Journal of Economic Geography (2014) pp. 1–23 doi:10.1093/jeg/lbu020

Coagglomeration of formal and informal industry: evidence from India

Megha Mukim[†]

World Bank, 1818 H Street NW, Washington DC 20433, USA
[†]Corresponding author: Megha Mukim. *email* <mmukim@worldbank.org>

Abstract

A large and growing informal sector is a major feature of developing countries. I analyze coagglomeration patterns between formal and informal manufacturing enterprises in India, and study (i) the causes underlying these patterns, and (ii) the positive externalities, if any, on the entry of new firms. I find that buyer–seller and technology linkages explain much of formal–informal coagglomeration. I also find that this sectoral, within-industry, coagglomeration matters mostly to small- and medium-sized formal firms births. Traditional measures of agglomeration remain important in explaining new industrial activity, whether in the formal or the informal sectors.

Keywords: Coagglomeration, informal sector, India
JEL classifications: R12, R30, O17
Date submitted: 30 January 2013 **Date accepted:** 7 May 2014

1. Introduction

The unorganized, or informal sector[1] is an important means of livelihood to millions of people in developing countries. Because of its very nature—it is unregulated by government—data collection and subsequent analysis lags far behind that for the formal sector. In India, the informal sector[2] often falls outside the scope for planned development efforts, and thus remains in the shadows with regard to productivity, social security and statistics.

Within non-farm industries, informal enterprises account for 43.2% of Net Domestic Product (NDP) and employ 71.6% of the workforce. In fact, informal activity accounts for a majority of employment and makes a significant contribution to NDP in several developing countries in Asia, Africa, Eastern Europe and Latin America—on average the informal share of employment ranges from 24% in transition economies to 50% in Latin America and over 70% in sub-Saharan Africa (see Marjit and Kar, 2009 and Jutting and De Laiglesia, 2009). The importance of the informal sector in India shows little sign of abating, while that of the formal sector remains stagnant. Yet little is known about the relationship, whether symbiotic or competitive, between the two

[1] A number of countries, including India, often use the terms 'unorganised sector' and 'informal sector' interchangeably. This article will stick with the term 'informal'.
[2] I will use 'sector' to indicate the division between formal and informal, and 'industry' to differentiate between different types of manufacturing subdivisions, such as furniture, paper, textiles, etc.

sectors. In India, new firm activity is strongly related to existing firm activity—however, it is often assumed that formal and informal sectors function independently of one another. Since systematic data on the transactions of firms in the informal sector are scarce, any knowledge of its relationship with formal firms is often based on qualitative case studies—Pieters et al. (2010, 2011) and Bairagya (2010) are exceptions. Given that policy-makers are interested in encouraging new economic activity, there is a need to understand better the links between formal and informal firms and whether these have any beneficial spillovers.

In this article, I investigate two main issues. First, to what extent and why informal firms cluster with those in the formal sector, and second, whether the coagglomeration of formal and informal firms has any effect on the rate of new firms setting up shop within a given location. I focus on manufacturing firms in India. Data on informal firms are taken from surveys conducted by the National Sample Survey Organization (NSSO) in 2000–2001 and 2005–2006. Data on formal firms are taken from the Annual Survey of Industries (ASI) for the 2 years corresponding to the NSSO surveys. I assume that firms' entry could be driven by the effects of agglomeration and/or coagglomeration, market access, the size of the market and a set of unobserved attributes, which could be specific to the location or to the industry. I study formal and informal firms across 22 two-digit manufacturing industries, 459 districts and 2 years (2000–2001 and 2005–2006).

This article will start with a brief overview of the existing literature—studies that focus on the relationship between formal and informal sectors, and those that provide evidence on the factors that drive firm location decisions. I will then describe patterns of coagglomeration between formal and informal manufacturing firms in India and subsequently investigate the factors that could help explain the observed trends. In Section 4, I will estimate the effect of coagglomeration and other types of clustering on new firms' location decisions and demonstrate how formal firms differ from their informal counterparts. Section 5 will discuss the results and Section 6 will conclude.

I provide a flavour of the main results upfront. Coagglomeration between formal and informal manufacturing firms in India is not very high and it has been falling over time. Industries such as apparel, furniture and metal-making are exceptions, where it has risen over the period of analysis. When formal and informal firms do coagglomerate, this is driven by buyer and seller linkages and by transfer of equipment and design from formal to informal firms. Buyer linkages are especially strong between smaller formal firms and their informal counterparts. New informal firms tend to choose locations based strongly on own-industry agglomeration. Input–output linkages between industries are important for entry of both, formal and informal firms. Large formal firms tend to avoid industrially diverse districts. Informal firms are attracted to markets, formal firms to locations with better access to markets elsewhere.

2. Related literature

The literature, both theoretical and empirical, on coagglomeration of firms is new,[3] but fast-growing. Ellison and Glaeser (1997) coined the term 'coagglomeration' to describe

[3] Ciccone and Peri (2006), Bacolod et al. (2008, 2009), Gabe and Abel (2013) study complementarities across workers in urban labour markets.

the tendency of various industries to agglomerate in close proximity to one another. Helsley and Strange (2012) provide a succinct but thorough review of the theory describing the foundations for coagglomeration. Most of existing literature focuses on coagglomeration dynamics across industries (for instance, whether textile producers coagglomerate with leather producers) and not across sectors within industries (for instance, whether formal and informal textile producers coagglomerate). There are few empirical papers that study patterns of coagglomeration across sectors within a given industry. For instance, Barrios et al. (2006) study coagglomeration between domestic and foreign firms in Ireland and find evidence of FDI-related local spillovers. He et al. (2012) provide detailed descriptions of coagglomeration between exporters, non-exporters and foreign firms within industries and across locations in China. Howard et al. (2012) investigate coagglomeration between low- and high-tech clusters in Vietnam and find that technology transfers and skills correlations explain the observed patterns.

What explains coagglomeration, and what sorts of benefits might this lead to? Firms could coagglomerate because of greater access to specialized inputs, labour–market pooling, and knowledge and technology spillovers (Ellison et al., 2010). Firms could also be attracted to the same location owing to the presence of natural endowments or public goods. The first part of this article focuses on coagglomeration of formal and informal firms within a given industry. There are a number of studies that have investigated the degree of complementarity or substitutability between the formal and informal sectors in India. Mitra (2009) finds that the incidence of informality is high in more industrialized Indian states, suggesting that informal activity could be complementary to the formal sector. Some working papers focus on studying the production linkages between the formal and informal sectors in India, and its effect on informal employment and on the size of the informal sector—see Pieters et al. (2010), Pieters et al. (2011) and Bairagya (2010). Pieters et al. (2010, 2011) find evidence of complementarities between the modern informal sector and firms in the formal sector, which are linked through outsourcing. Their analysis is conducted across two-digit manufacturing industries at the level of the state. States in India are often the size of small countries and in this article I use districts as the geographical unit of analysis.[4] Marjit (2003) uses a general equilibrium model and argues that only the capital-intensive segment of the informal sector is complementary to the formal sector. On the other hand, Dutta et al. (2011) find that informal activity serves as a substitute to the formal sector when they observe increased informal activity in states in India that suffer from higher levels of corruption (that inhibits formal activity).

The informal sector in India largely ignores labour regulations, officially recognized collective bargaining processes, taxes or institutional obligations. There is some research (Marjit and Kar, 2009) to show that informal manufacturing and self-employed units accumulate fixed assets and invest and that often they are able to do so in times when their formal counterparts are mired in complex regulations. Informal production can be an important input to the production of intermediate goods, processed exports and import substitutes, supported by supply side contracts with the formal sector, leading to beneficial spillovers for both sectors. For instance, informal carpet weavers in Agra operate alongside larger, more formal carpet designers and

[4] In urban regions, districts often correspond to cities and surrounding areas.

exporting firms in the city. Ranis and Steward (1999) illustrate how better formal–informal linkages could lead to growth within the informal sector. House (1984) and Arimah (2001) find that larger informal firms (in Kenya and Nigeria, respectively) often benefit from subcontracting or direct sales linkages with formal firms. Bairagya (2010) finds that in the presence of trade liberalization, production linkages between formal and informal manufacturing firms leads to an increase in the size of the informal sector. Overman and Venables (2005) and Duranton (2008) describe a number of case studies illustrating the generation of agglomeration economies from the interaction between formal and informal enterprises.

Firms, whether formal or informal, could also benefit from agglomeration—i.e. with firms from the same industry, from related industries or within a location with a diverse industry mix. Marshall (1919) theorized that clusters of firms, predominantly in the same industry, could take advantage of localization economies, such as the sharing of sector-specific inputs, skilled labour and knowledge. Firms in a given industry and those in related industries might agglomerate to enjoy the benefits such as inter-industry linkages, buyer–supplier networks and opportunities for efficient subcontracting (Venables, 1996). An overall large size of the urban agglomeration and its more diverse industry mix is also thought to provide external benefits beyond those realized within a single industry or due to a tight buyer–supplier network (Henderson, 2003).

The second part of this article will describe the effect of formal–informal coagglomeration versus that of other types of agglomeration on the generation of new activity, i.e. firm births, in the formal and the informal sectors. There are a number of papers that study firm births in the formal and the informal sectors in India—these are referred to as firms' location decisions (Lall and Mengistae, 2005; Lall and Chakravorty, 2005), or entrepreneurship (Ghani et al., 2011a, 2011b). Mukim and Nunnenkamp (2012) consider the location decisions of foreign firms in India and find that foreign investors tend to increasingly favour locations that already host other foreign investors. Lall and Mengistae (2005) find that formal manufacturing firm location choices in India are strongly influenced by the local business environment and own-industry agglomeration economies. Lall and Chakravorty (2005) find that new manufacturing activity is attracted to locations with existing industrial concentrations, exacerbating spatial inequalities. Lall et al. (2003, 2004) study the effect of agglomeration economies on the performance of existing formal manufacturing firms in India.

A priori, there is no reason to assume that informal sector activity remains unaffected by agglomeration economies. In the absence of access to formal credit facilities, or alternatively since they are untouched by changes in regulations, the importance of buyer–supplier linkages and informal networks of social interaction could be more important to informal firms than to firms operating in the formal sector. Mukim (2011) studies the location decisions of new informal firms, manufacturing and services, in India and finds that input–output linkages are important in determining spatial entry patterns. Ghani et al. (2011a) study the spatial determinants of new formal and informal firm births, i.e. entrepreneurial activity. They also find that input–output agglomeration economies matter for manufacturing firm entry, for both formal and informal firms. Ghani et al. (2011b) compare the spatial determinants of entry separately for male and female entrepreneurs in informal manufacturing and services in India. They find that the effect of female-owned firm agglomeration economies (particularly input–output linkages) is a strong predictor of new female entrepreneurial activity.

This article will build upon the results from earlier analysis in two ways. First, it will focus on the coagglomeration of informal and formal activity *within* manufacturing industries in India. While there are papers that have studied the production linkages between formal and informal firms in India, they have ignored other kinds of linkages (labour, technology-related) and have focussed their analysis at the level of states. Second, the article will contribute to the growing literature on the spatial determinants of entry patterns of formal and informal firms, and account explicitly for the effect of formal–informal coagglomeration versus that of traditional measures of agglomeration.

3. Formal–informal coagglomeration

3.1. Descriptive statistics

Informal firms are those that are not registered under the Factories Act of 1948 (which requires all firms engaged in manufacturing to register if they employ 10 workers or more and use power, or if they employ 20 workers or more). All public sector enterprises are assumed to be in the formal sector. Small formal firms (for instance, those employing 10–20 workers and using electricity) might be very similar to their larger informal counterparts. However, the data seem to suggest that formal firms are a very different animal.

In 2005–2006, the average number of employees per formal firm was 141 while the average informal firm had three employees. Informal firms are predominantly small-scale enterprises—between 90% and 93% of the establishments in the sample employ less than five workers, and between 68% and 70% employ no hired labour at all. Within the formal sector, small firms (i.e. less than five workers) comprise 5% of the sample, medium (i.e. employ more than 5, but less than 100 workers) comprise 71% and large (i.e. employ more than 100 workers) comprise 24%. Informal firms tend to belong mostly to industries such as apparel, food, textiles, wood and furniture. This includes cottage and household industries, khadi and village industries, handlooms, handicrafts, coir, sericulture, etc., set up all over the country in rural, semi-urban and urban environments. Formal firms are fairly well distributed across different manufacturing industries, such as non-metallic products, chemicals, basic metals and machinery and equipment.

Data on informal firms and employment are drawn from the Fifty-Sixth Round (July 2000–June 2001) and the Sixty-Second Round (July 2005–June 2006) of the National Sample Survey Organization. The surveys cover rural and urban areas, and the sample size varies from 100 units in remote areas to over 10,000 units in major cities. Data on formal firms are drawn from the Annual Survey of Industries, which conducts repeated cross-sectional annual surveys that cover all firms registered under the Factories Act of 1948. ASI survey data are available for each year starting in 1980. However, for the purposes of this article, data corresponding to the 2000–2001 and 2005–2006 surveys are used. This article focuses only on manufacturing industries that cover two-digit National Industrial Classification (NIC) 15 to 37.

The surveys cover around 558 districts in 2000–2001 and 576 districts in 2005–2006—an increase of 18 districts. These new districts are formed out of existing districts—mean district population fell from 1.88 million in 2000–2001 to 1.71 million in 2005–2006. For the purpose of analysis between the two time periods, I combine the new districts, essentially keeping the 2000–2001 district boundaries intact. For instance, Delhi is one

district in 2000–2001, but is divided into nine districts in 2005–2006—I use the combined data from the nine new districts so as to match this to the data from the original district boundaries.[5] In short, the analysis includes the original 558 districts, of which data are available for a total of 459 districts. After the 2005–2006 district boundaries are combined in line with the 2000–2001 boundaries, the mean district population for 2005–2006 rises to 2.11 million.

Do formal and informal firms cluster in the same locations? I calculate the Theil Index[6] for formal and informal firms and list the top 20 districts (Table 1). Only two districts (Mumbai and Ludhiana) are common to the top five districts across both sectors, and an additional two (Delhi and Ahmadabad) common to the top 10, indicating that there could be limited concordance across formal and informal clustering. To measure coagglomeration between the formal and informal sector directly, I follow Ellison and Glaeser (1997) and Ellison et al. (2010) in principle but modify their Coagglomeration Index somewhat. I compute a measure of formal–informal coagglomeration within each two-digit manufacturing industry, across each of the 2 years for which I have data at my disposal. In other words, the index for the coagglomeration of sectors i and j is:

$$\gamma_{ij}^{c} = \frac{\sum_{m=1}^{M}(S_{mi} - X_m)(S_{mj} - X_m)}{1 - \sum_{m=1}^{M} X_m^2},$$

Where m indexes districts, i indexes the formal sector and j the informal sector. S_{mi} is the share of formal sector employment contained in district m, S_{mj} is the share of informal sector employment contained in district m and X_m measures the aggregate size of district m, which is measured here as the mean employment share in the region across the two sectors (formal and informal).[7] The coagglomeration index is based on measures of formal and informal sector employment. Informal employment within industries across districts averaged 962,000, while formal employment averaged 2386. In addition, average district-level informal employment grew by 10.8% between 2000–2001 and 2005–2006, while formal employment grew by 9.3%.

Table 2 lists the coagglomeration measures for formal–informal sectors by industry for 2000–2001 and 2005–2006. Negative values of the index arise when formal and informal sectors are agglomerated in different areas. For example, when the coagglomeration index equals −0.0259 for 'Manufacture of Coke' in 2000–2001, it implies that informal coke manufactures tend to agglomerate in locations different from formal coke manufacturers. Industries like wood, food, apparel and fur, non-metallic mineral products tend to rank higher on the coagglomeration index from 1 year to the

[5] More difficult examples of combination include the new Dantewara district (in the state of Chhattisgarh) which was carved out of Bastar district in the same state—information that was gleaned from searches on Google.

[6] The Theil Index here provides an indication of the over or under-representation of district across a set of given industries, i.e. the distribution of new firms by two-digit industry across districts. The index belongs to the family of generalized entropy inequality measures wherein the values vary between 0 and ∞, with zero representing an equal distribution and higher values representing higher values of inequality. The value of the index increases in the inequality of the distribution of firm births by district with respect to total firm births: $T = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{x_m}{\bar{x}}.\ln\frac{x_m}{\bar{x}}\right)$, where x_m is the number of firm births in district m.

[7] In other words, the value of X_m equals the mean of district m's share of formal sector employment and its share of informal sector employment, i.e. mean$\left(\frac{\text{Employment}_{im}}{\text{Employment}_i}, \frac{\text{Employment}_{jm}}{\text{Employment}_j}\right)$.

Table 1. Contributions to the Theil Index

District	Informal	District	Formal
Mumbai	255.43	Bangalore Urban	549.39
Ludhiana	146.34	Mumbai city	433.85
South Tripura	100.84	Coimbatore	334.99
Kolkata	80.03	Vellore	246.75
Delhi	52.53	Ludhiana	180.36
Ahmadabad	47.11	Pune	163.48
Jaipur	44.08	Thane	143.64
South 24 Parganas	43.08	Pudukkottai	129.05
Coimbatore	42.63	Delhi	117.97
West Tripura	42.19	Surat	110.00
Surat	39.93	Ahmadabad	96.51
Thane	39.70	Chennai	93.01
North 24 Parganas	39.52	Thiruvallur	84.87
Haora	37.08	Guntur	83.19
Murshidabad	36.44	Kollam	81.97
Srinagar	34.17	Nizamabad	78.29
Hyderabad	34.00	Gurgaon	76.82
Varanasi	32.53	Gautam Buddha Nagar	75.07
Virudhunagar	31.18	Daman	68.29
Vellore	29.69	Rangareddi	68.27

Notes: Based on informal and formal sector employment for 2005–2006. The Theil Index provides an indication of the distribution of new firms by two-digit industry across district. The higher the value of the index, the higher the inequality of the distribution of firm births by district with respect to total firm births—indicating higher levels of clustering.

next. Averaged across industries, coagglomeration dropped from 0.0102 in 2000–2001 to 0.0042 in 2005–2006. Similar to the Ellison–Glaeser agglomeration index, the no-coagglomeration benchmark is when the value of the index is zero. In general, if the coagglomeration index is greater than 0.05, the sectors are considered to be highly concentrated. In 2000–2001, food products, tobacco, textiles and non-metallic minerals seemed reasonably coagglomerated with values averaging above the 0.03 mark, although these dropped to 0.01 in 2005–2006. And indeed, except for three industries—apparel, fabricated metal products and furniture, the measure of coagglomeration fell in 2005–2006, suggesting that formal and informal firms were less likely to coagglomerate.

3.2. Why do firms coagglomerate?

Theories of agglomeration predict that firms cluster to be in close proximity to goods, people and/or ideas. Or, they could agglomerate owing to the presence of natural advantages in a location. These explanations could apply equally to formal and informal firms within a given industry that tend to cluster next to one another—they might be linked through exchange or common use of goods, labour and/or technology. I describe below the different variables I construct to measure the possible explanations for formal–informal coagglomeration.

Table 2. Formal–informal coagglomeration

NIC	Industry name	2000–2001	2005–2006
15	Food products and beverages	0.0318	0.0173
16	Tobacco products	0.0300	−0.0001
17	Textiles	0.0314	0.0152
18	Apparel and fur	0.0118	0.0172
19	Leather, luggage and footwear	0.0181	0.0095
20	Wood, except furniture	0.0311	0.0174
21	Paper and paper products	0.0144	0.0066
22	Publishing, printing, recorded media	−0.0002	0.0142
23	Coke	−0.0259	−0.0064
24	Chemicals and chemical products	0.0214	0.0097
25	Rubber and plastic products	−0.0052	0.0068
26	Other non-metallic mineral products	0.0361	0.0166
27	Basic metals	0.0086	0.0064
28	Fabricated metal products	0.0138	0.0169
29	Machinery and equipment	0.0147	0.0119
30	Office	0.0252	−0.0540
31	Electrical machinery and apparatus	−0.0040	0.0113
32	Radio	−0.0219	−0.0100
33	Medical, watches, etc	0.0102	−0.0131
34	Vehicles	0.0037	−0.0012
35	Other transport	0.0125	0.0025
36	Furniture	0.0139	0.0158

Notes: Ellison–Glaeser (1997) coagglomeration index for formal and informal sectors within every two-digit manufacturing industry.

Buyer–seller linkages (proximity to goods): firms from a given sector could locate close to a firm from another sector if they bought goods from or sold goods to one another.[8] For example, Arimah (2001) provides evidence of linkages between the formal and informal sector in Nigeria in the form of subcontracting and the flow of consumer goods and raw materials. Carr et al. (2000) illustrate that formal enterprises often provide materials and inputs to informal enterprises, which then transform them to send back to formal enterprises. In such cases, proximity would lower costs for both formal and informal firms. To assess the importance of downstream linkages (i.e. seller linkages), I use data from the NSSO on informal firms that are subcontracting, i.e. working on a contract with firms in the private sector, co-operatives or with a contractor. In other words, $\text{Seller}_{j \to i, mkt}$ refers to the proportion of informal firms that are supplying or selling to the formal sector within a given district (m), industry (k) and year (t). To assess the importance of upstream (i.e. buyer) linkages, I use data from the ASI on formal firms that subcontract their inputs (excluding labour) from outside the firm. Thus, $\text{Buyer}_{i \to j, mkt}$ is the proportion of total input expenses that are being

[8] It could also be argued that buyer–seller linkages and coagglomeration are endogenous—in other words, firms may use the outputs of (or sell to) particular sectors simply because these sectors tend to coagglomerate (for other reasons).

subcontracted[9] within a given industry. In the regressions, I weight this variable at the district-level by the number of formal firms.

Labour–market linkages (proximity to people): agglomeration could occur because workers are able to move across firms and industries, or in this case, between firms in the informal and the formal sector. Rosenthal and Strange (2001) provide an overview of the importance of labour market pooling in explaining the spatial concentration of firms. To assess labour movement across the informal and formal sectors, I use data from the ASI on labour employed through a contract.[10] In other words, $\text{Labour}_{i \to j, mkt}$ is the proportion of employees that are being subcontracted from outside of the firm. In the regressions, I weight this variable at the district-level by the number of formal firms.

Technology spillovers (proximity to ideas): firms could also co-locate if this helped the transfer of technology within a cluster. Porter (1990) points out that knowledge sharing often takes place through the processes of buying and supplying. To assess the importance of technological linkages, I use data from the NSSO on the subset of informal firms that are subcontracting and who receive either equipment or design from the contractor. I construct a variable $\text{Tech}_{j \to i, mkt}$ that refers to the proportion of informal firms within a given industry that are receiving a technology transfer from the contractor.

Table 3 summarizes the mean values of each of the explanatory variables across districts and how these vary across different manufacturing industries from 2000–2001 to 2005–2006. In 2005–2006, on average, formal firms subcontracted 29% of their total employment—industries like tobacco, coke and non-metallic mineral products tended to subcontract more, while publishing, office equipment and textiles subcontracted less. The average proportion of inputs subcontracted was also high (36% in 2005–2006), with textiles, apparel and basic metals in the lead, and coke, chemicals and non-metallic minerals lagging behind. Interestingly, a simple pair-wise correlation reveals that labour–market linkages seem to be negatively correlated with buyer linkages, suggesting that firms that subcontract labour are less likely to subcontract inputs. The proportion of informal firms subcontracting, i.e. seller linkages as defined above, are relatively low at an average of 8% in 2005–2006. More sophisticated manufacturing products, such as vehicles, radio, publishing and rubber and plastics seem to subcontract more, while informal firms in industries like food and tobacco subcontract less. The proportion of informal firms supplied with equipment and design through subcontracting is the lowest, averaging around 2% for 2005–2006. Informal firms in industries such as other transport, medical/watches, publishing and machinery and equipment seem to have better technological linkages through their contractors, while radio, coke, tobacco and office equipment have the least linkages.

[9] This is a proxy for subcontracting by formal firms to informal firms—but it is equally possible that formal firms could subcontract inputs to other formal firms.

[10] Most papers explaining agglomeration tend to study the effect of the labour market in terms of several characteristics—size, education and/or skill-level, wages, productivity, mobility—separate from the effect of input–output linkages and technological externalities (see Overman and Puga, 2010, for an overview). In this article, I focus mainly on the extent to which formal and informal sectors coagglomerate because of hiring from the same industry-specific pool. I assume that formal-sector labour subcontracting captures the channel through which characteristics of the labour pool are attractive to both formal and informal sectors.

Table 3. Descriptive statistics

NIC	Industry name	Labour market linkages			Buyer linkages			Seller linkages			Technology linkages		
		2000	2005	Δ[a]	2000	2005	Δ	2000	2005	Δ	2000	2005	Δ
15	Food products, beverages	0.19	0.28	*38%*	0.30	0.37	*22%*	0.02	0.02	*5%*	0.01	0.01	*−39%*
16	Tobacco products	0.38	0.52	*31%*	0.41	0.38	*−9%*	0.03	0.02	*−51%*	0.01	0.00	*−82%*
17	Textiles	0.12	0.17	*37%*	0.20	0.52	*87%*	0.09	0.09	*−3%*	0.01	0.02	*28%*
18	Apparel, fur	0.19	0.29	*40%*	0.44	0.46	*4%*	0.08	0.06	*−38%*	0.01	0.02	*23%*
19	Leather, luggage, footwear	0.21	0.26	*18%*	0.31	0.35	*11%*	0.06	0.06	*2%*	0.01	0.01	*29%*
20	Wood	0.21	0.28	*25%*	0.21	0.28	*28%*	0.08	0.11	*26%*	0.01	0.02	*57%*
21	Paper, paper products	0.21	0.23	*12%*	0.09	0.41	*125%*	0.05	0.07	*26%*	0.01	0.01	*58%*
22	Publishing, printing, etc	0.08	0.15	*56%*	0.24	0.34	*37%*	0.21	0.19	*−8%*	0.03	0.03	*−2%*
23	Coke	0.25	0.41	*47%*	0.11	0.19	*53%*	0.03	0.01	*−82%*	0.03	0.00	*−200%*
24	Chemicals, chemical products	0.19	0.28	*37%*	0.18	0.27	*39%*	0.03	0.01	*−81%*	0.02	0.01	*−81%*
25	Rubber, plastic products	0.14	0.25	*55%*	0.21	0.40	*62%*	0.06	0.12	*59%*	0.01	0.02	*25%*
26	Non-metallic mineral products	0.33	0.44	*29%*	0.09	0.27	*96%*	0.02	0.05	*85%*	0.01	0.01	*96%*
27	Basic metals	0.25	0.30	*20%*	0.26	0.43	*50%*	0.10	0.10	*0%*	0.04	0.02	*−57%*
28	Fabricated metal products	0.20	0.34	*52%*	0.28	0.42	*40%*	0.11	0.10	*−16%*	0.02	0.02	*−15%*
29	Machinery and equipment	0.11	0.19	*57%*	0.30	0.41	*31%*	0.10	0.08	*−20%*	0.02	0.03	*23%*
30	Office	0.13	0.16	*21%*	0.32	0.31	*−3%*	0.17	0.07	*−86%*	0.00	0.00	
31	Electrical machinery etc	0.15	0.25	*52%*	0.21	0.36	*52%*	0.05	0.06	*23%*	0.01	0.02	*45%*
32	Radio	0.11	0.18	*52%*	0.23	0.35	*43%*	0.07	0.12	*56%*	0.00	0.00	
33	Medical, watches etc	0.07	0.16	*72%*	0.32	0.32	*2%*	0.06	0.09	*44%*	0.01	0.04	*151%*
34	Vehicles	0.15	0.26	*50%*	0.31	0.40	*25%*	0.15	0.11	*−37%*	0.01	0.01	*15%*
35	Other transport	0.15	0.27	*57%*	0.30	0.41	*31%*	0.14	0.11	*−18%*	0.05	0.05	*−14%*
36	Furniture	0.26	0.29	*11%*	0.28	0.35	*23%*	0.13	0.13	*7%*	0.02	0.02	*46%*
	Average	0.20	0.29	*37%*	0.23	0.36	*44%*	0.08	0.08	*1%*	0.02	0.02	*18%*

[a]Change refers to the percentage differential compared to the average value over 2000 and 2005.

To understand the extent to which each of these linkages might be driving the coagglomeration of formal and informal firms, I run the following regression:

$$\gamma^{c}_{ij,kt} = \text{Buyer}_{i\to j,mkt} + \text{Supplier}_{j\to i,mkt} + \text{Labour}_{i\to j,mkt} + \text{Tech}_{j\to i,mkt} + \text{District}_{m} + \text{Year}_{t} + \epsilon_{mkt},$$

Where, the dependent variable is coagglomeration between formal (i) and informal (j) firms in a given industry k in year t. The fact that some locations might be attractive to both informal and formal firms owing to certain types of natural advantages is captured by a district-specific dummy, and I also include a year-specific dummy. Since the dependent variable varies only by industry and year, I do not include industry dummies.

Table 4 summarizes the results. Coagglomeration between the formal and informal sectors seems to be driven largely by buyer and seller linkages and technology transfers. A unit increase in technological linkages, i.e. the proportion of informal firms that receive equipment or designs from the contractor, leads to a 0.0045 unit increase in coagglomeration—see model (3). Given that the average value of the coagglomeration index in 2005–2006 was 0.0042, this is a sizeable effect. Surprisingly, the effect of labour–market linkages is mostly insignificant. If stringent labour regulations in the formal manufacturing sector led firms to subcontract labour from the informal market, then that should have given formal firms more reason to coagglomerate with their informal counterparts. The results might indicate that formal firms are more likely to subcontract informal labourers that do not operate or work in informal enterprises. However, as I am unable to include industry-specific measures of more general labour–market characteristics I am unable to test this hypothesis. In addition, since I do not include industry-specific dummies, it might also be more or less feasible to subcontract to the informal sector in some industries than others, perhaps because they are regulated differently.

I also calculate separately the coagglomeration index and the explanatory variables for the subset of small and medium enterprises (SMEs) in the formal sector, i.e. firms with less than 100 employees, and then re-run the regression—see model (4). I find that the effect of buyer-linkages, i.e. the proportion of inputs being subcontracted by formal SMEs, is the only variable that matters—the effect of other variables is now irrelevant. This indicates that SMEs in the formal sector coagglomerate with informal firms primarily for the purpose of production linkages.

As a robustness exercise, I also drop the district-level dimension, re-compute the variables and re-run the regressions with variation limited to industry and year only.[11] I find that buyer linkages are no longer significant, but that seller linkages remain steadfastly significant and that their effect seems stronger. In addition, the technological linkages also exert a greater influence on coagglomeration. In separate regressions, I also include industry-level controls—such as average industry wage and size (i.e. the number of formal firms). The result seems to indicate that average industry-level wages (in the formal sector) positively affect formal–informal coagglomeration. In addition, the R^2 of the model with all firms is substantially improved.

[11] These results are available as a supplementary appendix.

Table 4. Explaining formal–informal sector coagglomeration

Variable	All firms			SMEs only
	(1)	(2)	(3)	(4)
Labour–market linkages	0.0028*	−0.0014	0.0005	−0.0020
	[0.002]	[0.001]	[0.002]	[0.002]
Buyer linkages	0.0104***	0.0068***	0.0069***	0.0015**
	[0.002]	[0.002]	[0.002]	[0.000]
Seller linkages	0.0041***	0.0040***	0.0042***	0.0011
	[0.001]	[0.001]	[0.001]	[0.001]
Technological linkages	0.0047***	0.0039**	0.0045**	0.0011
	[0.002]	[0.002]	[0.002]	[0.002]
Constant	0.0016***	0.0038***	0.0032	0.0011
	[0.000]	[0.000]	[0.002]	[0.002]
Year fixed effects	✗	✓	✓	✓
District fixed effects	✗	✗	✓	✓
Observations	2290	2290	2290	1735
R^2	0.036	0.181	0.287	0.120

Notes: Dependent variable is the coagglomeration index for formal–informal clustering for each manufacturing industry. Buyer linkages and labour–market linkages have been weighted at the district-level by the number of formal firms. Standard errors in square brackets clustered at the industry level. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

4. New firms, agglomeration and coagglomeration

Why should we care about coagglomeration between the formal and informal sectors? In particular, I would like to know if coagglomeration, relative to other sorts of agglomeration, affects firm births (or entrepreneurial activity) in formal and informal sectors.

The ASI and the NSSO provide information on firm births for the 2 years of interest—2000–2001 and 2005–2006. The NSSO asks informal firms if they have been in operation for less than 3 years—positive replies are coded as firm births. The ASI provides information on the year of initial production of the formal firm, thus identifying recent[12] entrants to the market. I create a repeated cross-sectional data set, pooled across the 2 years—2000–2001 and 2005–2006 for the dependent variable, i.e. the count of new firms per capita.

Table 5 summarizes the data on firm births. Clearly, not every two-digit-industry and district pair experience positive values of firm births and new firm entrants in the formal sector are rarer and fewer than those in the informal sector. Recall that the data are taken from repeated surveys of formal (ASI) and informal (NSSO) firms across industries and districts in India—I assume that the sample accurately captures firm-birth trends in the universe of firms. Thus, my assumption is that these zero values are

[12] In line with the NSSO survey, I take new entrants to the market for the given year and the two preceding years.

Table 5. Characteristics of the data

Data source	Year	Variable[a]	No.	Mean	Std. dev.	Min.	Max.
ASI	*2000–2001*	Count	5915	0.72	2.8	0	136
(formal sector)		Count > 0	1511	2.8	4.9	1	136
	2005–2006	Count	6192	0.4	2.08	0	89
		Count > 0	961	2.6	4.7	1	89
NSSO	*2000–2001*	Count	5915	228	1109	0	46,476
(informal sector)		Count > 0	2656	508	1612	1	46,476
	2005–2006	Count	6192	324	1730	0	57,801
		Count > 0	2748	731	2540	1	57,801

[a]This refers to the number of district–industry pairs. 'Count > 0' refers to the number of pairs with observed firm births. 'Count' refers to the total possible number of pairs.
Notes: Count of firm births for the formal sector (based on ASI data) and for the informal sector (based on NSSO data).

genuine, in that they indicate no new firm entrants in the given year, industry and district.

I would like to parcel out the effect of the agglomeration and economic geography variables, i.e. within-industry clustering, input–output or inter-industry clustering, industrial diversity and market access, from the effect of formal–informal coagglomeration on firm births. The estimation framework is based on a location decision model[13] in which entrepreneurs compare profitability across different locations. In the presence of agglomeration economies, births will tend to concentrate in these locations; otherwise births will tend to disperse.[14] I test the importance of different factors by estimating a model wherein the count of new firms (per capita) within a location is a function of factors common to the location (i.e. industrial diversity, market access and wealth) and those common to particular industries (i.e. intra and inter-industry clustering and formal–informal coagglomeration).

4.1. Specification of variables

The economic geography variables in this model are represented by market access (MA_{mt-1}), intra-industry or localization economies (σ_{mkt-1}), inter-industry linkages (τ_{mkt-1}), the Hirschman–Herfindahl Index for industrial diversity (U_{mt-1}) and coagglomeration between the formal and informal sectors (γ_{kt}). M_m, K_k and T_t are the district, industry and year dummies and are used wherever possible. To reduce clutter, the time subscript has been excluded in the descriptions of the variables below.

Localization economies are measured as the proportion of industry $k's$ employment in district m as a share of all of industry $k's$ total employment in the country. The higher

[13] See Section II.A in Rosenthal and Strange (2003) for an excellent exposition of the underlying model.
[14] This model will not test the effect of agglomeration on the location of existing firms, and other characteristics relating to their performance, such as productivity, employment or exit.

this value, the higher the expectation of intra-industry concentration benefits in the district.

$$\sigma_{mk} = \frac{E_{m,k}}{E_k}.$$

To evaluate the strength of input–output linkages for each industry, a summation of district industry employment weighted by the industry's input and output coefficients from the column and row vectors from the national input–output account are used:

$$\tau_{mk} = \sum_{l=1}^{n} (w_{ki} + w_{ko}) e_{mk},$$

where, τ_{mk} is the strength of the input–output linkages, w_{ki} is industry k's national input coefficient column vector, w_{ko} is industry k's national output coefficient row vector and e_{mk} is total employment for industry k in district m. The strength of industry $k's$input–output linkages within the district is based on the sum of the industry's links with all other industries (given by l) and excluding own-industry linkages (i.e. $l \neq k$). The measure examines district-level inter-industry linkages based on national input–output accounts. I also compute input and output linkages separately to use in my regressions, but I find that these variables are highly correlated (correlation coefficient = 0.83) and that this leads to serious multicollinearity problems. Thus, I use an input–output variable that captures the extent to which different industries buy from and sell to one another. These linkages, thus defined, are different from buyer and seller linkages previously defined—the latter measure the extent to which formal and informal *firms within a given industry* bought and sold goods from one another.

The Hirschman–Herfindahl Index is used to measure the degree of economic diversity. Unlike measures of specialization, which focus on one industry, the diversity index considers the industry mix of the entire regional economy. The largest value for U_m is one when the entire regional economy is dominated by a single industry. Thus a higher value signifies lower level of economic diversity.

$$U_m = \sum_k \left(\frac{E_{mk}}{E_m}\right)^2.$$

Market access, or the potential for opportunities for interactions with neighbouring districts is defined as:

$$MA_m = \sum_n \frac{S_n}{d_{m-n}^b},$$

where, MA_m is the accessibility indicator estimated for location m, S_n is a size indicator at destination n (district population), d_{mn} is a measure of distance between origin m and destination n, and b describes how increasing distance (orthodomic) reduces the expected level of interaction. In the original model proposed by Hansen (1959), b is an exponent describing the effect of the travel time between the zones. In this specification, the exponent value is set to 1. Transport is allowed to occur along a 500-km radius.[15]

[15] It would be unrealistic to measure market access by connecting a district to every other possible district in the country. Thus, I limit the measure by what can possibly be reached by motor vehicle within a day's journey. The 500-km radius also covers roughly the area of the state and neighbouring states.

The size of the district m is not included in the computation of market access—only that of neighbouring districts is taken into account.

I also include the percentage of high-income households (WE_m) within a district as an indicator of the measure of consumption expenditures within a district. The variable is constructed using household consumption data and refers to those households that belong to the highest monthly per-capita consumption expenditure (MPCE) group. The actual MPCE category differs depending on the year of the survey, the type of district (rural or urban) and the population of the district.

Coagglomeration (γ_{kt}) is defined in Section 3, and measures coagglomeration between the formal and informal sectors within each two-digit manufacturing industry (k).

It is also standard to control for location-level infrastructure such as transport, electricity, banks, etc. However, as physical infrastructure changes relatively slowly over time (in this case over a 5-year period), the effects should be more or less soaked up by district dummies.

My main estimating equation is as follows:

$$y_{mkt} = \alpha + x_{mkt-1} + \sigma_{mkt-1} + U_{mt-1} + \gamma_{kt} + MA_{mt-1} + We_{mt-1} + M_m + K_k + T_t + \epsilon_{mkt}.$$

I run separate regressions for firm births in the formal and the informal sectors. The dependent variable in the model is the count of new firms per capita in two-digit industry k in district m and at time t, while all the explanatory variables, with the exception of coagglomeration, in the model are defined at time $t-1$. Note that the explanatory variables are common across the two sets of regressions since these have been constructed taking total employment, i.e. across the formal and informal sector, into account. The simple descriptive statistics of the explanatory variables are presented in Table 6.

Data on these explanatory variables are drawn from the Employment and Unemployment Surveys—Round 55.10 (July 1999–June 2000) and Round 61.10 (July 2004–June 2005). These data, which are disaggregated by industry and district, allows me to construct the agglomeration variables. Data on wealth and population within the district are also drawn from the household surveys.

4.2. Endogeneity concerns

Although the dependent variable is the count per capita of firm births and the existing industrial structure is taken as given, the effect of unobservable factors could bias the coefficients. The underlying assumption within the model is that if a particular location offers some inherent features that improve the profitability of certain economic industries, this will increase firm births. Such inherent features may be related to natural endowments or regulatory specificities, but they could also have to do with essentially un-measurable factors such as local business cultures. It is important to distinguish whether firms are attracted by a common unobservable, whether they derive benefits from being located in close proximity to one another, or whether it is some combination of the two.

I include an arsenal of fixed effects, specifically, industry, district and year fixed effects, to control for omitted variables bias. District fixed effects successfully control for any time-invariant characteristics of the district, such as the presence of natural resources, climate, proximity to the coast—in short, all features of the natural

Table 6. Descriptive statistics

Variable	Expected sign	Varies by	Mean	
			1999–2000	*2004–2005*
Localization	+	District × industry × year	0.0040	0.0045
Input–output	+	District × industry × year	8773	8531
Herfindahl index	−	District × year	0.359	0.382
Coagglomeration[a]	+	Industry × year	0.012	0.005
Market access	+	District × year	870,570	904,737
Wealth[b]	+	District × year	5.98	5.58
Population	+	District × year	2,155,122	1,983,314

[a]Refers to 2000–2001, and 2005–2006.
[b]Refers to percentage of the population.
Notes: After cleaning, data are available for a total of 459 districts and 22 two-digit manufacturing industries.

geography of the district. Industry fixed effects control for any time-invariant characteristics particular to the industry. Year fixed effects control for any unobservable time-specific shocks. However, there could be industry-district conditions that could also affect the coefficients. I lack enough variation in the data to be able to include district × industry dummies, and so am unable to control satisfactorily for certain types of omitted variables bias. For instance, district-regulations that favour particular industries could bias the results—examples of these include the establishment of industrial zones for specific manufacturing industries (such as textiles, electronics, etc[16]), or the migration of workers with particular skills to a region (the influx of diamond cutters and traders into districts like Surat in Gujarat is an example of such industry-district specific changes in the past). Changes in the past should be picked up by measures of existing industrial agglomeration, but only industry-district dummies would successfully soak up these effects in the years of interest.

5. Results and discussion

The large number of zero firm births in the data bring with them computational problems. Rosenthal and Strange (2003) use tobit estimations to deal with the problem of zero births. The bias resulting from noisy estimates of fixed effects in nonlinear models is not a problem when the number of observations per fixed effect is large—in this case, the data lack enough variation when district-level dummies are introduced since there are 459 districts in the data. Thus, I run tobit regressions, wherein the response variable is the count of firm births per capita, with year and industry dummies only. I also use a probit model, wherein the response variable is a discrete variable representing the zero-positive outcomes for firm births defined at the level of districts, industries and the 2 years. I use the full range of district, industry and year fixed effects

[16] See updated list of zones here: http://en.wikipedia.org/wiki/List_of_Special_Economic_Zones_in_India.

and report marginal effects.[17] And, finally, I estimate a linear OLS model (wherein the response variable is the log of non-zero counts of new firms per capita) that drops observations with zero births, but the results of which remain unbiased with the use of district, industry and year dummies. Input–output linkages, the Herfindahl index, market access are in logs, while localization, coagglomeration and wealth are not. I run separate regressions for informal and formal firm births (Tables 7 and 8).

The tobit coefficients describe the linear effect on the uncensored latent variable and not the observed outcome. In other words, if localization were to increase by one unit, the expected count of new firms per capita would increase by 0.00214 while holding all other variables in the model constant (Table 7). The probit model reports marginal effects, i.e. a change in the probability of a positive firm birth for a given change in the explanatory variable. In other words, if localization were to increase by one unit, the probability of a positive number of firm births would increase by 1.508. The OLS coefficients can be interpreted as follows—if localization were to increase by one unit, informal firm births per capita would increase by 84.1%.

I find that on average, localization i.e. intra-industry clustering seems to positively affect the probability of informal firms entrants, and it also results in new firm births per capita. Input–output linkages have a strong positive and consistent effect on the probability of new informal firm entry and count of new firms per capita. This is in line with the results obtained by Ghani et al. (2011a) who find that the strength of buyer and supplier linkages has a positive effect on new employment in informal manufacturing in India. Coagglomeration between formal and informal firms seems to have no discernible effect, although it seems to negatively affect the probability of entry of new informal firms. Access to markets affects the count of new informal births per capita, but not the probability of positive births. There is little concordance between the tobit and OLS models, with the exception of the positive effect of input–output linkages. In other words, for informal firm births, different sets of variables seem to affect the uncensored and unobserved latent outcome (captured by the tobit), and the observed positive outcome (captured by the OLS).

Compared to the results for informal firm births, localization economies seem to have no effect on the probability or the count per capita of formal firm births (Table 8). Instead, a decrease in industrial diversity, given by a rise in the Herfindahl Index (higher levels = lower diversity) seems to positively affect both, the probability and the count per capita of firm births. Again, Ghani et al. (2011a) find a similar result for new employment in formal manufacturing. In addition, similar to the results for informal firms, input–output linkages across industries are a consistently important predictor of formal firm births. The effect of informal–formal coagglomeration increases the probability of births, but its effect on the count per capita is unclear. Accessibility i.e. being located close to larger, more populated districts is not a predictor of firm births, but it does positively affect the count per capita of firm births. The results of the tobit and OLS estimations are similar in terms of direction (if not magnitude) for most variables, with the exception of coagglomeration, for formal firm births per capita. On the other hand, the probit estimations indicate that coagglomeration, along with

[17] I use the dprobit command in Stata that calculates marginal probability effects, in this case, at the sample mean values of the regressors.

Table 7. Informal firm births

Variable	Tobit (1)	Probit (2)	OLS (3)
Localization	0.00214	1.508*	0.841***
	[0.00134]	[0.873]	[0.251]
Input–output	9.51e-05***	0.0360***	0.161***
	[9.69e-06]	[0.00661]	[0.0297]
Herfindahl index	−0.000159*	−0.0366	−0.0699
	[2.51e-05]	[0.0523]	[0.295]
Coagglomeration	−0.00172	−0.911*	−0.837*
	[0.00256]	[0.418]	[0.362]
Market access	3.65e-06	0.0132	0.419***
	[2.32e-05]	[0.0482]	[0.0116]
Wealth	1.25e-05*	0.00687**	0.0101
	[2.27e-06]	[0.00335]	[0.0124]
Year fixed effects	✓	✓	✓
Industry fixed effects	✓	✓	✓
District fixed effects	✗	✓	✓
Observations	5829	5584	3367
R^2	0.0266	0.287	0.458

The dependent variables are as follows: for model (1) it is the count of new informal firms per capita; for model (2) it is a dummy variable indicating zero and positive counts; for model (3) it is the log of count of new informal firms per capita. Marginal effects are reported for the Probit model. Numbers in brackets are standard errors, clustered at the district level; McFadden's Pseudo R^2 values are presented for models (1) and (2).
***$p<0.01$, **$p<0.05$, *$p<0.1$.

input–output linkages and lack of industrial diversity, is a good predictor of positive formal firm births.

Linkages to buyers and suppliers in related industries are a significant and consistent predictor of both informal and formal firm births (recall here that input–output linkages measure the extent to which industries buy from and sell to one another). Informal firm entry is affected by intra-industry clustering, while formal firm entry seems to be attracted to regions with less industrial diversity, but higher informal–formal coagglomeration within a given industry. The percentage of wealthy households positively affects informal firm activity, while access to larger markets positively affects formal firm activity. This suggests that informal firm births might have more to do with population patterns within a location, while formal firm births are drawn to locations with better access to other, larger, locations.

5.1. Differentiating by size

As a robustness check, I carry out the same exercise as above but differentiate between firms of different sizes. Perhaps large firms in the informal sector could be similar to smaller firms within the formal sector—and perhaps it is not formality, but size, that is the important delineating parameter.

I divide firms in the informal and the formal sector by the number of employees. For informal firms, I divide the sample of enterprises into those that are small (i.e. employ

Table 8. Formal firm births

Variable	Tobit (1)	Probit (2)	OLS (3)
Localization	7.41e-07	0.900	0.766
	[1.27e-05]	[0.750]	[2.361]
Input–output	1.04e-06***	0.0665***	0.133***
	[1.11e-07]	[0.00695]	[0.0292]
Herfindahl index	2.37e-06***	0.0723*	0.441**
	[2.60e-07]	[0.0230]	[0.223]
Coagglomeration	9.55e-05**	0.202*	1.673
	[2.95e-05]	[0.097]	[4.001]
Market access	5.16e-07**	0.00260	0.0219**
	[2.45e-07]	[0.0411]	[0.0165]
Wealth	3.64e-08	0.00188	−0.0120
	[2.38e-08]	[0.00302]	[0.0121]
Year fixed effects	✓	✓	✓
Industry fixed effects	✓	✓	✓
District fixed effects	✗	✓	✓
Observations	5829	4223	1303
R^2	0.0344	0.343	0.636

The dependent variables are as follows: for model (1) it is the count of new informal firms per capita; for model (2) it is a dummy variable indicating zero and positive counts; for model (3) it is the log of count of new informal firms per capita. Marginal effects are reported for the Probit model. Numbers in brackets are standard errors, clustered at the district level; Pseudo R^2 values are presented for models (1) and (2). ***$p<0.01$, **$p<0.05$, *$p<0.1$.

less than five workers) and large (i.e. they employ more than five workers). Almost 90% of the firms in the sample, thus defined, are small-scale enterprises. For formal firms, I divide the sample of enterprises into those that are small (i.e. employ less than five workers), medium (i.e. employ more than 5, but less than 100 workers) and large (i.e. employ more than 100 workers). If formal and informal firms are different primarily as a function of their size, then we should expect to see small informal firms behave like small formal firms, and large informal firms to behave like medium formal firms.

Since my interest lies in studying the factors that drive up the count of new firms per capita, and because I would like to control for unobservables at the level of districts, as well as industries and across time, I retain the simple OLS specification. The dependent variable is the log of the count of new firm births per capita.

A few interesting observations emerge (Table 9). Small firms in the formal sector do behave somewhat like their small counterparts in the informal sector—firm births across both types of firms are positively affected by intra-industry clustering, a result that was not observed when firm births in the formal sector were taken as a whole. Large (i.e. more than five employees) informal firms seem to shy away from wealthier locations, same as medium-sized formal firms (i.e. more than 5, but less than 100 employees). However, this is where the similarities end. Input–output linkages across industries are significant for all kinds of firms, except for small formal firms. Industrial diversity, or lack thereof, positively affects formal firm births irrespective of firm size.

Table 9. OLS regressions by size

Variable	Informal		Formal		
	Small	*Large*	*Small*	*Medium*	*Large*
Localization	0.702**	0.545	0.61***	−0.413	−0.884
	[0.267]	[3.251]	[0.095]	[1.113]	[1.862]
Input–output	0.0724**	0.145***	−0.0435	0.104***	0.194***
	[0.0355]	[0.0460]	[0.0330]	[0.0119]	[0.0261]
Herfindahl index	0.236	−0.563	1.158***	0.565***	0.432**
	[0.299]	[0.505]	[0.228]	[0.0989]	[0.189]
Coagglomeration	−8.813	−10.06	0.3***	0.24**	1.047
	[6.823]	[9.243]	[0.031]	[0.079]	[7.451]
Market access	−0.148	−1.271	0.288***	0.00522	0.469**
	[0.240]	[1.611]	[0.0876]	[0.0736]	[0.192]
Wealth	0.0153	−0.0330*	0.0105	−0.0184***	0.00752
	[0.0144]	[0.0196]	[0.0195]	[0.00514]	[0.0112]
Year fixed effects	✓	✓	✓	✓	✓
Industry fixed effects	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓
Observations	1805	1152	293	2776	935
R^2	0.400	0.466	0.901	0.673	0.737

The dependent variable is the log of count of new firms per capita. Numbers in brackets are standard errors, clustered at the district level.
***$p<0.01$, **$p<0.05$, *$p<0.1$.

Formal–informal coagglomeration positively affects small and medium formal firm births—recall that the effect of this variable was insignificant for the full sample of formal firms. It is possible that formal–informal clustering within a given industry might encourage new formal firms if these firms were more likely to buy from or sell to existing coagglomerations of formal–informal firms. Indeed, Holl (2008) finds that industry agglomeration in Spain tends to facilitate subcontracting between large and small manufacturing firms.

Broken down by firm size, the results for input and output economies buttress those observed for the general sample—the effect of these is positive and significant for firm births in both the formal and the informal sectors. The results suggest that larger enterprises are affected more strongly by the clustering of buyers and suppliers in related industries. This is in contrast to a study on Italian firms carried out by Lafourcade and Mion (2003) who find that small firms (i.e. less than 20 employees) are more sensitive to input–output linkages than large firms.

6. Conclusion

This article seeks answers to the following questions: (i) what is the extent of coagglomeration between formal and informal firms in India and what determines these trends; and (ii) to what extent does formal–informal coagglomeration, relative to traditional measures of agglomeration, have an effect on firm births within a location.

The empirical analysis finds that buyer–seller linkages between formal and informal firms within the same industry and technological spillovers are the most significant factors that explain coagglomeration. However, the effect of coagglomeration is not overwhelming in the face of traditional forms of clustering—such as intra-industry clustering (that matters most for new informal activity), or inter-industry clustering (that matters strongly for new formal and informal activity). However, coagglomeration does positively affect births of small and medium-sized firms in the formal sector. This could be a function of labour market regulations in the formal sector, or simply because these firms find it cost-effective for other reasons to outsource production or labour. The research also finds that formal and informal firms in India behave differently with regard to their response to agglomeration, market access and the extent of consumption. Informal firms are attracted to firms in similar industries, formal firms to districts with low levels of industrial diversity. Informal firms tend to start shop in more populated districts, while formal firms do so in districts that are closer to other larger districts.

The importance of this research is underscored by two inter-related factors. First, as Duranton (2008) points out, much more evidence is needed on informal manufacturing activity and its relationship to production externalities in order to design policies that would encourage economic activity in developing countries. Second, this article illustrates how agglomeration economies can arise from the interaction between informal and formal enterprises and to what extent these linkages affect the entry of new firms. The existence of positive production externalities would indicate that policies that try to limit the development of informal activities and attempt to shift the balance to formal activities might be fraught with danger and that it might be efficient to allow these clusters to grow.

There are several directions for further research. A valuable contribution would be to extend the analysis by studying the performance of new entrants over time and in the presence of agglomeration economies, since ultimately policy-makers would be interested in encouraging industrial activity that helps to sustain growth (whether in terms of productivity or jobs). Another interesting question would be the effect of formal–informal linkages on the productivity of the latter, and whether supplying to better or larger firms helps smaller or less formal firms move up the productivity ladder.

Supplementary material

Supplementary data for this article are available at *Journal of Economic Geography* online.

Acknowledgement

This article has been re-worked and re-focussed based on excellent suggestions provided by two anonymous referees and Kristian Behrens. I am also grateful to Diana Weinhold, Henry Overman, Stephen Gibbons, Holger Görg, Waltraud Schelkle, Maarten Bosker, Harry Garretson and Walker Hanlon for helpful comments and suggestions, to Michaela Fuchs at the European Regional Science Association conference in Sweden, to Elisabet Viladecans-Marsal at the Urban Economic Association conference in Denver, and to Vernon Henderson, Nathanial Baum-Snow, Adam Storeygard and other participants at the Urban Studies seminar at Brown University for useful discussions.

References

Arimah, B. C. (2001) Nature and determinants of linkages between informal and formal sector enterprises in Nigeria. *African Development Review*, 13: 114–144.

Bacolod, M., Blum, B., Strange, W. C. (2008) Skills and the city. *Journal of Urban Economics*, 65: 127–135.

Bacolod, M., Blum, B., Strange, W. C. (2009) Urban interactions: soft skills vs. specialization. *Journal of Economic Geography*, 9: 227–262.

Bairagya, I. (2010) Liberalization, informal sector and formal-informal sectors' relationship: a study of India. Paper prepared for the 31st General Conference of the International Association for Research in Income and Wealth, St Gallen, Switzerland.

Barrios, S., Bertinelli, L., Strobl, E. (2006) Co-agglomeration and growth. *Regional Science and Urban Economics*, 36: 467–481.

Carr, M., Chen, M. A., Tate, J. (2000) Globalisation and home-based workers. *Feminist Economics*, 6: 123–142.

Ciccone, A., Peri, G. (2006) Identifying human-capital externalities: theory with applications. *Review of Economic Studies*, 73: 381–412.

Duranton, G. (2008) Cities: engines of growth and prosperity for developing countries? Commission on Growth and Development Working Paper 12, World Bank.

Dutta, N., Kar, S., Roy, S. (2011) Informal sector and corruption: An empirical investigation for India. Discussion Paper No. 5579, IZA, Bonn, Germany.

Ellison, G., Glaeser, G. L. (1997) Geographic concentration of U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105: 889–927.

Ellison, G., Glaeser, E. L., Kerr, W. R. (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100: 1195–1213.

Gabe, T. M., Abel, J. R. (2013) Shared knowledge and coagglomeration of occupations. Federal Reserve Bank of New York Staff Report No. 612.

Ghani, E., Kerr, W., O'Connell, S. D. (2011a) Spatial determinants of entrepreneurship in India. NBER Working Paper 17514.

Ghani, E., Kerr, W., O'Connell, S. D. (2011b) Local industrial structures and female entrepreneurship in India. NBER Working Paper 17596.

Hansen, W. G. (1959) How accessibility shapes land use. *Journal of the American Institute of Planners*, 25: 73–76.

He, C., Ye, X., Wang, J. (2012) Industrial agglomeration and exporting in China: what is the link? *Regional Science Policy and Practice*, 4: 317–333.

Helsley, R. W., Strange, W. C. (2012) Coagglomeration and the scale and composition of clusters. Working Paper.

Henderson, J. V. (2003) The urbanization process and economic growth: the so-what question. *Journal of Economic Growth*, 8: 47–71.

Holl, A. (2008) Production subcontracting and location. *Regional Science and Urban Economics*, 38: 299–309.

House, W. J. (1984) Nairobi's Informal Sector: dynamic entrepreneurs or surplus labor? *Economic Development and Cultural Change*, 32: 277–302.

Howard, E., Newman, C., Tarp, F. (2012) Measuring industry agglomeration and identifying the driving forces. Working Paper No. 2012/84, United Nations University.

Jütting, J., De Laiglesia, J. R. (2009) *Is Informal Normal? Towards More and Better Jobs in Developing Countries*. Paris: OECD Development Centre.

Lafourcade, M., Mion, G. (2003) Concentration, spatial clustering and the size of plants: disentangling the sources of co-location externalities. CORE Discussion Paper No. 2003/91.

Lall, S. V., Koo, J., Chakravorty, S. (2003) Diversity matters: the economic geography of industrial location in India. Policy Research Working Paper 3072, World Bank.

Lall, S. V., Chakravorty, S. (2005) Industrial location and spatial inequality: theory and evidence from India. *Review of Development Economics*, 9: 47–68.

Lall, S. V., Shalizi, Z., Deichmann, U. (2004) Agglomeration economies and productivity in Indian industry. *Journal of Development Economics*, 73: 643–673.

Lall, S. V., Mengistae, T. (2005) The impact of business environment and economic geography on plant-level productivity: an analysis of Indian industry. Policy Research Working Paper No. 3664, World Bank.
Marjit, S. (2003) Economic reform and informal wage – a general equilibrium analysis. *Journal of Development Economics*, 72: 371–378.
Marjit, S., Kar, S. (2009) A contemporary perspective on the informal labour market: theory, policy and the Indian experience. *Economic and Political Weekly*, XLIV(14).
Marshall, A. (1919) *Industry and Trade*. London: Macmillan.
Mitra, A. (2009) Impact of trade on service sector employment in India. Institute of Economic Growth, Working Paper Series No. E/298.
Mukim, M. (2011) Industry and the urge to cluster: a study of the informal sector in India. SERC Discussion Paper No. 72.
Mukim, M., Nunnenkamp, P. (2012) The location choices of foreign investors: a district-level analysis in India. *The World Economy*, 35: 886–918.
Overman, H. G., Puga, D. (2010) Labor pooling as a source of agglomeration: an empirical investigation. In E. L. Glaeser (ed.) *Agglomeration Economics*. Ch. 4, pp. 133–150. University of Chicago Press. Centre for Economic Performance, London School of Economics and Political Science.
Overman, H. G., Venables, T. J. (2005) Cities in the developing world. Centre for Economic Performance Discussion Paper No. 695. London School of Economics.
Pieters, J., Moreno-Monroy, A. I., Erumban, A. A. (2010) Formal-informal linkages and informal sector dynamics: evidence from India. Paper prepared for the 31st General Conference of the International Association for Research in Income and Wealth, St Gallen, Switzerland.
Pieters, J., Moreno-Monroy, A. I., Erumban, A. A. (2011) Outsourcing and the size and composition of the informal sector: evidence from Indian manufacturing. Working paper draft (mimeo).
Porter, M. E. (1990) *The Competitive Advantage of Nations*. New York: The Free Press.
Ranis, G., Stewart, F. (1999) V-goods and the role of the urban informal sector in development. *Economic Development and Cultural Change*, 47: 259–288.
Rosenthal, S. S., Strange, W. C. (2001) The determinants of agglomeration. *Journal of Urban Economics*, 50: 191–229.
Rosenthal, S. S., Strange, W. C. (2003) Geography, industrial organization and agglomeration. *The Review of Economics and Statistics*, 85: 377–393.
Venables, A. J. (1996) Equilibrium locations of vertically linked industries. *International Economic Review*, 49: 341–359.