in the focal district-industry.

In most of our estimations, we use a combined input–output metric to model local conditions. We combine these metrics in anticipation of our upcoming instrumental variable estimations, where the combined average metric has less measurement error and is easier to identify in 1994 industrial conditions. In Column 5 of Table 6, we separate the input and output metrics. The input metric is the stronger of the two, both in terms of economic significance and statistical significance. This importance of the

Table 7. Conditional OLS estimations for services sector

Explanatory Variable	DV: Female-owned entrant share								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Estimations using unorganized services sample									
Log total incumbent employment in district-industry	0.002 (0.003)	0.003 (0.003)	0.000 (0.003)	0.002 (0.003)	-0.003 (0.003)	-0.003 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
Log female-owned incumbent businesses in district-industry					0.014+++ (0.002)	0.012+++ (0.002)	0.012+++ (0.002)	0.012+++ (0.003)	0.012+++ (0.003)
Local industrial conditions by incumbent firms									
Index of labor market strength for district-industry		-0.006 (0.005)		-0.017+++ (0.005)			-0.016+++ (0.005)	-0.014+++ (0.005)	-0.014+++ (0.005)
Index of labor market strength, female-owned businesses			0.014+++ (0.004)	0.018+++ (0.004)		0.004 (0.004)	0.008+ (0.004)	0.008+ (0.005)	0.008+ (0.005)
Sample weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	4292	4292	4292	4292	4292	4292	4292	4292	4292
Adjusted R ²	0.232	0.232	0.235	0.237	0.241	0.241	0.242	0.242	0.244
B. Estimations using complete services sample									
Log total incumbent employment in district-industry	-0.005+ (0.003)	-0.005 (0.003)	-0.006+ (0.003)	-0.005+ (0.003)	-0.009+++ (0.003)	-0.009+++ (0.003)	-0.008+++ (0.003)	-0.007+++ (0.003)	-0.008+++ (0.003)
Log female-owned incumbent businesses in district-industry					0.013+++ (0.002)	0.012+++ (0.002)	0.011+++ (0.002)	0.011+++ (0.002)	0.014+++ (0.002)
Local industrial conditions by incumbent firms									
Index of labor market strength for district-industry		0.000 (0.006)		-0.008 (0.007)			-0.007 (0.007)	-0.007 (0.005)	-0.007 (0.005)
Index of labor market strength, female-owned businesses			0.013+++ (0.003)	0.015+++ (0.004)		0.006 (0.004)	0.007+ (0.004)	0.007+ (0.004)	0.007+ (0.004)
Sample weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Exact sample as in Panel A	No	No	No	No	No	No	No	No	Yes
Observations	4458	4458	4458	4458	4458	4458	4458	4458	4292
Adjusted R ²	0.218	0.217	0.221	0.221	0.226	0.226	0.226	0.229	0.240

Notes: Regressions consider relative rates of female entrepreneurship across services district-industries in India in 2000. Conditional estimations include district and industry fixed effects. Regressions weight observations by an interaction of log district size and log industry size, excepting Column 9. Regressions report robust standard errors.

input metric also holds when controlling for the log female incumbent business count in the focal district-industry.

Table 7 analyzes the conditional estimates for the services sector. We again find a strong link from past district-industry female incumbent businesses to the gender ratio of subsequent entrants. The elasticity in Column 5 of Table 7 is again about half of that evident in Column 5 of Table 6. Given that the female entrant share is about twice as big in manufacturing as in services, this suggests that the two agglomeration effects are roughly comparable in relative terms. A similar pattern is evident in the total services sample. Finally, similar to manufacturing, we find evidence for Marshallian labor market connections being important. This connection operates more through the incumbent presence within the industry in question for services.

The patterns in Tables 4–7 are robust to many specification variants in addition to the variations described above when outlining our data and empirical strategy. One important test is that we find similar results when restricting to district-industries that have more than 50,000 employees, suggesting these outcomes are not sensitive to small sample sizes of female and male establishments. 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.

We have also undertaken several checks in specification design, including a modified form of the Rosenthal and Strange (2012) differencing methodology. The most important difference across our studies is that Rosenthal and Strange (2012) are able to examine agglomeration effects that operate within a one-mile radius of a Census tract, whereas we consider spatial units that are on average the size of two U.S. counties. As a consequence, we are not able to replicate their framework exactly. We can, however, borrow a piece of their methodology to further establish comparability. The final and most stringent analysis conducted by Rosenthal and Strange (2012) is a differenced estimation of the form:

$$\ln(\text{Female Entry}_{\text{dit}}/\text{Male Entry}_{\text{dit}}) = \pi_{\text{dt}} + \eta_{\text{it}} + \delta \cdot \ln(\text{Female Ownership}_{\text{dit}}) + \gamma \cdot \ln(\text{Male Ownership}_{\text{dit}}) + \varepsilon_{\text{dit}},$$

where we have combined their framework with the fixed effects for district-year and industry-year that we model in our conditional work.

Table 8 reports results for this estimation framework with our manufacturing and services samples. For each, we present three forms of the outcome variable. Within each triplet, we first show total employment as a single regressor, we then split employment in male- and female-owned establishments, and finally we add in the additional Marshallian metrics. These estimations strongly support our central results. In every case, the presence of female-owned businesses is strongly predicting a higher ratio of entry toward female entrepreneurs. We also find similar support when using incumbent business counts by gender, rather than incumbent employments, for explanatory variables.

5. Instrumental variable analysis using 1994 incumbent structures

Our final analysis turns to the question of identification. Our analysis thus far takes the incumbent industrial structures of districts as exogenous to predict the gender ratio of new entrants. There are several potential issues with this approach that could bias our

Table 8. OLS estimations of Rosenthal and Strange (2012) differencing analysis for manufacturing sector and services sector

Explanatory Variable	DV: ln(female-owned entrants/male-owned entrants)			DV: ln(employment in female-owned entrants/employment in male-owned entrants)			DV: female-owned entrants/male-owned entrants		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OLS estimations of Rosenthal and Strange (2012) differencing analysis for manufacturing sector ^a									
Log total incumbent employment in district-industry	0.008+++ (0.002)			0.006+++ (0.002)			0.010+++ (0.003)		
Log total incumbent employment in district-industry, female-owned	0.017+++ (0.001)	0.013+++ (0.002)		0.015+++ (0.001)	0.011+++ (0.002)	0.023+++ (0.002)	0.018+++ (0.002)		
Log total incumbent employment in district-industry, male-owned	-0.006+++ (0.002)	-0.007+++ (0.002)		-0.007+++ (0.002)	-0.007+++ (0.002)	-0.010+++ (0.003)	-0.011+++ (0.003)		
Local industrial conditions by incumbent firms									
Index of labor market strength for district-industry			-0.012+ (0.006)			-0.012+ (0.006)			-0.018+ (0.009)
Index of input-output strength for district-industry			0.014+ (0.008)			0.008 (0.009)			0.018 (0.012)
Index of labor market strength, female-owned businesses			0.019+++ (0.005)			0.019+++ (0.005)			0.026+++ (0.007)
Index of input-output strength, female-owned businesses			0.005 (0.005)			0.005 (0.005)			0.008 (0.007)
Observations	4336	4336	4336	4336	4336	4336	4336	4336	4336
Adjusted R ²	0.331	0.361	0.366	0.293	0.319	0.323	0.321	0.350	0.355
OLS estimations of Rosenthal and Strange (2012) differencing analysis for services sector ^b									
Log total incumbent employment in district-industry	0.002 (0.002)			0.002 (0.002)			0.002 (0.003)		
Log total incumbent employment in district-industry, female-owned	0.010+++ (0.001)	0.009+++ (0.002)		0.009+++ (0.001)	0.009+++ (0.002)	0.012+++ (0.002)	0.011+++ (0.002)		
Log total incumbent employment in district-industry, male-owned	-0.006+++ (0.002)	-0.005+++ (0.002)		-0.006+++ (0.002)	-0.005+++ (0.002)	-0.008+++ (0.003)	-0.007+++ (0.003)		
Local industrial conditions by incumbent firms									
Index of labor market strength for district-industry			-0.009+++ (0.004)			-0.008+++ (0.004)			-0.011+++ (0.005)
Index of labor market strength, female-owned businesses			0.006+ (0.003)			0.005+ (0.003)			0.008+ (0.004)
Observations	4292	4292	4292	4292	4292	4292	4292	4292	4292
Adjusted R ²	0.244	0.255	0.256	0.231	0.241	0.242	0.232	0.242	0.243

^aRegressions consider relative rates of female entrepreneurship across manufacturing district-industries in India in 2000 and 2005 with the Rosenthal and Strange (2012) differencing approach. Conditional estimations include district-year and industry-year fixed effects. Regressions weight observations by an interaction of log district size and log industry size. Regressions cluster standard errors by district-industry.

^bRegressions consider relative rates of female entrepreneurship across services district-industries in India in 2001 with the Rosenthal and Strange (2012) differencing approach. Conditional estimations include district and industry fixed effects. Regressions weight observations by an interaction of log district size and log industry size. Regressions report robust standard errors.

least squares results upward or downward. First, reverse causality may exist, where the anticipation of many female entrants encourages women in the prior period to own businesses to link to the new firms (Manski, 1993; Andersson and Koster, 2011). Second, omitted variable biases may be present. Our conditional estimations require that such omitted factors be specific to a district-industry. An example would be a very inspirational female business leader in the district-industry that encourages both past and future women in the district to engage in firm ownership. Alternatively, local policy initiatives may attempt to encourage women's entrepreneurship in sectors where women are under-represented and/or the incumbent structure is unsupportive. Finally, measurement error likely exists that biases our estimates of the importance of incumbent industrial structures. This measurement error could be due to the sampled micro-data (versus a census of local businesses) or incomplete metrics for how industries interact.

To analyze these challenges, we instrument in Tables 9–11 for incumbent industrial conditions in 2000 and 2005 using the industrial conditions that existed in 1994 in the manufacturing sector. The earliest year for which the female ownership question was asked is 1994. These data are only collected for the manufacturing sector.

Returning to the challenges facing the least squares estimates, the 1994 incumbent industrial conditions and the role of women are more exogenous to the entry conditions in 2000 and 2005 than the contemporaneous incumbent conditions. To the extent that reverse causality persisted, the anticipation of future entry would need to span about a decade in duration. Fernandes and Sharma (2011) discuss how the spatial locations of manufacturing firms in India's formal sector have adjusted substantially since the large-scale deregulations of the 1980s and 1990s. Prior to these deregulations, spatial location decisions for firms were set to a large degree by the government with the goal to promote general equality across regions. By reaching as far back into this regulated period as possible, we hopefully capture incumbent conditions that are not being determined by anticipation of female entry conditions after 2000.

For the omitted variable bias concern, the instrument partially helps. The instrument overcomes concerns of a special condition that emerged in a particular district-industry that favored female ownership (e.g. the inspirational female business leader, local government policies) if such an omitted factor is short term in duration. That is, the instrument addresses omitted factors that are localized to the 2000–2005 period and do not extend back to the early 1990s. To the extent that an omitted factor was very long in duration and specific to a district-industry (e.g. a local women's training institute that is specific to an industry), then the instrumental variable approach will not help if the factor was present in 1994, too.

We can test to some extent this residual concern. With our conditional framework, a key worry is situations where local norms or cultural differences create a permanent skewness in the role of women across the industrial structures of a district. Under such conditions, instruments that use 1994 industrial structures cannot correct for endogeneity issues present in 2000 and 2005. To build confidence in this dimension, we test whether the relative role of women's entrepreneurship in 1994 across industries within a district correlates systematically with other district traits. Strong correlations would cast doubt on the instrument's validity.

We first calculate the ratio of female entry rates by district in 1994 in industries with high rates of female business ownership nationally compared with industries with low rates of female ownership nationally. For this calculation, we divide industries at the

Table 9. Conditional IV estimations for manufacturing sector, first-stage results

Instrument	Incumbent industrial conditions in 2000 and 2005					
	Log total incumbent employment	Log female-owned incumbent businesses	Log total incumbent employment	Log female-owned incumbent businesses	Labor market strength, female-owned businesses	Input-output market strength, female-owned businesses
	(1)	(2)	(3)	(4)	(5)	(6)
Log total incumbent employment in district-industry in 1994	0.182 +++ (0.017)	0.096+++ (0.021)	0.180 +++ (0.017)	0.094+++ (0.021)	0.023++ (0.010)	0.031+++ (0.011)
Log female-owned incumbent businesses in district-industry in 1994	0.096+++ (0.014)	0.290 +++ (0.024)	0.073+++ (0.016)	0.263 +++ (0.028)	0.029+++ (0.011)	-0.001 (0.010)
Index of labor market strength, female-owned incumbent businesses in 1994			0.063++ (0.025)	0.101++ (0.041)	0.176 +++ (0.024)	0.031+++ (0.012)
Index of input-output strength, female-owned incumbent businesses in 1994			0.032 (0.048)	-0.012 (0.065)	0.026 (0.026)	0.199 +++ (0.068)
Observations	4336	4336	4336	4336	4336	4336
F statistic	105	101	56	53	31	17
Partial R ²	0.080	0.082	0.083	0.084	0.082	0.071
Shea Partial R ²	0.026	0.026	0.023	0.024	0.035	0.044

Notes: Estimations report first-stage results when instrumenting for district-industry incumbent industrial conditions in 2000 and 2005 with incumbent industrial conditions in 1994. Conditional estimations include district-year and industry-year fixed effects. Regressions weight observations by an interaction of log district size and log industry size. Regressions cluster standard errors by district-industry. IV estimations consider manufacturing sector only due to lack of 1994 data on services sector. The bold coefficients indicate the primary first-stage instruments.

Table 10. Conditional IV estimations for manufacturing sector, second-stage results

Explanatory Variable	DV: Female-owned entrant share					
	(1)	(2)	(3)	(4)	(5)	(6)
Log total incumbent employment in district-industry	-0.049+++ (0.018)	-0.020 (0.015)	-0.054+++ (0.019)	-0.049+++ (0.017)	-0.005+ (0.003)	-0.050+++ (0.017)
Log female-owned incumbent businesses in district-industry	0.065+++ (0.013)		0.058+++ (0.016)	0.054+++ (0.015)	0.027+++ (0.002)	0.066+++ (0.011)
Index of labor market strength, female-owned businesses		0.069++ (0.027)	0.014 (0.031)	0.016 (0.031)		
Index of input-output strength, female-owned businesses		0.055 (0.035)	0.035 (0.029)	0.044 (0.032)		
Log female-owned incumbent closures in district-industry					-0.003 (0.003)	-0.001 (0.003)
Estimation form	IV	IV	IV	IV	OLS	IV
Sample weights	Yes	Yes	Yes	No	Yes	Yes
Observations	4336	4336	4336	4336	4336	4336

Notes: Columns 1-4 report second-stage results when instrumenting for district-industry incumbent industrial conditions in 2000 and 2005 with incumbent industrial conditions in 1994. Conditional estimations include district-year and industry-year fixed effects. Regressions weight observations by an interaction of log district size and log industry size, excepting Column 4. Regressions cluster standard errors by district-industry. Table 9 reports first-stage results, and Table 6 reports OLS results. Columns 5 and 6 report OLS and IV results from an additional test when a metric of estimated female incumbent business closures between 1994 and 2000 is included as an additional regressor. IV estimations consider manufacturing sector only due to lack of 1994 data on services sector.

Table 11. IV estimations of Rosenthal and Strange (2012) differencing analysis

Explanatory Variable	DV: ln(female-owned entrants/male-owned entrants)	DV: ln(employment in female-owned entrants/employment in male-owned entrants)	DV: female-owned entrants/male-owned entrants
	(1)	(2)	(3)
Log total incumbent employment in district-industry, female-owned	0.043+++ (0.008)	0.040+++ (0.008)	0.059+++ (0.012)
Log total incumbent employment in district-industry, male-owned	-0.035+++ (0.011)	-0.038+++ (0.011)	-0.052+++ (0.015)
Sample weights	Yes	Yes	Yes
Observations	4336	4336	4336

Notes: Table reports second-stage results with the Rosenthal and Strange (2012) differencing approach when instrumenting for district-industry incumbent industrial conditions in 2000 and 2005 with incumbent industrial conditions in 1994. Conditional estimations include district-year and industry-year fixed effects. Regressions weight observations by an interaction of log district size and log industry size. Regressions cluster standard errors by district-industry. Table 8 reports OLS results. IV estimations consider manufacturing sector only due to lack of 1994 data on services sector.

median female business ownership rate nationally. Differences in this ratio reflect the degree to which female-owned businesses for a district are skewed toward or away from traditional women's sectors. Reflecting that many women own household-based businesses, we find similar metrics when using the extent to which industries nationally are operated in households. We also find similar results if we use relative rates of business ownership rather than relative entry rates.

We then test whether this skewness in 1994 correlates with other districts traits likely to cause or reflect differences in norms across districts for women's economic participation. The constructed entry metric is quite uncorrelated with our developed proxies for local norms. Caste and tribe differences are perhaps the strongest determinant of local norms in India, and Iyer et al. (2011b) find them to be important determinants for entrepreneurship in India. The relative entry rate's correlation with the district's population share from scheduled castes and tribes is just 0.02. Religion is another factor that shapes beliefs about the proper role for women, and the relative entry rate metric has -0.04 and 0.08 correlations with the share of the local population practicing Sikh and Islamic religions, respectively. The metric also has -0.01, 0.04 and -0.04 correlations with district female literacy rates, district fertility rates and district female labor force participation rates, respectively. As informative, the correlations remain very low at 0.03 and 0.05 when looking at the gaps in female literacy rates and labor force participation rates, respectively, compared with men in the district. These district-level traits are all taken from the 1991 Census. In addition to being economically small, these correlations are not statistically significant at a 10% level.

As a further test, we can use our NSSO data to calculate in 1994 the share of bank loans in the district made to female-owned firms. We use both the raw share and also a second version that is normalized by the overall share of female-owned firms in the district. These two traits give a sense of whether women are over- or under-represented in bank lending, perhaps reflective of general economic and business status in the local area. Again, the correlations with the relative entry rate metric are very low at -0.04

and -0.02 , respectively. These low correlations suggest that omitted local norms are not likely to be reflected in industry differentials within districts that could persist and bias our conditional instrumental variables estimations at the district-industry level.

To our third concern, the instrumental variables will help overcome measurement error in the agglomeration regressors. This measurement error is due to the sampled nature of our data and the imperfect design of our metrics. The instrumental variables can help relieve the standard downward bias in coefficient values that measurement error produces. On a related note, the instrumental variables approach can further help with potential concerns, highlighted in footnote 15, about measuring our entry rates and incumbent conditions from the same surveys. The data split that we use of young firms and incumbents has been used by multiple researchers on India, and we have not identified any sampling biases in this approach. Nonetheless, the instrumental variables approach identifies off of conditions from the 1994 survey to predict entry in 2000 and 2005.

Table 9 provides the first-stage results. The first two columns provide the first-stage results where we instrument for just the total incumbent employment in the district-industry and the count of female-owned incumbent firms in the district-industry. Columns 3–6 are the first-stages from an extended specification where we also include the two Marshallian indices. The construction of the Marshallian indices in 1994 mirrors that in the later surveys. The lagged conditions strongly predict the incumbent conditions in 2000 and 2005, with the strongest elasticity for each agglomeration metric being its direct counterpart in 1994. The instruments pass all weak instruments criteria and tests.

Table 10 reports the second-stage outcomes using our main specification. Column 1 presents results from just instrumenting for the total incumbent employment in the district-industry and the count of female-owned incumbent firms in the district-industry. The results confirm an important role for female-owned incumbent businesses in encouraging female entrepreneurs. The elasticity of 0.065 is about twice the least squares magnitude of 0.027 estimated in Column 6 of Table 6.

Column 2 presents the results from just instrumenting for the total incumbent employment in the district-industry and the two Marshallian indices. The first-stage results for this estimation are similar to those reported in Table 9. We again find evidence for important interactions through both channels. The instrumented elasticities are again about twice the magnitude of those estimated in Column 3 of Table 6. The labor market channel is statistically significant, whereas the input–output channel falls just short of being precisely measured at the 10% confidence level.

The third column of Table 10 instruments for all four agglomeration regressors, and the fourth column repeats this specification without the estimation weights. We continue to find that the log count of incumbent female-owned businesses in the district-industry is the key factor for encouraging a higher female entrepreneurship share in the district-industry. Once controlling for this core measure, the labor market channel is not found to be of further economic or statistical importance. The input–output measure maintains a strong point estimate, but it is not precisely measured. These results compare to Column 8 of Table 6, and one would not reject a null hypothesis that the least squares results are correctly estimated.

Table 11 presents instrumental variables results when using the Rosenthal and Strange (2012) differencing methodology. We instrument for the incumbent employment in female- and male-owned establishments in the district-industry by the levels of

these variables in 1994. The unreported first-stage estimations remain very strong and are available upon request. These results paint a very similar picture to Table 10, with female incumbent businesses being very important for subsequent entry rates. We have also examined instrumenting for the Marshallian interactions. Similar to Columns 3 and 4 of Table 10, these inter-industry interactions are not significant once the focal industry's main effect has been instrumented for.

Overall, these instrumental variable specifications support the conclusion that female entrepreneurship follows from incumbent female-owned businesses in a district-industry that encourage subsequent entry. Marshallian channels are important, but they mostly appear to be operating through the district-industry agglomeration for female business owners. Although our approach does not rule out every potential bias that may exist, it does suggest that the most worrisome endogeneity or omitted factors are not behind our least squares estimates. The results also suggest that measurement error or an omitted factor biased downward the least squares elasticities. As the instrument focuses attention on the very persistent parts of the incumbent industrial distribution, the associated local average treatment effect may also be higher than that present in transitory components.

These instrumental variables results are also supported by some simple placebo exercises. First, if it is true that female incumbent businesses provide a special impetus for future female entrepreneurs, forward entry rates for male-owned businesses should not be predictive of incumbent women's industrial structures, excepting perhaps some small crowd-out effects. As a test of this logic, we regress the log count of female-owned incumbent businesses on the log count of male-owned entering businesses and the log employment of the district-industry. The estimated coefficients are -0.031 (0.023) and 0.564 (0.026), respectively.

Second, our instrumental variables strategy proposes that the industrial conditions in 1994 are only impacting entrepreneurship in 2000 and 2005 through how the 1994 conditions shaped the incumbent firm structure in 2000 and 2005. If this is true, we should not expect to see predictive power for businesses that existed in 1994 but were closed by 2000. Unfortunately, our data are repeated cross-sections, and thus we cannot measure closures exactly. We can, however, calculate a proxy for closures as $\max(\text{total business count in 1994} - \text{incumbent count in 2000}, 0)$. This calculation models that in cases of substantial district growth, we do not know if any closures occurred. But we can estimate the net change in cases with more limited growth.

In the last two columns of Table 10, we report least squares and instrumental variables estimations that include this bounded change measure as an additional regressor. Due to the log zero issue, we add one to the value before taking logs so that its form matches our other variables. These controls have no effect on our results. We show this result with closures for female-owned businesses, and we find the same result when also controlling for closures of male-owned businesses. This provides some assurance for the instrumental variables results acting through the proposed industrial legacies rather than other channels.

6. Conclusions

Economic growth and development depends upon successfully utilizing the workforce, both male and female. Despite its recent economic advances, India's gender balance for

entrepreneurship remains among the lowest in the world. Improving this balance is an important step for India's development and its achievement of greater economic growth and gender equality. Although achieving economic equality sometimes requires tough choices (e.g. progressive taxation that may discourage effort), the opposite is true here. Unlocking female entrepreneurship will promote a broader dynamic economy and economic growth generally. This study quantifies the connection that female entrepreneurs have to favorable incumbent industrial structures. The central message is the high degree to which existing female business ownership enables future female entry.

More generally, the results of this article are encouraging for studies that model the incumbent industrial structures of cities and entry rates (e.g. Glaeser and Kerr, 2009; Jofre-Monseny et al., 2011; Dauth, 2011; Mukim, 2011; Ghani et al., 2011b). Earlier studies focus on linking the total entry rate in a city-industry to favorable incumbent conditions. This study has taken this conceptual device one step further by tracing out a specific set of entrepreneurship and localized interactions—that is, incumbent female business owners being especially helpful for higher relative rates of subsequent female entry. This study helps validate the technique with a more detailed application, and we hope that future research considers other ties among firms (e.g. racial ties and networks among past work colleagues). There appears to be strong potential for combining more detailed identities of workers and firms in studies of agglomeration economies.

Several important questions remain for future research. First, we need to identify the extent to which this female business concentration in India is due to exclusion/segregation versus choice. A complete analysis of this issue, especially in a developing economy, requires careful attention to time. The theory behind network dynamics and their economic efficiency is complex (e.g. Munshi and Rosenzweig, 2005, 2006; Munshi, 2011). Even if the initial female business concentrations were due to economic exclusion, they may be much more efficient today. The Indian economy provides interesting tests, but our data will likely need to be complemented to do so (e.g. we need better insights into whether the new female entrepreneurs are being spawned out of the existing female businesses).

Second, we need a better understanding of the output and income consequences for female business owners and entrepreneurs versus males. The basic statistics are not pretty. The value-add per worker in female-owned businesses in the manufacturing sector for 2000 and 2005 is roughly a third of that for male-owned business; in services, the female-to-male ratio is about one-half to two-thirds. These differentials exist in many states and industries, so simple explanations such as industry choice are incomplete. In current research, we are extending the district-industry framework to analyze the role of local industrial structures for these gaps. We also hope to link these features to political representation of women (e.g. Iyer et al., 2011a).

Finally, we need to develop a better understanding of how these gender networks influence aggregate efficiency. A number of studies estimate the economic returns to density and agglomeration (e.g. Ciccone and Hall, 1996; Rosenthal and Strange, 2004), and Behrens et al. (2011) and Venables (2011) provide recent models linking entrepreneur city selections to sorting and agglomeration economies. Ottaviano (2011) emphasizes how further refinements to our micro-foundations of economic geography models are needed to improve the macro-performance of these models. An important message of this article is that these linkages and spillovers across firms can depend a lot on common traits of business owners. Likewise, although not studied in this article, interactions between the informal and formal sectors may not be as strong

as interactions within each sector. Further research needs to identify how these economic forces vary by the composition of local industry. This will be especially helpful for evaluating the performance of industry concentrations in developing economies and guiding appropriate policy actions.

Acknowledgments

We thank Ahmad Ahsan, Muhamad Amin, Mehtabul Azam, Rajeev Dehejia, Arti Grover, Debasree Das Gupta, Lakshmi Iyer, Henry Jewell, Henry Overman, Arvind Panagariya, Agnes Quisumbing, Hyoung Gun Wang, and two referees 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 article. We are particularly indebted to Shanthi Nataraj for sharing her wisdom regarding the industrial survey data.

Funding

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.

References

- Aghion, P., Burgess, R., Redding, S., Zilibotti, F. (2008) The unequal effects of liberalization: evidence from dismantling the license raj in India. *American Economic Review*, 98: 1397–1412.
- Ahsan, A., Pages, C. (2007) *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.
- Akerlof, G., Kranton, R. (2000) Economics and identity. *Quarterly Journal of Economics*, 115: 715–753.
- Amin, M., Mattoo, A. (2008) *Human Capital and the Changing Structure of the Indian Economy*. World Bank Policy Research Working Paper 4576.
- Andersson, M., Koster, S. (2011) Sources of persistence in regional start-up rates—evidence from Sweden. *Journal of Economic Geography*, 11: 179–201.
- Ardagna, S., Lusardi, A. (2008) *Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints*. Working Paper.
- Behrens, K., Duranton, G., Robert-Nicoud, F. (2011) *Productive Cities: Sorting, Selection, and Agglomeration*. Working Paper.
- Besley, T., Burgess, R. (2004) Can labor regulation hinder economic performance? Evidence from India. *Quarterly Journal of Economics*, 119: 91–134.
- Bönte, W., Falck, O., Heblich, S. (2009) The impact of regional age structure on entrepreneurship. *Economic Geography*, 85: 269–287.
- Chari, A. V. (2008) *The Aggregate Productivity Effects of Entry and Output Restrictions: An Analysis of License Reform in India*. Working Paper.
- Chinitz, B. (1961) Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review*, 51: 279–289.
- Ciccone, A., Hall, R. (1996) Productivity and the density of economic activity. *American Economic Review*, 86: 54–70.
- Combes, P.-P., Duranton, G. (2006) Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36: 1–28.
- Dahl, M., Sorenson, O. (2007) *Home Sweet Home: Social Capital and Location Choice*. Working Paper.
- Datta, S. (2011) The impact of improved highways on Indian firms. *Journal of Development Economics*, 99: 46–57.
- Dauth, W. (2011) *The Mysteries of the Trade: Interindustry Spillovers in Cities*. Working Paper.

- Dehejia, R., Panagariya, A. (2010) *Services Growth in India: A Look Inside the Black Box*. Working Paper.
- Desmet, K., Ghani, E., O'Connell, S., Rossi-Hansberg, E. (2012) *The Spatial Development of India*. Working Paper.
- Doms, M., Lewis, E., Robb, A. (2010) Local labor force education, new business characteristics, and firm performance. *Journal of Urban Economics*, 67: 61–77.
- Drennan, M., Kelly, H. (2011) Measuring urban agglomeration economies with office rents. *Journal of Economic Geography*, 11: 481–507.
- Dufló, E. (2005) *Gender Equality in Development*. MIT Working Paper.
- Dufló, E. (2011) *Women Empowerment and Development*. NBER Working Paper 17702.
- Dunlop, J., Velkoff, V. (1999) *Women and the Economy in India*. U.S. Department of Commerce Report WID/98-2.
- Duranton, G., Overman, H. (2005) Testing for localization using micro-geographic data. *Review of Economic Studies*, 72: 1077–1106.
- Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In V. Henderson, J. F. Thisse (eds) *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2063–2117. Amsterdam: North-Holland.
- Ellison, G., Glaeser, E., Kerr, W. (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100: 1195–1213.
- Eriksson, R., Lindgren, U. (2009) Localized mobility clusters: impacts of labour market externalities on firm performance. *Journal of Economic Geography*, 9: 33–53.
- Estrin, S., Mickiewicz, T. (2011) Institutions and female entrepreneurship. *Small Business Economics*, 37: 397–415. (See also IZA Working Paper 4577.).
- Evans, D., Leighton, L. (1989) Some empirical aspects of entrepreneurship. *American Economic Review*, 79: 519–535.
- Falck, O., Fritsch, M., Heblich, S. (2011) The phantom of the opera: cultural amenities, human capital, and regional economic growth. *Labour Economics*, 18: 755–766.
- Fernandes, A., Pakes, A. (2010) *Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Survey Data*. Working Paper.
- Fernandes, A., Sharma, G. (2011) *Together We Stand? Agglomeration in Indian Manufacturing*. Working Paper.
- Figueiredo, O., Guimaraes, P., Woodward, D. (2002) Home-field advantage: location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, 52: 341–361.
- Figueiredo, O., Guimaraes, P., Woodward, D. (2009) Localization economies and establishment size: was Marshall right after all? *Journal of Economic Geography*, 9: 853–868.
- Ghani, E. (ed.) (2010) *The Services Revolution*. Oxford, UK: Oxford University Press.
- Ghani, E., Kerr, W., O'Connell, S. (2011a) Promoting entrepreneurship, growth, and job creation. In E. Ghani (ed.) *Reshaping Tomorrow*, pp. 168–201. Oxford, UK: Oxford University Press.
- Ghani, E., Kerr, W., O'Connell, S. (2011b) *Spatial Determinants of Entrepreneurship in India*. NBER Working Paper 17514.
- Glaeser, E., Kerr, W. (2009) Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? *Journal of Economics and Management Strategy*, 18: 623–663.
- Glaeser, E., Kerr, W., Ponzetto, G. (2010) Clusters of entrepreneurship. *Journal of Urban Economics*, 67: 150–168.
- Greenstone, M., Hornbeck, R., Moretti, E. (2010) Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118: 536–598.
- Hasan, R., Jandoc, K. (2010) *The Distribution of Firm Size in India: What Can Survey Data Tell Us?* ADB Economics Working Paper 213.
- Iyer, L., Mani, A., Mishra, P., Topalova, P. (2011a) *The Power of Political Voice: Women's Political Representation and Crime in India*. HBS Working Paper 11-092.
- Iyer, L., Khanna, T., Varshney, A. (2011b) *Caste and Entrepreneurship in India*. Working Paper.
- Jofre-Monseny, J., Marín-López, R., Viladecans-Marsal, E. (2011) The mechanisms of agglomeration: evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*, 70: 61–74.

- Kathuria, V., Natarajan, S., Raj, R., Sen, K. (2010) *Organized versus Unorganized Manufacturing Performance in India in the Post-Reform Period*. MPRA Working Paper No. 20317.
- Kerr, W., Kominers, S. (2010) *Agglomerative Forces and Cluster Shapes*. NBER Working Paper 16639.
- Khanna, T. (2008) *Billions of Entrepreneurs: How China and India Are Reshaping Their Futures—and Yours*. Boston, MA: Harvard University Press.
- Klapper, L., Amit, R., Guillén, M. F. (2010) Entrepreneurship and firm formation across countries. In J. Lerner, A. Schoar (eds) *International Differences in Entrepreneurship*, pp. 129–158. Chicago, IL: University of Chicago Press.
- Klapper, L., Parker, S. (2011) *Gender and Business Environment for New Firm Creation*. World Bank Research Observer.
- Klepper, S. (2010) The origin and growth of industry clusters: the making of Silicon Valley and Detroit. *Journal of Urban Economics*, 67: 15–32.
- Lall, S. (2007) Infrastructure and regional growth, growth dynamics and policy relevance for India. *The Annals of Regional Science*, 41: 581–599.
- Lall, S., Mengistae, T. (2005) *The Impact of Business Environment and Economic Geography on Plant Level Productivity: An Analysis of Indian Industry*. Working Paper.
- Lall, S., Shalizi, Z., Deichmann, U. (2004) Agglomeration economies and productivity in Indian industry. *Journal of Development Economics*, 73: 643–673.
- Lall, S., Wang, H., Deichmann, U. (2011) *Infrastructure and City Competitiveness in India*. Working Paper.
- Mammen, K., Paxson, C. (2000) Women's work and economic development. *Journal of Economic Perspectives*, 14: 141–164.
- Manski, C. (1993) Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60: 531–542.
- Marshall, A. (1920) *Principles of Economics*. London, UK: MacMillan and Co.
- Michelacci, C., Silva, O. (2007) Why so many local entrepreneurs? *Review of Economics and Statistics*, 89: 615–633.
- Minniti, M. (2010) Female entrepreneurship and economic activity. *European Journal of Development Research*, 22: 294–312.
- Minniti, M., Naudé, W. (2010) What do we know about the patterns and determinants of female entrepreneurship across countries? *European Journal of Development Research*, 22: 277–293.
- Mukim, M. (2011) *Industry and the Urge to Cluster: A Study of the Informal Sector in India*. Spatial Economics Research Centre Discussion Paper 0072.
- Mukim, M., Nunnenkamp, P. (2011) The location choices of foreign investors: a district-level analysis in India. *The World Economy*, 35: 886–918.
- Munshi, K. (2011) Strength in numbers: networks as a solution to occupational traps. *Review of Economic Studies*, 78: 1069–1101.
- Munshi, K., Rosenzweig, M. (2005) Economic development and the decline of rural and urban community-based networks. *The Economics of Transition*, 13: 427–443.
- Munshi, K., Rosenzweig, M. (2006) Traditional institutions meet the modern world: caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96: 1225–1252.
- Nataraj, S. (2011) The impact of trade liberalization on productivity: evidence from India's formal and informal manufacturing sectors. *Journal of International Economics*, 85: 292–301.
- Ottaviano, G. (2011) 'New' new economic geography: firm heterogeneity and agglomeration economies. *Journal of Economic Geography*, 11: 231–240.
- Overman, H., Puga, D. (2010) Labor pooling as a source of agglomeration: an empirical investigation. In E. Glaeser (ed.) *Agglomeration Economics*, pp. 133–150. Chicago, IL: University of Chicago Press.
- Parker, S. (2009) *The Economics of Entrepreneurship*. Cambridge, UK: Cambridge University Press.
- Rosenthal, S., Strange, W. (2004) Evidence on the nature and sources of agglomeration economies. In V. Henderson, J. F. Thisse (eds) *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2119–2171. Amsterdam: North-Holland.
- Rosenthal, S., Strange, W. (2012) Female entrepreneurship, agglomeration, and a new spatial mismatch. *Review of Economics and Statistics*, 94: 764–788.

- Rosenthal, S., Strange, W. (2001) The determinants of agglomeration. *Journal of Urban Economics*, 50: 191–229.
- Schoar, A. (2009) The divide between subsistence and transformational entrepreneurship. In J. Lerner, S. Stern (eds) *Innovation Policy and the Economy*, Vol. 10, pp. 57–71. Chicago, IL: University of Chicago Press.
- van Stel, A., Storey, D., Thurik, R. (2007) The effect of business regulations on nascent and young business entrepreneurship. *Small Business Economics*, 28: 171–186.
- Venables, A. (2011) Productivity in cities: self-selection and sorting. *Journal of Economic Geography*, 11: 241–251.
- World Bank (2011) *Poverty and Social Exclusion in India*. Oxford: Oxford University Press.
- World Bank (2012) *Gender Equality and Development*. World Development Report.
- Zucker, L., Darby, M., Brewer, M. (1998) Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review*, 88: 290–306.

Appendix

Table A1. Establishment size distributions for manufacturing and services

Size distribution	Mfg. 1994	Mfg. 2000	Mfg. 2005		Services 2001	Services 2006
Establishment count distribution						
Organized sector (20+ employees)	0.008	0.006	0.006	> 5 employees	0.045	0.043
Unorganized sector	0.992	0.994	0.994	1–5 Employees	0.955	0.957
Employment weighted distribution						
Organized sector (20+ employees)	0.211	0.172	0.193	> 5 employees	0.239	0.267
Unorganized sector	0.789	0.828	0.807	1–5 Employees	0.761	0.733

Notes: Authors' calculations based on NSS and ASI data (various rounds).