The London School of Economics and Political Science

ESSAYS IN TRADE AND ECONOMIC GEOGRAPHY

Megha Mukim

A Thesis submitted to the Department of International Development of the London School of Economics and Political Science for the Degree of Doctor of Philosophy.

December 2011

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Abstract

This thesis tests the predictions of theoretical models of trade and economic geography using micro-data from India. As part of a large, poor and rapidly developing country, Indian households receive a disproportionate share of attention from development economists. However, there remain large gaps in the understanding of its other micro-entities – firms.

In Chapter 1, I use detailed panel-level data on 8,253 manufacturing firms from 1990 to 2008 and demonstrate how firms that export differ from their counterparts who cater to the domestic market. After identifying the extent to which the act of exporting drives these differences, I provide evidence that Indian exporters performed better than non-exporters at the outset, and that exporting positively impacts further productivity increases.

In Chapters 2, 3 and 4, I focus on how economic activity in India organises itself along economic geography factors. Chapter 2 studies firms in the Indian informal sector, who have largely escaped close scrutiny before. Using data from national sample surveys on over 4 million manufacturing and services enterprises, I find that firms choose to locate in particular districts across the country. I show that existing agglomeration within these locations, such as that of intermediate buyers and suppliers, is driving the location decisions of new firms. In Chapter 3, using previously inaccessible data on inward FDI, I find that foreign investors also show evidence of clustering and that existing agglomeration and the business environment jointly drive this behaviour. In Chapter 4, I collect data from the Indian Patent Office and my analysis concludes that regional innovation is largely a function of public research and development and economic clustering.

In summary, this thesis uses new data and robust methodological approaches to provide important economic insights into the workings of firms in India and the factors affecting their productivity and their location decisions.

Acknowledgements

I would like to thank my supervisor, Diana Weinhold, for always calling a spade a spade. Her no-nonsense approach did little for my confidence, and consequently, wonders for my productivity. Robert Wade provided generous advice and guidance along the way. I am also grateful to my PhD Committee, namely Henry Overman and Alan Winters, for valuable feedback and suggestions.

A number of people have advised me, prodded me on and helped me in a myriad of ways – most of these are thanked in individual essays. Countless others have shared their *gyaan* chatting in corridors or at lunch or at the pub. I am also grateful to a number of academics, from the LSE and other universities, who took out the time to sit down and talk through my preliminary ideas with me – they were under no obligation to do so and doubtless had many other pressing duties.

Karan Singh helped with data. Philipp Krause bucked me up at my initial upgrade defence and encouraged me every step of the way since.

Funding from the Department of International Development and the Asia Research Centre at the London School of Economics is gratefully acknowledged.

This thesis is dedicated to my mother and father – they have made everything possible.

INTRODUCTION

This dissertation consists of four chapters focussing on aspects of trade and agglomeration across firms and districts in India. Each chapter analyses an empirical question using data from India. While the first chapter lies squarely in the field of empirical trade economics, the second, third and fourth chapters relate more closely to the economic geography literature. In Chapter 1, I address the impact of exporting on productivity using firm-level data. In Chapters 2 and 3 I study the factors driving the location decisions of informal sector firms and of foreign investors respectively. Chapter 4 analyses how patenting activity is related to specific attributes of a given location. This introduction will provide a brief summary of each of the chapters in the thesis and highlight their key findings and contributions.

In Chapter 1, I examine the link between exporting and productivity for manufacturing firms in India. I use data for over 8,000 firms for the period 1990-2008 from Prowess, a database that tracks firms over time. The database provides me with information on inputs and outputs of firms with which I compute a measure of firm-level productivity. I also know whether and to what extent firms in the sample export. With this data, I am able to isolate the effect that runs from participation in export markets to productivity. This long data panel also allows me to differentiate between the within-industry (or cross-sectional) effect and the within-firms (or time-series) effect. I also show that controlling for any unobserved characteristics of the firm is important, since these could be driving the differences between exporters and non-exporters.

The literature demonstrates that exporters are more productive than non-exporters. However, two mechanisms could underlie this difference – exporters could be more productive than non-exporters because they export, or because they were more productive from the outset. Following other influential papers in the literature, I first control for the self-selection effect and then compare exporting firms with non-exporting firms. I use propensity score matching to identify a control group, i.e. a firm that does not export but is similar to a given exporter in every observable respect. I find that exporters are indeed more productive than non-exporters. This methodology provides me with the within-industry average treatment effect. My estimates suggest that entry into export markets increases the productivity of exporters by 11.03 percentage points. However, the marginal effect of continuing to export is statistically insignificant and after five years of exporting, the treatment group (i.e. exporters) are no longer different from the control group (i.e. matched non-exporters).

The paper also estimates the treatment-on-treated effect. I exploit the long panel characteristics of the data and include firm fixed-effects, which allow me to control for any unobservables at the level of the firm and identify the effect of exporting on a given firm's productivity. To isolate the effect that runs from entry into export markets to productivity I use effectively applied tariffs faced by firms. The intuition is that a fall in tariffs would encourage firms at the margin to start exporting, which would, in turn, affect their productivity. After instrumenting with effectively applied tariffs I find that starting to export increases productivity by 24.71 percent. However, the robustness tests lead me to conclude that the effect of continuing to export remains marginal.

In Chapter 2, I study the location decisions of firms in the informal sector in India, in manufacturing and services industries. The informal or the unorganised sector accounts for 43 percent of the country's non-farm net domestic product and employs over 70 percent of its workforce. Yet, there is little or no understanding of what factors drive these enterprises to locate and cluster in particular parts of the country. I take data for services enterprises from the 57th Round (2001-02) and manufacturing enterprises from the 62nd Round (2005-06) of the National Sample Survey Organisation. The total number of new enterprises in each sector exceeds 2 million, and since I use count models, my dependent variable is firm births by 2-digit industry and district. I use data from the employment and unemployment surveys for the years 1999-2000 and 2004-2005 to construct intra-industry and inter-industry (input and output) measures of agglomeration that vary by 2-digit industry and district. I also control for other geographical factors such as industrial diversity, market access and some attributes of infrastructure.

I find that intra-industry clustering is of overwhelming importance to informal firms, both manufacturing and services, after controlling for industry characteristics. In addition, linkages to buyers and suppliers are also important, while the general business environment seem to be largely extraneous to firms' location decisions. To control for the effect of any unobservable sources of natural advantage that may be driving these results, I introduce district fixed-effects in my estimation that successfully control for any time-invariant characteristics. I find that although the magnitude of the effect of my agglomeration variables is smaller, the direction is unaffected and in most cases, the coefficients remain significant. I also investigate if firms in the informal sector appear to be guided by different motivations to a location than firms in the formal sector, and I further explore co-agglomeration of formal and informal activity, across manufacturing and services industries.

These results are very much in line with intuition since informal firms might be expected to behave differently from those in the formal sector – they may be less encumbered by regulatory structures, they could be less sensitive to wage levels since they mostly rely on own or family-labour, and they could me more mobile, across locations and industries. Indeed I find that the importance of networks of social interaction with intermediate goods' suppliers or final goods' buyers within a location out weighs the importance of infrastructure or institutional factors. And since the effect of public policy on generating such clusters seems unclear, I conclude that government might be limited in its ability to encourage relocation of informal firms.

Chapter 3 also explores the attributes of districts in India that drive the location decisions of firms – except that it studies the case of foreign investors in India. My co-author and I collect data on over 19,500 foreign investment projects approved in India between 1991-2005. The data on investments gives us information on the value, industry and importantly, final location of the investment project. Using count and conditional logit models, the paper investigates to what extent factors such as agglomeration, local business conditions, institutional conditions and the presence of previous foreign investors affect the choice of district. In line with other papers in this literature, we find that foreign investors have a strong preference for locations where other foreign investors are present. This result remains robust across different years, sectors and different types of FDI. Indeed, in a number of cases investors seem to follow not only other foreign investors but also those from the same country of origin.

The paper also studies the importance of different types of infrastructure - educational, financial, transport and power, and of institutions such as labour regulations and social unrest. The impact of access to banks and that of the ease with which states can hire and fire workers seems to matter. There is also evidence to indicate that districts in the neighbourhood of large metro areas do not benefit, in terms of attracting more FDI, from having easier access to larger markets than more remote districts. The paper is unable to deal with omitted variables bias since the model does not reach concordance with the introduction of location fixed-effects. We conclude that path dependence might constrain the influence of regional policymakers.

And lastly, in Chapter 4 I analyse patent applications from firms, individuals and research institutions to determine the extent to which research and development expenditures and clustering of economic activity within a location affect innovative activity. Patent data for India is available from the US, EU and Indian Patent Offices at the level of states in the country. Since states in India are often the size of small countries and can be many times as populous, this paper attempts to carry out the exercise at the level of districts. Data is collected from the weekly journals published by the Indian Patent Office, which contain information on the date of the application, name and address of the investor and the international patent classification code. Owing to limited data on explanatory variables taken from the NSSO, the analysis is restricted to pooled cross-sections. The main variables of interest are private R&D expenditures within and in neighbouring locations, intra-industry clustering and human capital, measured by the proportion of the population with a high-school degree, or with a scientific degree.

I find that private R&D expenditures, education and intra-industry, all have a positive and statistically significant effect on the rate of patenting activity within a district. Although the regressors are lagged, I include year, industry and district fixed effects to deal with any omitted variables bias. I find that the effect of R&D remains broadly stable, that the effect of intra-industry clustering also remains significant and that industrial diversity continues to matter. The effect of education also remains broadly stable across these specifications. I conclude that public policy might be hard pressed to influence private R&D expenditures but that it could play an important role in investing in education to raise the level of human capital.

These papers contribute to the existing literature in a number of ways. India is a large, rapidly developing country and whilst it has been of much interest to development economists, the lack of sound data has hampered research in other fields. For instance, the question of whether or not exporting affects productivity has been around for a number of years, and this is a first paper that offers robust empirical findings using data from firms in India. Chapter II studies the case of informal firms in India and the factors that affect their location decisions. These are firms that account for an overwhelming proportion of GDP and employment in the country, and yet that have been ignored in previous firm location studies. Chapter III uses a new dataset on the activities of foreign investors in India to determine their location choices – lack of data has hindered others from tackling the same question. And Chapter IV uses data collected from the Indian Patent Office to determine what aspects of a location might encourage patenting activity.

This dissertation makes valuable contributions in aspects other than providing a better understanding of firm behaviour in India. Chapter 1 is one of the few attempts in the empirical trade literature that separates the within-industry effect from the within-firms effects. While economists care about the effects of exporting on average industrial productivity and to what extent this is driven by the reallocation of productivity across firms, they also care about what happens to average firm-level productivity over time. The analysis in Chapters 2, 3 and 4 is also carried out at the level of the district – a spatial unit that corresponds well to American counties or to the EU NUTS 3 classification, and importantly to Marshall's notion of agglomeration. I use household and enterprise data from the National Sample Survey Organisation, which allows me to construct measures of agglomeration across the 604 districts in the country. Studies that have analysed FDI stocks and flows or patenting activity in India have done so primarily at the level of states.

I also use different techniques to address possible endogeneity bias in these chapters. In Chapter 1, I use propensity score matching and instrumental variables techniques to control for the self-selection of more productive firms into the export market. Indeed, the use of effectively applied tariffs as an instrument to control for self-selection could easily be replicated in other studies. In Chapters 2, 3 and 4 I use industry and location fixed effects to control for omitted variables bias.

In summary, this dissertation makes some key contributions to the existing empirical literature. It does this by using existing data innovatively and by collecting, extracting and assembling new data to address important questions in the field of trade and of economic geography. Not only does the dissertation apply cutting-edge techniques to deal with troubling questions of endogeneity, it also illustrates the shortcomings and the applicability of these methods in different settings. And lastly, it provides important economic insights into the workings of firms in India and the factors affecting their productivity and their location decisions.

TABLE OF CONTENTS

CHAPTER 1

Does Exporting increase Productivity? Evidence from India*

* I am grateful to Diana Weinhold, TN Srinivasan and James Harrigan for helpful comments and suggestions, to David Weinstein, Amit Khandelwal and other participants at the Trade Colloquium at Columbia University, to Ana Fernandes, HL Kee and other seminar participants at the World Bank, to Patrick Conway and other participants at the Rocky Mountain Empirical Trade Conference, and to Michael Ferrantino and other participants at the US International Trade Commission for useful discussions.

Abstract

This paper identifies two separate effects – that of exporting on productivity across firms, and that of starting to export on productivity within firms. It uses detailed panel data from 1990 to 2008 for over 8,000 manufacturing firms in India across 4-digit product categories. The findings show that exporting is associated with a jump in productivity, both within industries and within firms, but that this effect tapers over time. Entry into export markets has a positive effect on firm performance in the very beginning, but there is no evidence of sustained learning-by-exporting.

JEL Classification: F14, D24
Keywords: Exporting, Productivity, Within Firms versus Across Firms, India

1.1 Introduction

In this paper, I examine the link between exporting and productivity for manufacturing firms in India. I study firm-level data in India at the National Industrial Classification (NIC) level for the period 1990-2008. With the exception of the last few years, this period saw significant entry into exports markets – see Figure 1.A.1. I find three main results. First, I demonstrate that firms' productivity is positively related to their participation in export markets, across firms and within firms. Second, I show that part of the increase in productivity is accounted for by learning-by-exporting, after having controlled for any self-selection into exporting. And third, I find that the learning-by-exporting effect is the highest immediately subsequent to entry into exporting and then begins to level off.

An important empirical finding of the literature (Roberts and Tybout 1997, Clerides et al 1998, Bernard and Jensen 1999, Van Biesebroeck 2006, Alvarez and Lopez 2005) is that exporters are more productive than non-exporters. There are two mechanisms that can explain this difference - the first is related to self-selection and the second to learning-by-exporting. Exporters may be more productive than their counterparts, who only supply the domestic market, simply because more productive firms are able to engage in export activity and compete in international markets. The second, more important mechanism, and one which this paper will focus on, is learning-by-exporting by firms, in other words, post-entry productivity benefits. The idea being that when firms enter export markets they gain new knowledge and expertise, which allows them to improve their level of efficiency. However, the two effects are not mutually exclusive – it is possible that high productivity firms that enter the export market continue to improve their productivity because of their exposure to exporting.

The remainder of the paper is organised as follows: The next section provides a descriptive overview of the theoretical and empirical literature on exporting and firm productivity. Section 1.3 describes the data and shows that exporters are in fact different from non-exporters in a number of ways. Section 1.4 outlines the empirical specification for unbiased production function estimates, and the identification of gains from exporting. Section 1.5 carries out robustness checks and Section 6 concludes.

1.2 Theory and Evidence

The potential link between trade and economic growth has been fundamental to international and to development economics. This paper concerns itself with the question of whether firms achieve higher productivity by becoming exporters.

The empirical and theoretical literatures have moved forward in fits and starts. Earlier endogenous growth models (Grossman and Helpman 1991, Rivera-Batiz and Romer 1991) predicted that international technology diffusion through exposure to export markets could boost within-plant productivity[1]. The traditional export-led growth hypothesis (Kaldor 1970, Dixon and Thirlwall 1975) posited that external demand would enable firms to exploit economies of scale[2] leading to productivity growth. Firms could also invest in productivity-enhancing technology in anticipation of larger export markets (Yeaple 2005).

Another microeconomic channel is the reallocation of economic activity across firms within industries. Models by Helpman and Krugman (1985) and Krugman (1994) predicted that average productivity could rise if resources were shifted to industries with lower average costs. Heterogeneous firm models (Melitz 2003 and Bernard et al 2003) also argue that the existence of trade costs allows only the most productive firms to enter export markets. As low productivity firms exit, output and employment are reallocated towards higher productivity firms and average industry productivity increases. In other words, it is the reallocation of activity across firms, and not within-firm productivity growth, that drives industry-level productivity.

A slew of papers test the predictions of these theoretical models, and their empirical findings demonstrate that differences between exporters and non-exporters could arise whether or not exporting enhanced productivity.

---

[1] Industrial-level productivity could also rise if individual firms' new technological learning spilled-over and positively affected the total stock of knowledge for all firms, thus raising aggregate productivity.

[2] Firms move to a lower point on the average cost curve since a rise in output is accompanied by a less than proportionate rise in average costs.

There are some studies that find little of no evidence of learning-by-exporting. See Table 1.A.1 for a summary of some key studies. For instance Bernard and Jensen (1999) find that the benefits from exporting for American firms are unclear. Although employment, growth and profitability are higher for exporters, productivity and wage growth is not superior. Kim (2000) finds only marginal increases in productivity following trade liberalisation in Korea. Delgado et al (2002) find evidence of higher productivity for exporters versus non-exporters and attribute this to the self-selection of more productive firms into the export market. They do not find much evidence to support the learning-by-doing hypothesis, and if so, only for younger exporters. Castellani (2002) finds evidence of productivity gains associated with increases in export intensity. Other studies, such as Isgut (2001) and Clerides et al (1998), the latter using data from several countries, also conclude in favour of the self-selection and against the learning-by-exporting hypothesis. Only the most productive firms have a sufficient cost advantage to overcome transportation costs and compete internationally. Exporters are more productive than non-exporters, not because there are any benefits associated with export activities, but they are simply more productive at the outset. Some studies look at trade liberalisation in general and not just exporting. Hung et al (2004) find that exporting activity itself does not seem to promote productivity in the US, and that it is import competition that attributed for the largest part of labour productivity growth in manufacturing during 1996-2001. Pavcnik (2002) and Fernandes (2007) find that it is trade liberalisation more generally that has a strong positive impact on firm-level productivity, in Chile and Columbia, respectively. Indeed the former is one of the few papers in the literature that is able to identify both, the within-industry and the within-plant effects.

However, some studies do find empirical support for post-entry productivity gains. For instance, Kraay (1999) for China, Bigsten et al (2004) for sub-Saharan Africa, and Aw et al (2000) for Taiwan, find evidence of learning-by-exporting. Loecker (2007) finds that Slovenian export entrants become more productive once they start exporting, and that the productivity gap between exporters and their domestic counterparts widens over time. He also finds that productivity gains are higher for firms exporting to higher-income regions. Biesebroek (2005) finds evidence that sub-Saharan exporters are more productive than their counterparts who only serve the domestic market, and that the former enjoy increasing rates of productivity growth. Other examples of some studies are Castellani (2002), Baldwin and Gu (2003, 2004), Blalock and Gertler (2004), Girma et al. (2004) and Greenaway and Kneller (2008). Park et al (2010) study firms in China and find that the productivity gains are greater for firms that export to more developed markets. However, not all studies are able to describe the source of these learning effects. A notable exception is Baldwin and Gu (2004). From their analysis of Canadian plants they conclude that exporters learn from participation in export markets through channels that include new innovations, as well as technology transfer from abroad and investments in absorptive capacity such as human capital.

A recent working paper by Tabrizy and Trofimenko (2010) studies the effect of exporting on productivity for firms in India. They find no evidence of post-entry productivity gains and conclude that productivity differences between exporters and non-exporters are explained only by self-selection. There are some crucial differences between this paper and their study. Although the sample of firms comes from the same data source, their period of study covers only a ten-year period (1998-2008) whilst this paper studies firms over a twenty-year period (1989-2008). Additionally their measure of productivity, which has been calculated using Levinsohn-Petrin techniques, controls for simultaneity bias but does not control for endogenous exit. As capital-intensive firms are better able to weather a negative productivity shock and thus more likely to survive in the market, and since exporters also tend to be more capital-intensive than non-exporters, this suggests that their productivity estimate does not account for the large downward bias on the capital coefficient. They also deflate firm-level data using a national wholesale price index and not industry-specific input and output deflators. Most importantly, their paper describes pre- and post-entry productivity differentials between exporters and non-exporters mainly by using dummy variables for different types of exporting behaviour.

It is interesting that for the same countries different studies find confirming or conflicting evidence of learning-by-exporting effects, albeit, not always for overlapping periods in time. For instance, both papers that study the United States (Bernard and Jensen 1999, and Hung et al 2004) find no evidence of learning-by-exporting. There are three studies for Germany, of which the first finds a positive result (Bernard and Wagner 1997), while the later two (Wagner 2002, and Arnold and Hussinger 2004) find no effect. Both studies for the United Kingdom (Girma et al 1004 and Greenaway and

Kneller 2008) find positive effects. For Columbia, the first two studies (Clerides et al 1998 and Isgut 2001) find no evidence, while a later study (Fernandes 2007) finds some evidence. It seems that if a study focuses on a developed country or if it uses cross sectional analysis, it is less likely to find learning-by-exporting effects.

It is also possible that firms in developing countries that are further away from the technological frontier and that export to other, perhaps more developed markets, also make larger strides in productivity increases. According to the Global Economic Prospects report (World Bank 2008), progress in developing countries reflects the absorption of pre-existing technologies and not at-the-frontier inventions. The technology achievement index[3] for developing countries clearly shows that India is a laggard with a score of 0.04 – the highest value is 0.25 for the United States. Thus, one might expect Indian exporters to enjoy large productivity increases as compared their other domestic counterparts, whilst US exporters may not be much more productive than non-exporters.

1.3 Data and preliminary analysis

1.3.1. Data description

Firm-level data on output and inputs is drawn from the Prowess database. Prowess is a corporate database that contains normalised data built on a sound understanding of disclosures of over 20,000 companies in India. The database provides financial statements, ratio analysis, fund flows, product profiles, returns and risks on the stock market etc. The Centre for Monitoring of the Indian Economy (CMIE), which collects data from 1989 onwards, assembles the Prowess database. The database contains information on 23,168 firms in agriculture, mining, manufacturing and services for the years 1989 – 2008, yielding a total of 437,283 observations. On average there are 8

[3] The technology achievement index is published by the United Nations Development Programme and combines (a) the indicators of human skills (mean years of schooling in the population age 15 and older and enrollment ratio for tertiary-level science programs); (b) the diffusion of old innovations (electricity consumption per capita and telephones per capita) and of recent innovations (Internet hosts per capita and high- and medium-tech exports as a share of all exports); and (c) the creation of technology (patents granted to residents per capita and receipts of royalties and license fees from abroad). The index is constructed as simple averages of these indicators within subgroups and then across groups.

years of data on each firm. However, data are either not available or are reported as missing values for a number of observations for different variables such as sales, capital stock and wages. This paper focuses mainly on firms in the manufacturing sector, since their exporting behaviour is more easily observable. After cleaning the data[4], the final dataset contains 8,253 firms for the years 1989-2008, yielding a total of 69,286 observations.

Table 1.1: Data Summary

	Mean	Std. Dev	Min	Max
Firms (#)	8,253			
Observations (#)	69,286			
Sales	1,103	10,556	0	842,770
Gross Assets	1,881	19,614	0	1,352,683
Investment	24	250	0	14,239
Wage Bill	92	901	0	56,568
Raw Materials Bill	523	4,591	0.03	269,567
Electricity Bills	50	345	0	14,989
Age	19	18	0.25	173

Note: Sales, assets, wages, raw materials and electricity are reported in INR ’000s.

There is a large degree of firm heterogeneity in terms of size and age – see Table 1.1. Firms in the sample also include both exporters and non-exporters – a total of 5,191 firms enter the export market at least once over the period of study. Some caveats should be mentioned here. It is not mandatory for firms to supply data to the CMIE, and one cannot tell exactly how representative of the industry is the membership of the firms in the organisation. Prowess covers 60-70 percent of the organised sector in India, 75 percent of corporate taxes and 95 percent of excise duties collected by the Government of India (Goldberg et al 2010[5]). Large firms, which account for a large percentage of industrial production and foreign trade, are usually members of the CMIE and are more likely to be included in the database. This also explains why more than 60% of the firms in the sample enter the export market at some point over their lifetime. And so, the analysis is based on a sample of firms that is, in all probability, taken disproportionately from the higher end of the size distribution. As Tybout and

[4] I exclude observations for which data on sales, gross assets and wages are missing. I check whether these values are systematically missing for particular industries, years or types of firm (by age, and by type of ownership), and find that this is not the case.
[5] Quoted in earlier version of NBER working paper.

Westbrook (1994) point out, a lot of productivity growth comes from larger plants, and so a more comprehensive study might have found smaller average residual effects.

1.3.2 Preliminary Analysis: Are exporters different?

At the outset I am interested in knowing whether the facts found in the literature – that exporters differ from non-exporters – also hold for firms in India. By regressing firm characteristics on a dummy for whether the firm exports, a number of studies have documented that exporters differ from non-exporters in important ways. Following Bernard and Jensen (1999) and others, I run the following OLS regression that tells me whether firms that export are different from those that don't:

$$\ln x_{it} = \alpha + \beta EXP_{it} + \gamma Controls_{it} + \sum_t \delta_t Time_t + \sum_k \lambda_k Ind_k + \sum_j \xi_j District_j + \varepsilon_{ikt} \qquad (1)$$

where x refers to the characteristics of firm i at year t active in industry k in district j, EXP is an export dummy equal to one when the firm is an exporter and zero otherwise. Firm-specific controls include the size (number of employees), the age and the type of firm (private domestic, private foreign, public or mixed). I also control for industry, year and location effects, where subscripts k, t and j run through the number of industries (Ind), years ($Time$), and districts ($District$) respectively. In total, there are 22 2-digit NIC industries, 20 years (1989-2008) and 265 districts. The coefficient β reveals to what extent exporters differ from non-exporters, within the same year, industry and district. The results are presented in column (1) of Table 1.2.

However, this regression doesn't say anything about whether there is something about the act of exporting that makes exporters different from non-exporters. Indeed, if I re-run regression (1), comparing exporters with non-exporters *before* the former started to export, I find that exporters were different from the outset – see column (2) of Table 1.2.

Thus, to know if exporting is truly associated with any changes in firm characteristics, I run the following OLS regression that tells me whether participation in export markets is associated with differential characteristics for *a given firm*:

$$\ln x_{it} = \alpha + \beta EXP_{it} + \gamma Controls_{it} + \sum_i \delta_i Firm_i + \sum_t \delta_t Time_t + \varepsilon_{it} \qquad (2)$$

where x refers to the characteristics of firm i at time t active in any particular industry and location. EXP is an export dummy equal to one when the firm exports and zero otherwise. Firm-specific controls include the size and the age of the firm. The difference is that since I am now mainly interested in within-firm variation with regards to exporting over time, I include firm and year fixed effects. The coefficient β reveals whether a given firm is different with regard to exporting. The results are reported in column (3) of Table1.2.

The results show that exporters are indeed different from non-exporters: they have a higher wage bill (128 per cent higher), operate on a larger scale, add higher value, sell more and invest more than non-exporters. However, this is not very informative in itself because exporters were different from the outset. Thus, what is pertinent is that, for a given firm, participation in export markets is associated with a higher wage bill (57 per cent higher), more assets, more value added, higher sales and more investment.

Table 1.2: Firm characteristics and exporting

Firm Characteristic	β			R^2		
	(1)	*(2)*	*(3)*	*(1)*	*(2)*	*(3)*
Average wage bill	1.28***	0.55***	0.57***	0.73	0.77	0.90
Gross Assets	1.12***	0.55***	0.41***	0.68	0.75	0.92
Gross value added	1.33***	0.62***	0.50***	0.68	0.75	0.86
Sales	1.53***	0.59***	0.81***	0.65	0.70	0.84
Investment	1.02***	0.41***	0.27***	0.60	0.72	0.65
Productivity	0.11***	0.08**	0.26***	0.76	0.75	0.83
Observations (min/max)	34,429/41,118	12,743/16,407	38,759/41,920			

*** p<0.01, ** p<0.05, * p<0.1. Notes: Specification (1) refers to Equation (1), wherein industry, district and year fixed-effects are included; Specification (2) is similar, except that exporters are compared with non-exporters before they started to export; Specification (3) refers to Equation (2) wherein firm and year fixed effects are included.

While others in this field have compared exporters to non-exporters within a given industry, location and/or year and found differences (see Table 1.A.1), the data for firms in India clearly shows that these differences could easily be driven by intrinsic pre-exporting differences. And that only by studying firm characteristics for a given firm can one truly identify the association, if any, between the act of exporting and productivity.

It should be kept in mind that nominal values are deflated using NIC 2-digit level output and input specific price indices. Since more productive firms are likely to have a lower-than-average firm specific price, the use of industry price indices might systematically underestimate the output of more productive firms and therefore underestimate their productivity. On the other hand, if exporters were more likely to use better quality inputs and materials, then using industry-specific deflators would overestimate productivity. The converse would be true for less productive firms. In the absence of firm-specific prices I am unable to overcome this bias.

The strong positive association between a given firm's characteristics and its participation in export markets could reflect the decision of better firms to self select into the export market and/or it could reflect the effect of exporting on the firm. Ultimately I am interested in the effect of exporting on productivity - as Paul Krugman said 'Productivity isn't everything, but in the long run it is almost everything'[6]. The next chapter will deal with the computation of productivity and will then go on to disentangling the effect of exporting on productivity by controlling for the self-selection effect, across firms and within firms.

1.4 Empirical Specification

Following the influential papers of Bernard and Jensen (1999) and Clerides et al (1998), the literature has used mainly two methods to measure learning-by-exporting effects. The first method consists of separating the sample into mutually exclusive groups, such as exporters and non-exporters, to assess differences in plant performance between these groups (see Loecker 2007, Greenaway and Kneller 2008, Girma el al 2004). The

[6] Paul Krugman (1994) The Age of Diminished Expectations, MIT Press.

second method of measurement of learning-by-exporting effects consists of one or more dummies for lagged export participation in a regression explaining some measure of firm performance. For example, Clerides et al (1998) regress average variable costs on lagged export participation controlling for real exchange rate, lagged capital stock and lagged average variable costs. Kraay (1999) regresses three alternative measures of performance (labour productivity, TFP, and unit costs) on lagged export participation, lagged performance and firm fixed effects. Bigsten et al (2004) and Van Biesebroeck (2004) estimate production functions with a lagged export participation dummy added as a shifter of total factor productivity.

The measure of plant performance used in this paper to assess the presence of learning-by-exporting effects is total factor productivity (TFP)[7]. I will use both methods, propensity score matching to separate the sample into two mutually exclusive groups and instrumental variables to estimate the causal effect of exporting on firm-level productivity.

1.4.1 Estimating Productivity

The estimation of production functions can be affected by two different sources of bias. Since firms' inputs and outputs are simultaneously chosen, inputs will be correlated with any shocks, say demand or productivity shocks, that would be captured in the error term and coefficient estimates will be biased. Under fairly general assumptions[8], Levinsohn and Petrin (2003) show that under simple OLS estimations the labour co-efficient will be upward biased and the capital co-efficient will be downward biased, implying that productivity estimates will be upward biased for more capital-intensive firms (such as exporters). On the other hand, the selection problem is generated by the relationship between the unobserved productivity variable and the shutdown decision. In this case, firms' choices on whether to exit the export market depend on their productivity. Olley and Pakes (1996) obtain consistent production function estimates

[7] TFP measures the economic and technical efficiency with which resources are converted into products.

[8] Levinsohn and Petrin (2003) consider the bias in three different cases: when only labour responds to the shock and capital is not correlated with labour (the labour co-efficient will be biased upwards, and the capital co-efficient will be unbiased); when only labour responds to the shock and capital and labour are positively correlated (the labour co-efficient will be biased upwards, and the capital co-efficient will be biased downwards); when labour and capital respond to the shock, the two are positively correlated and labour responds more strongly to the shock (the labour co-efficient will be biased upwards and the capital co-efficient will be biased downwards).

controlling for the fact that firms' choices on whether to exit the market depends on their productivity[9].

This paper follows Olley and Pakes (1996) – henceforth referred to as OP – to obtain consistent production function estimates. The OP approach uses investment to control for the simultaneity between inputs and outputs. Consider the following production function:

$$Y_{it} = A_{it} L_{it}^{\beta_l} I_{it}^{\beta_i} K_{it}^{\beta_k}$$

where Y_{it} is output, A_{it} is total factor productivity, and L_{it}, K_{it}, I_{it} represent labour, capital and investment, respectively. TFP is modelled as:

$$A_{it} = \exp(\omega_{it} + \varepsilon_{it})$$

where ω_{it} is a firm-specific productivity shock known to the firm manager, but unknown to the econometrician, and ε_{it} is a zero-mean productivity shock realised after variable inputs have been chosen.

The production function of a given firm is described as follows, where output is expressed as a function of the log of inputs and shocks:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_i i_{it} + \omega_{it} + \varepsilon_{it} \tag{3}$$

As mentioned above, there is a possibility that the coefficients on the variable input, i.e. labour, are upwardly biased and that there is a corresponding downward bias in the coefficient on the quasi-fixed input, i.e. capital. To obtain consistent production function estimates, I use the OP procedure. Estimation proceeds in two stages. In the first stage, the coefficients on labour are obtained by semi-parametric techniques. It is assumed that a firm's demand for investment increases monotonically with productivity, conditional on capital. Then the inverse of the investment demand function depends only on observable inputs and capital, and its non-parametric estimate can be used to

[9] Olley and Pakes (1996) use investment as a proxy to control for the simultaneity problem, i.e. when inputs are endogenous to productivity.

control for unobservable productivity, removing the simultaneity bias. Since productivity is assumed to affect capital with a lag[10], there is no simultaneity problem in estimating the coefficient on capital. Loecker (2010) argues that if firms that export are also firms that invest more, then using the OP procedure would overestimate the capital coefficient and underestimate the returns from exporting. I do not include past exporting experience when estimating productivity.

The OP procedure also controls for the endogeneity of firm exit by computing survival probabilities for the firm. The probability that the firm survives in the market depends on lagged values of capital and the proxy for productivity. These probabilities control for the selection bias and are based on some threshold of productivity below which a firm exits the market. The survival probabilities are then introduced into the production function to generate the coefficient on capital. In Appendix A, I discuss the estimation algorithm for getting reliable estimates of the production function in more detail. Nominal values have been deflated using output (sales) and input (labour, capital, investment) deflators[11].

I re-run Equations (1) and (2) with the computed measure of unbiased productivity as the dependent variable. Recall that the first regression controls for district, industry, year and group fixed effects and that the second controls for year and firm fixed effects. I find that exporters are 11 per cent more productive than non-exporters. More importantly, I find that for a given firm, exporting is associated with a 26 per cent increase in productivity. I also run Equation (2) disaggregated by 2-digit NIC industry and find that the positive association between exporting and productivity is broadly positive across firms in all manufacturing industries, with the exception of tobacco products – see Table 1.A.3.

Another way to see this relationship is graphically. I rescale time in such a way so that $t = 0$ refers to the year when the firm begins to export, $t = 1$ is when the firm has exported for a year and so on and so forth. The sample consists of only those firms that

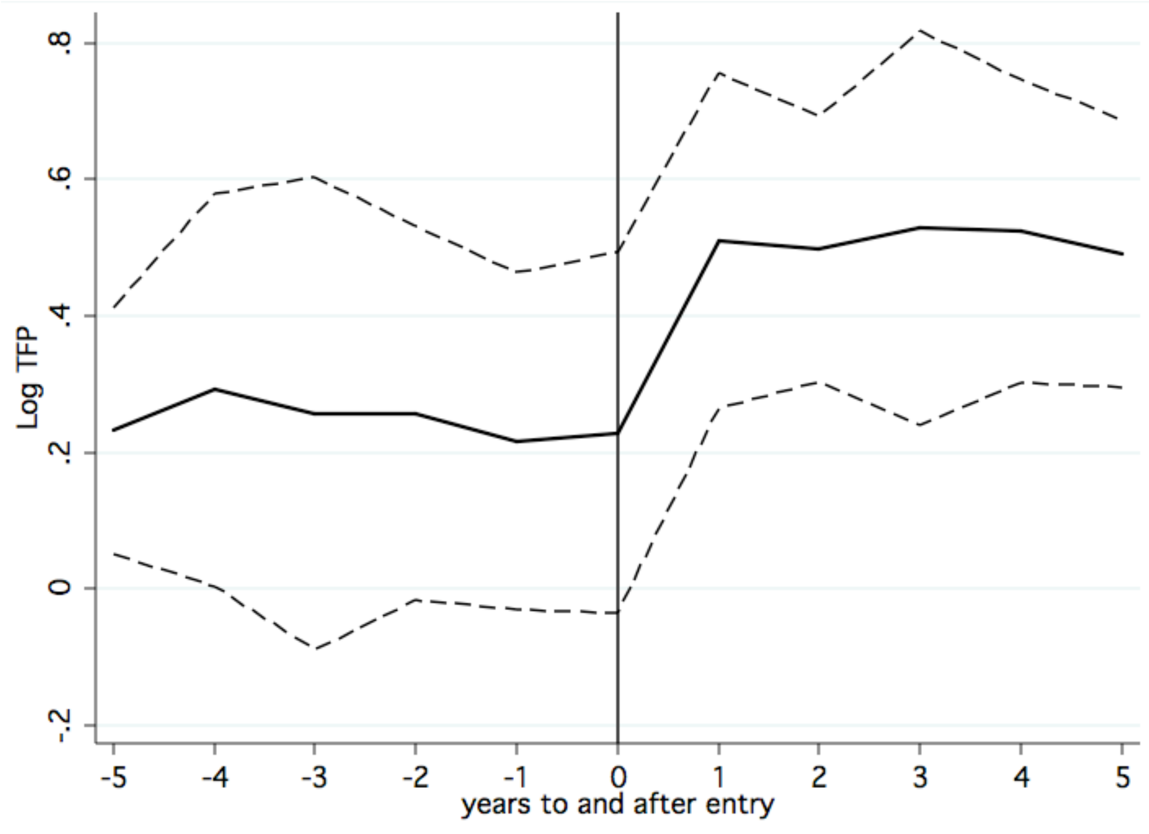
[10] More specifically, it is assumed that productivity follows a Markov process: $\omega_{it} = E[\omega_{it} \mid \omega_{it-1}] + \xi_{it}$ where ξ_{it} represents the unexpected part of current productivity to which capital does not adjust.
[11] Where the input deflators are constructed by their weight in the Consumer Price Index basket, and the data is taken from the Central Statistical Organisation.

export at least once, i.e. 5,191 firms. In the graph the productivity of these firms is averaged for the 5 years before entry into export markets, for the year of entry and then for the subsequent 5 years of exporting. I re-run Equation (2), i.e. with year and firm fixed effects, since I am mainly interested in how productivity for a given firm changes after its entry into export markets and how it changes with continued participation in export markets. The number of firms over which the TFP estimate is averaged falls for every additional year of participation since fewer and fewer firms export continually over the 5 year period. Most firms export for 1-2 years only, after which they either exit the sample altogether or they re-enter export markets – in either case, once they stop exporting, they are no longer included in the sample.

The fitted values that describe the relationship between exporting and productivity are depicted graphically in Figure 1.1. The bold lines refer to the estimated coefficient β, which equals 1 when the firm exports, i.e. at the point of entry and for the 5 subsequent years that the firm continues to export, and which equals 0 for the 5 years before entry. This is after controlling for firm characteristics (size and age) and including firm and year fixed-effects. The dashed lines show the 99% upper and lower bound confidence intervals around the parameter estimates. The levels of TFP jump at the time of entry into export markets and then taper off and do not grow over time. In other words, for firms in the sample that export, average TFP is associated with an increase when the firm first enters the export market and subsequent years of exporting do not seem to be associated with any further increases.

I emphasise here that the regressions carried out in Equations (1) and (2) earlier do not control for endogenous self-selection of more productive firms into exports, but they do find that (1) exporters are different from non-exporters within the same industry, location and year, and that (2) for a given firm, exporting is associated with a jump in productivity.

Figure 1.1: TFP and entry into export markets



1.4.2 Identification of productivity gains from exporting

In the following sub-sections I will identify two separate estimates of exporting on productivity, using propensity score matching and instrumental variables techniques. The first estimate is a within-industry estimate that provides me with the effect of exporting on aggregate industrial productivity after controlling for the self-selection of more productive firms into the export market. This estimate of productivity is directly comparable with other empirical papers in the literature that also compare productivity differentials between exporters and non-exporters within a given industry (and often within a given location and year). The second estimate is a within-firm estimate that provides me with the effect of entry into export markets on aggregate firm productivity after controlling for the self-selection problem. In other words, I will identify the effect of exporting on a given firm over time and the coefficient would no longer be constrained to be the same across all firms. In summary, this section will separate the cross-section variation across firms using Propensity Score Matching (PSM), and the time-series variation within firms using Instrumental Variables (IV).

Across-Firm Estimate (Propensity Score Matching)

Following Girma et al (2004) and Loecker (2007), I control for the self-selection of more productive firms in a given industry into export markets (i.e. the Melitz effect) by creating control groups using matching techniques based on average treatment models as suggested by Heckman et al (1997). The aim of this methodology is to evaluate the causal effect of exporting on productivity by matching export starters with non-exporters. The identifying assumption in estimating the treatment effect (i.e. exporting) comes from the introduction of the state variable – lagged productivity – in the matching procedure. The method constructs a counterfactual that allows me to analyse how productivity of a firm would have evolved if it had not started exporting. The main problem in this type of analysis is that one does not observe the counterfactual and therefore it is necessary to match the exporting firm with a control group of similar firms that do not export.

A straightforward method to assess the impact of exporting would be to compare average outcomes of productivity between firms that exported and those that did not. However, this methodology fails to control for any differences in pre-export characteristics across firms, which could severely bias the estimates if, for instance, more productive or larger firms were more likely to export. Another option might be to run a regression of productivity, the outcome variable, on a dummy variable for whether the firm was active in export markets, conditional on time-lagged observables at the level of the firm, such as age, size and past productivity. This method imposes arbitrary functional form (mostly linear) assumptions concerning the treatment effects and the control variables, that may or may not be accurate and which matching avoids. This is valuable since these functional form restrictions are usually justified neither by theory nor by data (Angrist and Pischke 2008). In addition, the predictor variables in an ordinary least-squares (OLS) regression are included with regards to their exogeneity and suitability to predict the outcome variable, i.e. productivity. In the propensity score matching (PSM) method, the matching covariates or the exogenous variables are chosen with regards to their ability to predict participation, and not their suitability to predict the final outcome. In other words, the PSM technique creates treatment and control groups based on the propensity of firms to export conditional on pre-export characteristics, while the OLS technique would predict the productivity of firms, conditional on pre-export characteristics. The parameters in an OLS regression will be biased even in large samples unless the right-hand side variables are exogenous.

The main problem with the PSM technique is that the matching procedure is based on observables only, and any bias owing to unobservables cannot be ruled out. Jalan and Ravallion (2003) illustrate how qualitative fieldwork can help validate the choice of covariates, which in turn could help minimise the selection bias. Balancing tests could help to test for systematic differences in the covariates between the treatment and comparison groups. Additionally, owing to the need to assure comparability in terms of initial characteristics, some firms have to be dropped owing to lack of sufficiently similar matches. This could create a possible sampling bias in the inferences about the impact to the extent that firms that export (i.e. that receive the treatment) are dropped to achieve common support[12].

However, PSM remains a popular tool for carrying out evaluations more generally, and it is particularly useful for this analysis for a number of reasons. A practical advantage is that it does not require randomisation data and is less costly and time-consuming to implement. The propensity score tries to create the observational analogue of an experiment in which every participant (here, firm) has the same probability of participation (here, in export markets). Unlike regressions, matching does not presume linearity and also helps identify problems with the support of the covariates. This is because, compared to OLS regressions that are run on the entire sample, PSM techniques help to reduce bias in the computation of the impact estimates by restricting the sample to where the conditional probability of participation is the same between participant and comparison groups. Ravallion (2008) provides some examples of studies where PSM techniques compare both, favourably and unfavourably, with OLS and randomisation methods.

The main aim in the following exercise is to evaluate the causal effect of exporting on the performance indicator – here, TFP. Following Loecker (2007) I rescale the time periods in such a way that a firm starts exporting at $s = 0$. Let ω_{is} be the outcome at time s - the productivity of firm i at period s - following entry in export markets at $s = 0$ and the variable $START_i$ takes on the value one if the firm i starts to export. The

[12] See the discussion of the problem of non-overlapping support bias in Heckman et al (1997).

causal effect can be verified by looking at the difference: $(\omega_{is}^{1}-\omega_{is}^{0})$, where the superscript denotes the export behaviour. The crucial problem is that ω_{is}^{0} is not observable. I follow the micro-econometric evaluation literature (Heckman et al 1997) and I defined the average effect of export entry on productivity as:

$$E[\omega_{is}^{1}-\omega_{is}^{0} \mid START_i = 1] = E[\omega_{is}^{1} \mid START_i = 1] - E[\omega_{is}^{0} \mid START_i = 1] \quad (4)$$

The key difficulty is to identify a counterfactual for the last term in Equation (4). This is the productivity effect that entrants in export markets would have experienced, on average, had they not exported. What is mainly of interest is the magnitude of the 'impact', labelled in red in Figure 1.A.3 and the main problem is the calculation of the counterfactual that is to be deducted from the total change.

This counterfactual is estimated by the corresponding average value of firms that remain non-exporters: $E[\omega_{is}^{o} \mid START_i = 0]$. An important feature of the construction of the counterfactual is the selection of a valid control group. In order to identify this group it is assumed that all the differences in productivity (except that caused by exporting) between exporters and the appropriately selected control group is captured by a vector of observables, including the pre-export productivity of a firm. The intuition behind selecting the appropriate control group is to find a group that is as close as possible to the exporting firm in terms of its predicted probability to start exporting. More formally, I apply the propensity score matching method as proposed by Rosenbaum and Rubin (1983). This boils down to estimating a probit model with a dependent variable equal to one if a firm starts exporting and zero elsewhere on lagged observables including productivity.

The probability of starting to export is modelled as follows. $START$ is a dummy variable that equals one at the time a firm starts exporting. The probability of starting to export, i.e. the propensity score, can be represented as follows:

$$\Pr(START_{i,0} = 1) = F(\omega_{i,-1}, CONTROLS_{i,-1}) \quad (5)$$

where $F(.)$ is the normal cumulative distribution function. The re-scaling of the time periods implies that the probability of starting to export is regressed on variables prior to this period $s = 0$ and I use the subscript '-1' to denote this. The most important variable in estimating the propensity score estimation clearly is the lagged productivity variable. Differences in productivity will be conditioned on pre-export levels of productivity and the size and age of the firm. I also include a full set of industry and year dummies to control for common aggregated demand and supply shocks. I use nearest-neighbour one-to-one matching, with replacement[13].

Let the predicted export probability for firm i (which is an eventual exporter) be denoted by p_i. The matching is based on the method of the nearest neighbour, which selects a non-exporting firm j that has a propensity score p_j closest to that of the export entrant. This results in a group of matched exporting and non-exporting firms needed in order to evaluate the causal impact of exporting on productivity. Following both Girma et al (2004) and Loecker (2007) I match within each 2-digit NIC sector and therefore create control groups within narrowly defined sectors as opposed to matching across the entire set of firms. This is likely to be important as the marginal effect of various variables on the probability of starting to export may differ substantially between different sectors due to different technological and market conditions that firms face in different industries. This implies that I estimate the probability to start exporting for each industry separately, allowing the coefficients to vary within the various industries. However, I am unable to control for other differences between firms that produce the same product, for instance, quality, mark-ups, employee skill sets etc.

Once I have this counterfactual in hand I use a difference-in-differences (DID) methodology[14] to assess the impact of exporting on productivity. Following Loecker (2007) the estimator of the learning-by-exporting effect (β_{LBE}) is calculated in the following way. Assume N firms that started exporting and a set C of control firms, with ω^1 and ω^c being the estimated productivity of the treated and the controls respectively. Denote $C(i)$ as the set of control units matched to a firm i with a propensity score of p_i. The number of control firms that are matched with an

[13] Matching with replacement tends to reduce bias, and can be performed in cases where the control group is smaller than the treatment group.

[14] Another option would be to use a double DID wherein the changes over time across the treatment and counterfactual groups would be compared. This method would help to contain the bias owing to unobservables in the data. However, owing to selective retrenchment of firms from export markets over time, I would be faced with a potentially larger bias, since I am unable to follow-up with firms that drop out of the sample over time.

observation i (starter) is denoted as N_i^c and the weight $w_{ij} = \frac{1}{N_i^c}$ if $j \in C(i)$ and zero otherwise. In this way every firm i that started exporting is matched with N_i^c control firms. I stress that the matching is always performed at the time a firm starts exporting and $s = \{1,2,.....,S\}$ denotes the time periods after the decision to start exporting, i.e. at $s = 0$. I introduce two estimators getting at the productivity effect at every time s (Equation 6) and a cumulative productivity effect (Equation 7). The first estimator, β_{LBE}^s at every period s after the decision to start exporting, is given by:

$$\beta_{LBE}^s = \frac{1}{N_s} \sum_i \left(\omega_{is}^1 - \sum_{j \in C(i)} w_{ij} \omega_{js}^c) \right) \tag{6}$$

In words, I estimate the productivity premium of firms that started exporting at each period s compared with (a weighted average of) productivity of a control group based on nearest neighbour matching at every period s. However, since I am also interested in how starting to export impacts the productivity trajectory of a firm, I estimate the average cumulative treatment effect. This is the productivity gain gathered over a period S after the decision to start exporting. The second estimator, β_{LBE}^S, is given by:

$$\beta_{LBE}^S = \frac{1}{N_S} \sum_i \left(\sum_{s=0}^{S} \omega_{is}^1 - \sum_{s=0}^{S} \sum_{j \in C(i)} w_{ij} \omega_{js}^c \right) \tag{7}$$

This provides me with an average cumulative productivity gain at every time period and plotting these estimated coefficients over time gives me a relation between time (s) and the productivity gain. The estimate in Equation (7) gives us the productivity premium new exporters have gathered over time. This implies that the entire productivity path of export entrants is compared to that of the control group, whereas the estimate in Equation (6) estimates the productivity premium at the each time period s.

The firms in the treatment group in the analysis have been matched at the point of entry into export markets with firms in the control group. The control group consists of firms that never export over the period of comparison, i.e. 6 years – the year of entry into export markets and the 5 years thereafter. It is possible, however, that the counterfactual group contains firms that don't export for the duration of the comparison, but which might enter the export market in the future, and which might have exported in the past. Thus, exporters are matched with non-exporters at the point of the formers' entry into export markets; after which, the two sets of firms are followed over time and the average difference in productivity is reported – for each year of exporting, and for each cumulative period of exporting. In other words, if an exporter A^{e} is matched with a non-exporter A^{ne} at time $s=0$, the PSM procedure will compare the productivity differential between the two firms at each additional year of exporting (i.e. $s=1,2...5$). Firms that export are only matched once over the period of analysis, i.e. at the point of entry in export markets. If firms drop out of export markets (i.e. they stop exporting), they are no longer included in the analysis and do not enter the control group of non-exporters.

I find that the first year of exporting is associated with a large increase in productivity (11.03 per cent), but that each subsequent year of exporting is no longer associated with a rise in productivity – see Table 1.3. However, the cumulative effect of exporting on productivity continues to remain positive and significant up until the fourth year of exporting, at which point the productivity increase associated with exporting is approximately 4.46 per cent. By the fifth year of exporting, the cumulative productivity differential is no longer significant. Thus, entry into export markets is associated with a rise in productivity, but the marginal effect of continuing to export seems to be insignificant.

Table 1.3: Estimated learning by exporting effects

s	1	2	3	4	5
Outcome: productivity					
β^{s}_{LBE}	0.1103***	-0.0420	-0.0302	-0.0329	0.0277
	[0.038]	[0.034]	[0.040]	[0.039]	[0.052]
# treated	4324	3538	3204	2786	2445
# controls	3394	2971	2690	2344	2073
Outcome: cumulative productivity					
β^{S}_{LBE}	0.1103***	0.0814***	0.1007***	0.0446*	0.0072
	[0.038]	[0.029]	[0.029]	[0.024]	[0.025]

*** p<0.01, ** p<0.05, * p<0.1
Standard errors in square brackets

Also, the cumulative estimator β^{S}_{LBE} is not equal to the sum of the pure time estimator β^{s}_{LBE} due to the unbalanced data. Formally, $\sum_{s}\beta^{s}_{LBE} \neq \beta^{S}_{LBE}$ since N varies with s. Firms also start to drop out of the sample as the total number of years of exporting (i.e. the value of s) increases. This could be because less productive firms that are unable to survive exit the export market, and/or firms that began to export in recent years drop out because the sample period comes to an end. If the former is true, the average treatment effect will be upwardly biased. However, if the latter is true then we should not expect to see any bias. In the first five years of exporting, I lose between 40 percent of firms because the sample period comes to an end. The remainder of the firms exit the export market, with the largest number of firms exiting the market in the first two years of exporting. And so, the coefficients could be biased upwards since I am unable to control for endogenous exit.

The propensity score matching method also assumes that there exists a region of ‘common support’, where the treated and control propensity scores overlap, and over which a robust comparison can be made. A straightforward way to check the overlap is a visual analysis of the density distribution of the propensity scores across the treated and control groups before and after matching (see Figure 1.A.4). Firms that fall outside of the region of common support are disregarded and for these firms the treatment effect cannot be estimated. With matching the proportion of such firms is small, and since the region of common support is vastly improved the estimated effect on the remaining firms can be viewed as representative. In Table 1.A.4 I provide the t-test of the equality of means between the treated and the matched control group. I find that there is no statistically significant difference in the means of the conditioning variables between the treated and the control group, and that matching reduces the bias by 19 percent for productivity to 95 percent for sales.

Within-Firm Estimate (Instrumental Variables)

The problem with propensity score matching techniques is that it eliminates the selection problem based only on observables, and that it assumes away every possible problem with the error terms – this could include endogeneity and/or measurement error. Additionally, this estimate only provides me with the productivity differential between exporters and non-exporters within a given industry, and says nothing about the within-firm effect of exporting on productivity. In this section, I will use an instrument to deal with the methodological shortcomings of the matching procedure, and to identify that part of the effect of exporting on productivity that is owing to time-series variation[15].

A good instrument would be a variable that affects the productivity of the firm only through its effect on the firm's decision to export, and which would be exogenous to changes in firm-level productivity. I use effectively applied tariffs faced by exporting firms. These tariffs are defined by destination country for each industrial sector at the NIC 4-digit level from 1990 to 2008. Since these are tariffs imposed by destination countries, individual firms in India should not be able to affect the level of tariffs. And in theory, a fall in tariffs should positively affect a firm's decision to enter the export market and thus affect firm-level productivity.

One might argue that increases in firm-level productivity in India could have a direct impact on tariffs faced by exporting firms if destination markets raised their tariffs to protect their domestic markets. If this were the case, one would expect to see a positive relationship between exporting and tariffs. But in fact tariffs and exports are negatively related (see Figure 1.A.5), and so it would seem more likely that lower tariffs drive higher export volumes. Alternatively one may argue that an increase in firm-level productivity leads to an increase in exports, which in turn creates pressure from exporting firms from the originating country, i.e. India, on destination markets to lower their tariffs. In other words, that exports are driving tariffs. However, India's share of the global export market is small, and since independence has varied between 0.5 and 2.5 percent, and between 0.5 and just over 1 percent over the period of study. When disaggregated by sector the range varies between 0.32 and 0.55 percent. When further disaggregated by the top 20 country destinations (that make up almost 75 percent of the total market share of Indian exports), the range varies between 5.72 and 0.09 percent (see Figures 1.A.7 and 1.A.8). So, a more likely story would be that Indian exporters are in fact price takers in global markets[16]. It has been argued that India is able to punch

[15] DiPrete and Gangl (2004) provide a valuable comparison between PSM and IV techniques and illustrate how these can be seen as complementary analyses, and not necessarily substitutes.
[16] It could be argued that Indian exporters may be able to influence tariffs in products where they have a small share of the destination market precisely because the latter might not care

above its weight at the WTO, for instance through anti-dumping investigations. In fact in 2002, India overtook the US to become the highest initiator of anti-dumping cases at the WTO. However, the point of anti-dumping cases is mainly to try and protect domestic markets from cheap imports. For instance, most anti-dumping cases were brought out against China, Brazil, and Taiwan etc. It could, however, be the case that countries may still use anti-dumping as a clever negotiating tool to increase market access.

I use effectively applied tariff rates, in levels and changes, taken from the Trade Analysis and Information System (TRAINS)[17]. The classification system used is International Standard Industrial Classification (ISIC Revision 3), at the 4-digit level, as this corresponds on a one-to-one basis to the Indian National Industrial Classification (NIC) system at the 4-digit level. I re-run Equation (1) with productivity as the dependent variable and I instrument the dummy variable for starting to export with export-weighted tariffs. My instrument, i.e. tariffs, varies by industry and year, whilst my instrumented variable, i.e. the dummy variable for when the firm starts exporting, varies by firms over time. I am unable to find an instrument that is specific to firms, but I use tariffs at the 4-digit industry level to narrow the effect as best I can. As in my earlier OLS specifications, I include firm fixed-effects, which also controls for the differential effects that changes in industry-level tariffs would have on firms depending on their individual characteristics. I also include year fixed-effects. Thus, the instrument is time varying and applies within firms.

It could be argued that a change in tariffs at the 4-digit industry level has little to do with individual firms' decisions to export, and simply affects the industry-level propensity to export. Indeed, if I were interested in the average effect of exporting on industry-level productivity, the instrument, i.e. effectively applied tariffs would apply at the level of the industry. However, it could also be the case that a change in industry-level tariffs affects firms in that industry differently, depending on the individual characteristics of the firm. It's possible that not all firms would respond to an industry-level change in tariffs in the same way – some firms might be more inclined than others

about providing better access, especially if in return it were able to gain concessions in other sectors.

[17] TRAINS data is made available through the World Integrated Trade Solution (WITS) database.

to adjust export behaviour. For instance, Bown and Porto (2010) study the effect of an increase in preferential market access for the Indian steel industry and find that some firms within the industry, such as those which historically had export ties to developed markets, responded more quickly than others in order to increase their exports. Indeed, as their analysis shows, aggregating variables at the industry-level fails to capture the differences across firms, some of which are large producers who were active for a number of years prior to the shock and others that were relatively new entrants to the market[18].

The decisions to export are endogenous choices of the firm and it is easy to imagine ways in which the export status could be correlated with unobserved firm characteristics that directly affect both the level and the growth rate of firm productivity. For example, dynamic firm managers may be more aggressive in entering export markets and also more aggressive in making productivity-enhancing investments. An industry-level drop in tariffs would improve access for all firms within an industry; however, it is likely that some firms might be better poised to exploit the change than others. Regressing industry-level productivity on an aggregate firm-level decisions to enter export markets would mask such individual differences across firms. For instance, we could take the example of the textile industry, at the level of the 4-digit National Industrial Classification 1730, i.e. Manufactures of Knitted and Crocheted Apparel. TFP, defined for a firm for a given year, could be regressed on a dummy variable for whether this was the first year of exporting for that textile firm, with or without firm fixed-effects. In the former case, the average effect of starting to export on productivity is 10.53 percent, while in the latter it is a mere 1.62 percent (the respective R^2 are 0.707 and 0.064). When aggregated across all textile firms within the 4-digit NIC industry, it seems that the effect of entry on productivity is much smaller than if firm-specific unobservables are taken into account. Since both regressions, i.e. with and without firm fixed-effects, control for the age and size of the firms, this could be an indication that other factors, say managerial capacity or supplier relationships, might be important when computing the average effect of entry into export markets on the productivity of textile firms. If some textile firms were better prepared to exploit a change in market access than others,

[18] For instance, when aggregated across all firms, it seems that the share of sales associated with the preferential products seems to fall in response to the increase in market access. This could be because new entrants in the market sell only a small share of preferential products, compared to more established firms, which brings down the aggregate average for all firms.

which in turn would predict the propensity of the firm to start exporting, these differences should be controlled for to arrive at the true effect of export entry on productivity.

Recall, that productivity here has been calculated after having controlled for the potential simultaneity of input choices and unobserved exit, and so I am mainly trying to control for any reverse causality between productivity and the decision of the firm to start exporting. I run the regression using changes and levels of weighted tariffs, using two-stage least squares techniques. I check the exogeneity of the export status using the Durbin-Wu Hausman specification test, and find that the results of the IV estimates are preferable. The instrumented coefficient remains positive and significant. The F-statistic is above the rule-of-thumb value of 10 in the case of changes in tariffs, but this is not the case for tariff levels.

Table 1.4: IV Estimation

Predictors	OLS	IV (Tariffs)	
		Tariffs (Δ)	*Tariffs (level)*
Start	0.1588***	0.2471*	0.8417
	[0.030]	[1.036]	[2.344]
Age	-0.0236***	-0.0202	-0.0223***
	[0.005]	[0.013]	[0.008]
Size	0.0000	0.0000	0.0000
	[0.000]	[0.000]	[0.000]
Firm & Year FE	*Yes*	*Yes*	*Yes*
#	16,444	16,442	15,899
R^2	0.814	0.814	0.815

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in square brackets (clustered at the firm level[19]). The dependent variable is firm-level productivity.

I also find that instrumenting for starting to export raises the coefficient dramatically. With OLS techniques starting to export is associated with a 15.88 percent increase in productivity. However, instrumenting with changes in tariffs raises this effect to 24.71

[19] The residuals could be correlated across firms or time. For instance, economy-level shocks would cause correlation between firms at a moment in time, and persistent firm-specific shocks could cause correlation across time. Thus, I also cluster standard errors by year, but this does not affect the results.

percent. Similarly, when instrumenting with tariff levels the co-efficient rises to 84.17 percent, although the coefficient is not statistically significant. My endogenous variable, exports, is weakly correlated with my instruments, whether changes or levels of tariffs (see Table 1.A.5 for first-stage results). In general, the weaker the correlation between the instrument and the variable being instrumented, the greater is the population variance of the coefficient. The increase in the instrumented coefficient could also be owing to heterogeneity in export effects, implying that the marginal return to exporting is higher for lower productivity firms, who are also more affected by changes in effectively applied tariffs. Van Biesebroeck (2005) finds a similar result using ethnicity of the owner as an instrument, and Card (2001) elaborates on such econometric results when using instrumental variables.

1.5 Robustness and Other Exercises

1.5.1 Intensive Margin

Another interesting sub-question is whether productivity is determined not just by participation in export markets, but also by the intensity of that participation. In other words, firm productivity could be affected by the extensive margin of exporting (i.e. participation), but additionally it could be affected by the intensive margin of exporting wherein firms increase their years and/or intensity of exports. As pointed out by Park et al (2010) one might expect to see continued improvements in productivity with more years of exporting if, for instance, productivity-enhancing investments are lumpy and firms make additional investments following a few years of exporting. Research by Blalock and Gertler (2004) and Kraay (1999) finds that firms grow more productive with continued participation in export markets. Others, such as Castellani (2002) and Girma et al (2004) find that export intensity is also related to larger increases in firm-level productivity. In this section, I examine how productivity changes are related to continued participation (and discontinuation of that participation) and the intensity of that participation in export markets.

First, I examine within the same regression, how entry into, continued stay in and exit from export markets impacts firm-level productivity. I modify Equation (2) as follows:

$$\ln TFP_{it} = \alpha + \beta_1 start_{it} + \beta_2 stop_{it} + \beta_3 continue_{it} + \gamma Controls_{it} + \sum_i \delta_i Firm_i + \sum_t \delta_t Time_t + \varepsilon_{it} \quad (8)$$

In the original Equation (2), log TFP of the firm was regressed on a dummy variable that equalled 1 if the firm engaged in exports for that given year and 0 otherwise. In the modified version in Equation (8), log TFP of the firm is regressed on 3 different variants of the export participation of a firm - a firm could be present in export markets in a given year because it began to export in that year, or because it continues to export in that given year, or it could have exported in the previous year but discontinued participation in the given year. Thus, in Equation (8), $start_{it}$ is a dummy variable that equals 1 when the firm first enters the export market and zero otherwise, $stop_{it}$ is a dummy variable that equals 1 if the firm exits the export market and zero otherwise, $continue_{it}$ is a dummy variable that equals 1 if a firm is already present within the export market and continues to export and zero otherwise (i.e. it equals zero if the firm enters the market or exists the market). Since a firm could 'continue' to export for more than one year, this variable is interacted with the number of years of exporting, for 5 years and 10 years respectively. Firm-specific controls include the size (number of employees) and the age of the firm. In contrast to other papers that study between-firm effects, since I am mainly interested in within-firm changes, I include firm fixed-effects.

Table 1.5: Exporting and the Intensive Margin

Variables	**Productivity** *(1)*	**Productivity** *(2)*
Start	0.0967***	
	[0.007]	
Continue1	0.0151***	
(5 years)	[0.002]	
Continue2	0.0098***	
(10 years)	[0.001]	
Stop	-0.1093***	
	[0.001]	
Log (Exports)		0.0472***
		[0.006]
Age	-0.0195***	-0.0242***
	[0.001]	[0.000]
Size	0.0000	0.0000

	[0.000]	[0.000]
Firm & Year FE	*Yes*	*Yes*
#	35,379	25,378
R^2	0.836	0.922

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in square brackets (clustered at the firm level)

According to the results (see Table 1.5), the year of entry into the export market is associated with an increase in productivity for the firm. In other words, starting to export is associated with a 9.67 per cent increase in productivity. However, the positive effect of continuing to export grows smaller over time, although it remains statistically significant. The variable 'continue' is interacted with the number of years of exporting in the short (5 years) and in the medium term (10 years)[20]. Continuing to export for 5 years is associated with an average increase in productivity of 1.51 per cent, and 0.98 per cent for 10 years of continued participation. In other words, on average, a firm that exports continuously for 10 years is 0.98 per cent more productive than when it first started to export.

What is interesting is that exiting the export market has a negative effect on firm productivity. Exiting the market is associated with a 10.93 per cent fall in productivity compared to the previous year. It should be kept in mind that although the OP method controls for endogenous exit from the sample, it does not control for endogenous exit from export markets. And so the negative coefficient on 'stop' could be an overestimate since I am unable to observe firms that exit both, the export market and the sample, at the same instant.

In the second exercise, following Castellani (2002) and Girma et al (2004) I regress firms' TFP on exports as a proportion of sales, controlling for the age and size of the firm and with the inclusion of firm fixed-effects (see Column 2 of Table 1.5). I find that the value of total exports has a positive and significant effect on firms' productivity – implying that the more that a firm exports, the higher the productivity of the firm. In this case, a percentage point increase in export volume leads to a 4.72 percent in productivity for my sample of firms.

[20] The possible range is (1,20), with firms exporting for an average of 4-5 years.

1.5.2 Spillovers

The sample also provides me with information on whether firms belong to particular ownership groups. These are large business groups that own a number of firms, usually operating in similar industries, along the vertical or horizontal scale of production. There are a total of 583 business groups where the mean number of firms within each group is 6.5, the minimum 2 and the maximum 127[21]. As an example, take the 'Rane Group', which has 12 firms that produce steering systems, engine valves, brake linings etc. In any given year, anywhere between 1 to 5 firms within the Rane Group export.

One might expect knowledge spillovers to be high for firms within the same business group since there would be few or no restrictions on technology sharing. Technological spillovers are one of the channels through which firms with access to foreign markets became more productive. It could be theorised that non-exporting firms might be able to access better technologies or production processes or designs if other firms within the same business group had access to foreign markets through exporting.

I regress productivity of non-exporters in a given business group on the total number of exporting firms within the same business group in that year, controlling for the location, age and size (number of employees) of firms and again I include firm and year fixed-effects since I mainly interested in the effect for a given firm. I also carry out this exercise at the 2-digit NIC level to study how spillovers might matter between firms in industries with higher technological relatedness.

[21] The business group with the most firms is the Tata Group.

Table 1.6: Business Groups

NIC	NIC Sector Name	Exporters (sum)	L.Exporters (sum)	Observations	R^2
15-36	All	0.0055		2,103	0.770
			0.0081	1,879	0.792
15	Manufacture of food products and beverages	-0.0036		492	0.658
			-0.0147	448	0.715
17	Manufacture of textiles	-0.0567		146	0.728
			-0.0341	124	0.730
21	Manufacture of paper and paper products	-0.0278		81	0.746
			-0.0104	73	0.742
22	Publishing, printing and reproduction of recorded media	-0.0475		26	0.977
			-0.0069	22	0.952
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.0755		30	0.786
			0.0627	28	0.906
24	Manufacture of chemicals and chemical products	0.0271		348	0.688
			0.0190	304	0.653
25	Manufacture of rubber and plastic products	-0.0592***		93	0.847
			-0.0778***	86	0.855
26	Manufacture of other non-metallic mineral products	-0.0095		188	0.533
			-0.0104	175	0.559
27	Manufacture of basic metals	0.0053		144	0.643
			0.0085	128	0.664
28	Manufacture of fabricated metal products, except machinery and equipments	0.1984***		29	0.995
			0.1680***	25	0.998
29	Manufacture of machinery and equipment N.E.C.	-0.0235		84	0.852
			-0.0007	74	0.862
30	Manufacture of office, accounting and computing machinery	0.0854		23	0.872
			0.0109	18	0.975
31	Manufacture of electrical machinery and apparatus N.E.C.	0.0083		108	0.732
			0.0217*	96	0.867
32	Manufacture of radio, television and communication equipment and apparatus	-0.0053		52	0.843
			0.0337	48	0.904
34	Manufacture of motor vehicles, trailers and semi-trailers	0.0669***		116	0.862
			0.0504***	102	0.860
35	Manufacture of other transport equipment	-0.0482		52	0.619
			0.0326	46	0.543

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors clustered at firm level (not shown)

I find that on average (see Table 1.6), the number of exporters within a business group seems to have no statistically significant effect on the productivity of non-exporters within the same business group. I find similar results if I use lagged values. However, when I drill down to particular industry groups I find more interesting results. For manufacturers of rubber and plastic products (NIC 25), the productivity of a given non-exporter falls by 5.92 percentage points if an additional firm within the same business group exports, suggesting that firms within the same business group grow at the expense of others. On the other hand, the effect of additional exporters on the productivity of non-exporters in a given business group is positive for motor manufacturers and for fabricated metal producers – the productivity of a given non-exporter increases by a whopping 6.69 and 19.84 percent respectively if an additional firm within the same business group exports.

1.6 Concluding remarks

This paper analyses the effect of exporting on firm-level productivity over a period that saw a large increase in the number of exporting firms. Descriptive statistics find that exporting is positively associated with size, capital intensity and value addition, within and across firms. Ultimately, however, it is the causal effect of exporting on a given firm's performance that is of interest. In the paper I identify the effect of entering export markets after controlling for the self-selection of more productive firms into such markets. Since I care mostly about the within-firm effect of exporting, wherever possible, I use firm-fixed effects. This is in stark contrast to the earlier literature that studies across-firm effects.

To identify the within-industry effect of exporting on productivity, I use propensity-matching techniques. I construct a set of 'control firms' and then evaluate the effect of the 'treatment', i.e. exporting. I find that exporting does indeed lead to a positive and significant effect on the productivity of firms that begin to export. I also find that the marginal effect of continuing to export is insignificant, although the cumulative effect of exporting after a few years of exporting is still positive. Since I match firms within a given industry, this methodology allows me to identify the within-industry effect of exporting. Using matching, I control for a Melitz-type effect wherein more productive firms are also more likely to become exporters.

I then move on to identifying the effect of exporting on *a given firm*, i.e. the within-firm productivity premium. I use effectively applied tariffs faced by a given firm in world markets to control for reverse causality. The decision of the firm to enter the export market is instrumented with export-weighted tariffs – the intuition being that a fall in tariffs would reduce the fixed cost of entry for firms at the margin. Again, I find that entry is associated with a large increase in productivity. The large jumps in productivity immediately subsequent to entry leads me to suspect that firms in my sample might anticipate entry into export markets, making productivity-enhancing investments which then allow them to recoup large productivity premiums in the first few years of exporting. There is no evidence of continued learning-by-exporting effects, whether within-industry or within-firms.

I also study the effect of the intensive margin of exporting for firms, for continued participation and for higher intensity of participation. I find that the gains from exporting are the highest in the first few years of entry and then begin to taper off. In other words, after tracking firm performance for up to 5 years, I find that there is no evidence of further productivity benefits, except in those firms that are most exposed to export markets. I also find evidence that productivity gains are reversed when the firm decides to exit the export market. I check for any evidence of spillovers from exporting to non-exporting firms within the same business group, and find that there may be positive externalities for some industries, but that on average there are little or no spillovers.

My results for the within-industry effects are in line with previous findings in the literature. Loecker (2007) uses propensity-score matching techniques to identify the effect of exporting on firms in Slovenia and finds that the annual productivity premium from exporting varies between 8 to 13 per cent. Using similar techniques, this paper finds that average productivity premium for firms that start to export are 11 per cent, but that these premiums are no longer significant over time. Biesenbroeck (2005) incorporates lagged exports within the production function and finds robust evidence for a positive effect of exporting on productivity. He also uses the ethnicity of the firm owner as an instrument for the decision of firms to export, and again finds evidence for a causal and positive impact. Using different techniques, his estimate of the effect of exporting on productivity varies between 25 to 28 per cent for firms in Africa. I use effectively applied tariffs as instruments, but the instrument is applied within-firms and not across-firms, and my estimate of the effect of starting to export is around 24 percent. In other words, I find that the effect of exporting on within-firm productivity is much higher than that for within-industries.

The paper uses different ways to get at the effect of exporting on productivity. It starts with a set of simple descriptive statistics, wherein the productivity of a given firm for a given year is regressed on a dummy variable for whether that firm exports in that year, controlling for the size, age and time-invariant unobservables at the level of the firm. This regression illustrates that exporting is associated with a 26 percent increase in average firm-level productivity. However, this regression does not control for the self-selection of more productive firms into exporting. The paper then uses two different methodologies to control for the endogeneity of firm entry – it matches exporters with like firms and calculates the productivity differential over time, and it uses effectively applied tariffs as an instrument to predict entry into export markets. Using propensity score matching techniques, I find that exporters become 11 percent more productive than similar non-exporters after entry into export markets. Using instrumental variables techniques, I find that entry into export markets is associated with a 24 percent increase in average firm-level productivity.

In the paper, I also try and assess the impact of continuing to export, i.e. the productivity differential beyond the first year of entry into export markets. I do this in two ways at different stages in the paper. First, the propensity score matching exercise allows me to follow exporters over time and compare their performance to the matched non-exporters. And second, in the robustness exercises, productivity of a given firm is regressed on dummy variables for whether that firm is a new entrant, is continuing to export, or has exited the market, controlling for firm characteristics and unobservables[22]. In the first case, I find that that exporters remain 4.46 percent more

[22] It should be kept in mind that the robustness exercises do not control for endogeneity of entry into (or exit from) export markets, and only provide an indication of the association between continuing to export and firm-level productivity.

productive than their matched non-exporter counterparts after the fourth year of exporting. However, since exporters were 11 percent more productive in the year of entry, this would indicate that the productivity differential between exporters and their domestic counterparts seems to be falling over time. In the second exercise, the results reveal that continuing to export for 5 years is associated with a 1.51 percent increase, and for 10 years is associated with a 0.98 percent increase in productivity. However, the year of entry is associated with a much larger, 9.67 percent, increase in productivity. Both these exercises indicate that new entrants into export markets see a marked increase in productivity, but that these increases do not seem to be sustained over time.

A few other papers (Damijan and Kostevc 2006, Damijan et al 2004) also find a similar result – productivity improvements, although present, are far from permanent and tend to dissipate shortly after initial entry. One of the explanations for this occurrence may be a simple utilization of excess capacity caused by the sudden availability of a larger product market. In other words, there is little evidence for any sustained learning-by-exporting and it may be that the initial productivity hike is solely a consequence of a scale effect whereby the firm takes advantage of a larger market to place its additional output. In essence, the hike in productivity only reflects the fact that firms can take advantage of their spare capacity in new markets. One way to check if this might be the case would be if sales and/or capital show a substantial change in the year of entry. To explore this possibility further, I plot the average sales and capital (gross assets) in the years before and after entering export markets (see Figure 1.A.9). I find that the year of entry is associated with a marked increase in sales and capital. Thus, there is some indication that the increase in productivity growth could be attributed to the initial utilisation of excess capacity, although conclusive evidence for this would rely on the availability of capacity utilisation data. As firms proceed to increase their size in order to accommodate the increased sales the productivity hike diminishes quickly in the data. Thus, this would suggest that the observed improvements in productivity in the year of entry into export markets might primarily be a reflection of a growth in inputs.

This study makes a few contributions to an already crowded literature. It is an attempt to understand the productivity premium from exporting for a poor, rapidly developing country. There has been no previous robust methodological work on answering the age-old, yet classic, question of the effect of exporting on firm-level productivity for a large and increasingly important country like India. Developing countries that are growing briskly are also where one would expect the gains to be the highest. Evidence on the determinants and computation of firm-level productivity in low-income countries is also rare. In addition, I use an instrument, i.e. effectively applied tariffs that has previously not been used in the literature to control for the self-selection effect within firms. The use of tariffs nicely isolates the effect of entry into export markets on productivity, and since these tariffs are hardly unique to the case of India, they could easily be applied in other settings.

And lastly and most importantly, to my knowledge, this is a first attempt to identify the effect of participation in export markets that separates the across-firm effect from the within-firm effect. The existing empirical literature focuses almost primarily on controlling for the self-selection of more productive firms within a given industry. In contrast, in this paper I identify the effect of starting to export within firms. Although the question of whether exporting raises aggregate industrial productivity is a tremendously interesting one, it seems perilous to ignore the effects on aggregate firm productivity.

References

Alvarez, R. and Lopez, R.A. (2005), 'Exporting and performance: evidence from Chilean plants', Canadian Journal of Economics, 3(4): 1384-1400.

Angrist, J. and Pischke, J.S. (2008), Mostly Harmless Econometrics, Princeton University Press, New Jersey.

Arnold, J. and Hussinger, K. (2004), 'Export behaviour and firm productivity in German manufacturing: A firm-level analysis', Discussion Paper No. 04-12, Centre for European Economic Research (ZEW).

Aw, BY., Chung, S and Roberts, MJ. (2000) 'Productivity and turnover in the export market: micro-level evidence from the Republic of Korea and Taiwan (China)', World Bank Economic Review, 14(1): 65– 90.

Baldwin, JR. and Gu, W. (2003), 'Export market participation and productivity performance in Canadian manufacturing', Canadian Journal of Economics, 36(3): 634–657.

Baldwin, J.R. and Gu, W. (2004), 'Trade liberalisation: Export-market participation, productivity growth and innovation', Oxford Review of Economic Policy, 20(3): 372–392.

Bernard, A. and Jensen, J.B. (1995), 'Exporters, jobs and wages in US manufacturing: 1976–1987', Brookings papers on economic activity, Microeconomics, 67–119.

Bernard, A. and Jensen, J.B. (1999), 'Exceptional exporter performance: cause, effect or both?' Journal of International Economics, 47: 1-25.

Bernard, A., Eaton, J., Jensen, J.B., and Kortum, S.S. (2003), 'Plants and productivity in international trade', American Economic Review, 93(4): 1268-1290

Bernard, A. and Jensen, J.B. (2004), 'Why some firms export', Review of Economics and Statistics, 86(2): 561–569.

Bernard, A., Jensen, J.B., Redding, S. and Schott, P.K. (2007), 'Firms in international trade', NBER Working Paper No. 13054.

Van Biesebroeck, J. (2005), 'Exporting raises productivity in sub-Saharan African manufacturing firms', Journal of International Economics, 67: 373-391.

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J.W., Oduro, A., Oostendorp, R., Pattillo, C., Soderbom, M., Teal, F. and Zeufack, A. (2004), 'Do African manufacturing firms learn from exporting', Journal of Development Studies, 40 (3): 115– 141.

Blalock, G. and Gertler, P.J. (2004), 'Learning from exporting revisited in less developed setting', Journal of Development Economics, 75(2): 397–416.

Blundell, S. and Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', Journal of Econometrics, 87: 29-52.

Bown, C. and Porto, G. (2010), 'Exporters in developing countries: Adjustment to foreign market access after a trade policy shock', Working Paper (Link)

Card, D. (2001), 'Estimating the return to schooling: progress on some persistent econometric problems', Econometrica, 69(5): 1127–1160.

Castellani, D. (2002), 'Export behaviour and productivity growth: Evidence from Italian manufacturing firms', Weltwirtschaftliches Archiv, 138: 605–628.

Clerides, S., Lach, S. and Tybout, J. (1998), 'Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco', Quarterly Journal of Economics, 113(11): 903-947.

Damijan, J.P., Polanec, S. and Prasnikar, J. (2004), 'Self-selection, export market heterogeneity and productivity improvements: Firm level evidence from Slovenia', LICOS Discussion Paper No. 148, Centre for Transition Economies. Katjolieke Universiteit Leuven.

Damijan, J.P. and Kostevc, C. (2006), 'Learning-by-Exporting: Continuous productivity improvements or capacity utilisation effects? Evidence from Slovenian firms', Review of World Economics, 142 (3): 599-614.

Delgado, M.A., Farinas, J.C. and Ruano, S. (2002), 'Firm productivity and export markets: a non-parametric approach', Journal of International Economics, 57: 397-422.

DiPrete, T.A. and Gangl, M. (2004), 'Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments', Sociological Methodology, 34 (1): 271-310.

Dixon, R.J. and Thirlwall, A.P. (1975), 'A model of regional growth differences on Kaldorian lines', Oxford Economic Papers, 27: 201-214.

Fernandes, A.M. (2007), 'Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries' Journal of international economics, 71: 52-71.

Girma, S., Greenaway, D. and Kneller, R. (2004), 'Does exporting increase productivity? A microeconometric analysis of matched firms', Review of International Economics, 12(5): 855–866.

Goh, A.T. (2000), 'Opportunity cost, trade policies and the efficiency of firms' Journal of Development Economics, 62(2): 363-383.

Goldberg, P, Khandelwal, A., Pavcnik, N. and Topalova, P. (2010), 'Multiproduct firms and product turnover in the developing world: Evidence from India', Review of Economics and Statistics, 92(4): 1042-1049.

Greenaway, D. and Kneller, R. (2008), 'Exporting, productivity and agglomeration', European Economic Review, 52(5): 919-939.

Grossman, G. amd Helpman. E. (1991), *Innovation and Growth in the World Economy*, MIT Press: Cambridge.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', Review of Economic Studies, 64:605–654.

Helpman, E. and Krugman P. (1985), *Market Structure and Foreign Trade*, MIT Press:

Cambridge.

Hung, J., Salomon, M. and Sowerby, S. (2004), 'International trade and US productivity', Research in International Business and Finance, 18: 1-25.

Isgut, A.E. (2001), 'What's different about exporters? Evidence from Colombian manufacturing', Journal of Development Studies, 37 (5): 57– 82.

Jalan, J. and Ravallion, M. (2003), 'Estimating the benefit incidence of an Antipoverty Program by Propensity Score Matching', Journal of Business and Economic Statistics, American Statistical Association 21 (1): 19-30.

Kaldor, N. (1970), 'The case for regional policies', Scottish Journal of Political Economy, 17: 337-348.

Kim, E. (2000), 'Trade liberalization and productivity growth in Korean manufacturing industries: price protection, market power, and scale efficiency', Journal of Development Economics, 62(1): 55-83.

Kraay, A. (1999), 'Exports and economic performance: Evidence from a panel of Chinese enterprises', Revue d'Economie du Developpement,1(2): 183-207.

Levinsohn, L. and Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', Review of Economic Studies, 70(2): 317-341.

Loecker, J.D. (2007), 'Do exports generate higher productivity? Evidence from Slovenia', Journal of International Economics, 73: 69-98.

Loecker, J.D. (2010), 'A note on detecting learning by exporting', NBER Working Paper 16546.

Melitz, M. (2003), 'The impact of trade on intra-industry reallocations and aggregate industry productivity', Econometrica, 71 (6): 1695-1725.

Olley, G.S. and Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', Econometrica, 64 (6): 1263-1297.

Park, A., Yang, D., Shi, X. and Jiang, Y. (2010) 'Exporting and firm performance: Chinese exporters and the Asian financial crises', Review of Economics and Statistics, 92(4): 822-842.

Pavcnik, N. (2002), 'Trade liberalisation, exit and productivity improvements: Evidence from Chilean plants', The Review of Economic Studies, 61(1): 245-276.

Ravallion, M. (2008), Evaluating Anti-Poverty Programs, Chapter 59, Handbook of Development Economics, T. Paul Schultz and A. Strauss (ed.), Elsevier.

Rivera-Batiz, L. and Romer, P. (1991), Economic integration and endogenous growth, Quarterly Journal of Economics, 106: 531-555.

Roberts, M. and Tybout, J. (1997), 'The decision to export in Colombia: An empirical model of entry with sunk costs', American Economic Review, 87(4): 545–564.

Srinivasan, T.N. And Archana, V. (2009), 'India in the global and regional trade: determinants of aggregate and bilateral trade flows and firms' decision to export', Working Paper No. 232, Indian Council for Research on International Economic Relations (ICRIER).

Tabrizy, S.S. and Trofimenko, N. (2010), 'Scope for export-led growth in a large emerging economy: Is India learning by exporting?", Kiel Working Paper No. 1633, Kiel, Germany.

Tybout, J.R. and Westbrook, M.D. (1994), 'Trade liberalisation and the dimensions of efficiency change in Mexican manufacturing industries', Journal of International Economics, 39: 53-78.

Wagner, J. (2002), 'The causal effects of exports on firm size and productivity: First evidence from a matching approach', Economics Letters, 77: 287–292.

World Bank (2008), 'Global Economic Prospects: Technology diffusion in the developing world', Washington DC.

Yasar, M., Raciborski, R. and Poi, B. (2008), 'Production function estimation in Stata using the Olley and Pakes method', The Stata Journal, 8(2): 221-231.

Yeaple, S.R. (2005) 'A simple model of firm heterogeneity, international trade, and wages', Journal of International Economics, 65: 1-20.

Appendix 1.A

1.A.1 Estimating Productivity

Following Olley and Pakes (1996) it is assumed that in year t the manager observes the firm's current productivity ω_{it} before choosing labour l_t and investment i_t to combine with the quasi-fixed input, capital k_t for the production of output y_t. Output is expressed as follows:

$$y_t = \beta_0 + \beta_l l_t + \beta_i i_t + \beta_k k_t + \omega_t + \varepsilon_t \tag{A1}$$

Inputs are divided into a freely variable ones (l_t,i_t) and the state variable capital (k_t). The error term is assumed to be additively separable in a transmitted component (ω_t) and an i.i.d. component (ε_t). The key difference between the former and the latter is that the former is a state variable and hence impacts the firm's decision rules, while the latter has no impact on the firm's decisions.

Since ω_t is known to the manager but unknown to the econometrician and may be positively correlated with l_t and i_t, it generates a potential simultaneity bias that is addressed by the following estimation procedure. The firm's variable input demands, derived from profit maximisation, depend on privately known productivity and capital. Investment's demand function is given by:

$$i_t = i(\omega_t, k_t)$$

and it must be monotonic in all ω_t for all relevant k_t to qualify as a valid proxy – implying that conditional on capital the demand for investment increases with productivity. Assuming that monotonicity holds, the input demand function can be inverted to obtain ω_t as a function of investment and capital, as below. Note that this function depends on observables only.

$$\omega_t = \omega(i_t, k_t)$$

Consider the problem of self-selection. Firms with larger capital stocks can expect higher returns on capital even in the face of lower levels of productivity, and will choose to stay longer in the market. Thus the self-selection generated by the exit behaviour implies that the expectation of productivity will be decreasing in capital, leading to a negative bias in the capital coefficient.

The first stage of the estimation proceeds by rewriting Equation (A1) in a partially linear form:

$$y_t = \beta_l l_t + \phi(i_t, k_t) + \varepsilon_t \tag{A2}$$

where,

$$\phi(i_t, k_t) = \beta_0 + \beta_i i_t + \beta_k k_t + \omega(i_t, k_t) \tag{A3}$$

Since $E[\varepsilon_t \mid i_t, k_t] = 0$, taking the difference between Equation (A2) and its expectation conditional on investment and capital generates the following expression:

$$y_t - E[y_t \mid i_t, k_t] = \beta_l(l_t - E[l_t \mid i_t, k_t] + \varepsilon_t \qquad \text{(A4)}$$

Equation (A4) is estimated by OLS (no constant) to obtain consistent parameter estimates for labour. The conditional expectations in Equation (A4) are the intercepts of locally weighted least squares (LWLS) regressions of output and labour on (i_t, k_t). After obtaining estimates for β_l, we estimate the function $\phi(.)$ as a LWLS regression of $y_t - \hat{\beta}_l l_t$ on (i_t, k_t). If one were only concerned with the marginal productivities of the variable inputs (but not the co-efficient on the proxy variable) one could stop here. To obtain a capital co-efficient, a plant-level measure of productivity a more complete model for $\phi_t(.)$ will be required since capital enters it twice. To estimate β_k, in addition to the estimates of β_l obtained from the partially linear model, estimates of the survival probabilities are also used. These probabilities are given by:

$$\Pr\{\chi_{t+1} = 1 \mid \underline{\omega}_{t+1}(k_{t+1})J_t\} \equiv P_t \qquad \text{(A5)}$$

where χ_t is defined as the indicator function and is equal to zero if the firm exits the market, and J_t refers to the information available at time t. In the implementation the probability of survival is estimated by fitting a probit model of χ_{t+1} on the state and proxy variables, as well as their squares and cross products.

In the next stage the expectation of $y_{t+1} - \beta_l l_{t+1}$ conditional on information at time t and survival is given as:

$$y_{t+1} - \beta_l l_{t+1} = \beta_k k_{t+1} + g(P_t, \phi_t - \beta_k k_k) + \xi_{t+1} + \varepsilon_{t+1} \qquad \text{(A6)}$$

where, $\xi_{t+1} = \omega_{t+1} - E[\omega_{t+1} \mid \omega_t, \chi_{t+1} = 1]$, and is the unexpected productivity shock and is independent and identically distributed (i.i.d.). The unknown function $g(.)$ is approximated by a second-order polynomial in $\phi_t - \beta_k k_k$ and P_t[23]. The estimates of β_l, ϕ_t and P_t are substituted in (A6) for the true values of β_l, ϕ_t and P_t, to then obtain estimates of β_k by minimising the sum of squared residuals in equation (A6). Since the estimation routine involves three steps, the stat command implemented uses the clustered bootstrap errors, treating all observations for a single firm as one cluster.

Following Loecker (2007) with the coefficients of the production function in hand, I then recover a productivity measure for the firm i in industry j at time t[24]:

$$\omega_{ijt} = y_{ijt} - \beta_{lj} l_{ijt} - \beta_{kj} k_{ijt}$$

[23] Readers interested in the details of the estimation system can refer to Yasar et al (2008).

[24] It should be noted here that this measure of productivity if not the true unobserved productivity shock. It also includes the i.i.d. component which is assumed to be zero on average.

Tables

Table 1.A.1: Summary of empirical findings on exports and productivity

Study	Country	Sample	Methodology	Evidence*	
				$\omega^{e} > \omega^{ne}$	LBE effects
Aw and Hwang (1995)	Taiwan	2,832 firms; 1986	Translog production function, cross-section	√	x
Bernard and Wagner (1997)	Germany	7,624 firms; 1978-92	Panel Data	√	√
Clerides et al (1998)	Colombia, Mexico, Morocco	All firms; 1981-91, 1986-90, 1984-91	FIML of cost functions; Panel data	√	√[1]
Kraay (1999)	China	2,105 firms; 1988-92	Dynamic panel	√	√
Bernard and Jensen (1999)	US	60,000 plants; 1984-92	Linear probability with fixed effects	√	x
Kim (2000)	Korea	36 sectors; 1966-1988	Translog production function; cross-sections	√	x
Isgut (2001)	Colombia	6453 plants; 1981-1991	Difference-in-Differences methodology (with dummies)	√	x
Delgado et al (2002)	Spain	1,766 firms; 1991-96	Nonparametric analysis of productivity distributions	√	x
Castellni (2002)	Italy	2,898 firms; 1989-94	Cross-section	√	√[2]
Wagner (2002)	Germany	353 firms; 1978-89	Panel data; Matching	√	x
Alvarez and Lopez (2005)	Chile	5,000 plants; 1990-96	Ordered probit; pooled data	√	√
Baldwin and Gu (2003)	Canada	8215 firms; 1974-1996;	System GMM; Cross-sections	√	√
Arnold and Hussinger (2004)	Germany	389 firms; 1992-2000	Olley and Pakes production function; Matching techniques	√	x
Bigsten et al (2004)	Cameroon, Ghana, Zimbabwe	289 firms; 1992-1995	Maximum likelihood, System GMM methods; Panel data	√	√
Girma et al (2004)	UK	8,992 firms; 1988-1999	Matched samples	√	√

Hung et al (2004)	US	40 industries; 1996-2001	Difference-in-Differences Methodology; Panel data	√	x
Blalock and Gertler (2004)	Indonesia	20,000 firms; 1990-1996	Translog, Olley and Pakes, Levinsohn and Petrin production function;	√	√
Van Bisebroeck (2005)	Sub-Saharan Africa	1916 firms (9 countries); 1992-1996	GMM, Maximum likelihood, Olley and Pakes production function; Cross-section	√	√
Loecker (2007)	Slovenia	7915 firms; 1994-2000	Olley and Pakes production function; Matched samples	√	√
Fernandes (2007)	Colombia	6474 plants; 1977-1991	Levinsohn and Petrin for production function; Panel data	√	√
Greenaway and Kneller (2008)	UK	11,225 firms; 1988-2002	Matched samples	√	√
Park et al (2010)	China	3,339 firms; 1995, 1998, 2000	Instrumental Variables Technique	√	√

* $\omega^{e} > \omega^{ne}$ **:** Exporters more productive than non-exporters; LBE: Learning-by-exporting effects. 1: Some learning from exporting in the case of Morocco. 2: Learning associated with export intensity. Source: Girma et al (2004), modified and updated.

Table 1.A.2: Capital and Labour Coefficients (by Industry)

NIC	NIC Sector Name	Olley-Pakes			OLS		
		Capital	*Labour*	#	*Capital*	*Labour*	#
15	Manufacture of food products and beverages	1.0256***	0.2055***	7,564	0.7470***	0.3182***	8,831
16	Manufacture of tobacco products	0.7768**	0.2802***	154	0.7828***	0.3638***	195
17	Manufacture of textiles	0.7123***	0.3469***	6,398	0.6441***	0.3858***	7,883
18	Manufacture of wearing apparel; dressing and dyeing of fur	0.7421***	0.3848***	851	0.6738***	0.4743***	1,027
19	Tanning and dressing of leather; manufacture of luggage, handbags and footwear	0.1178	0.4476***	550	0.4740***	0.5791***	687
20	Manufacture of wood and of products of wood and cork, except furniture	0.5592***	0.4966***	293	0.4338***	0.7010***	360
21	Manufacture of paper and paper products	0.3629**	0.4996***	1,716	0.3609***	0.5891***	1,968
22	Publishing, printing and reproduction of recorded media	0.6958***	0.3982***	548	0.5495***	0.4679***	692
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.6992***	0.4302***	619	0.5686***	0.5638***	735
24	Manufacture of chemicals and chemical products	0.7428***	0.4490***	12,143	0.5286***	0.5308***	14,326
25	Manufacture of rubber and plastic products	0.8023***	0.4534***	3,619	0.5484***	0.5463***	4,384
26	Manufacture of other non-metallic mineral products	0.5144***	0.4920***	2,499	0.5274***	0.5773***	2,983
27	Manufacture of basic metals	0.8639***	0.2580***	5,738	0.5827***	0.3569***	6,773
28	Manufacture of fabricated metal products, except machinery and equipments	1.0148***	0.2689***	1,644	0.6635***	0.3629***	1,968
29	Manufacture of machinery and equipment N.E.C.	1.0025***	0.3401***	4,072	0.6799***	0.3958***	5,013
30	Manufacture of office, accounting and computing machinery	0.9153***	0.6034***	477	0.3873***	0.6965***	591
31	Manufacture of electrical machinery and apparatus N.E.C.	1.1370***	0.3449***	2,238	0.6669***	0.4295***	2,694
32	Manufacture of radio, television and communication equipment and apparatus	0.9436***	0.6574***	1,489	0.4887***	0.6974***	1,863
33	Manufacture of medical, precision and optical instruments, watches and clocks	1.0387***	0.4980***	708	0.5121***	0.6117***	856
34	Manufacture of motor vehicles, trailers and semi-trailers	0.7570***	0.5300***	3,262	0.4936***	0.5983***	3,655
35	Manufacture of other transport equipment	1.0306***	0.4041***	557	0.5877***	0.4551***	660
36	Manufacture of furniture; manufacturing N.E.C.	1.2380***	0.1309***	888	1.0892***	0.1825***	1,142

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Notes: The Table reports production function estimates using Olley-Pakes and simple OLS methodologies, by each 2-digit NIC industry. With constant returns to scale the sum of the coefficients should equal 1, and if higher, this implies increasing returns to scale for the given industry.

Table 1.A.3: Exporting and TFP (by Industry)

NIC	NIC Sector Name	β	Errors	#	R^2
15	Manufacture of food products and beverages	0.2686***	[0.036]	4,955	0.708
16	Manufacture of tobacco products	-1.1438***	[0.248]	126	0.789
17	Manufacture of textiles	0.2609***	[0.030]	4,757	0.618
18	Manufacture of wearing apparel; dressing and dyeing of fur	0.6911***	[0.128]	483	0.564
19	Tanning and dressing of leather; manufacture of luggage, handbags and footwear	1.0234***	[0.167]	347	0.701
20	Manufacture of wood and of products of wood and cork, except furniture	0.2577***	[0.093]	228	0.749
21	Manufacture of paper and paper products	0.3966***	[0.058]	1,181	0.560
22	Publishing, printing and reproduction of recorded media	0.1837**	[0.072]	373	0.766
23	Manufacture of coke, refined petroleum products and nuclear fuel	-0.0323	[0.092]	479	0.619
24	Manufacture of chemicals and chemical products	0.3354***	[0.022]	8,945	0.618
25	Manufacture of rubber and plastic products	0.3084***	[0.040]	2,681	0.635
26	Manufacture of other non-metallic mineral products	0.1969***	[0.043]	2,016	0.557
27	Manufacture of basic metals	0.3733***	[0.040]	3,918	0.628
28	Manufacture of fabricated metal products, except machinery and equipments	0.4486***	[0.062]	1,060	0.705
29	Manufacture of machinery and equipment N.E.C.	0.0813***	[0.028]	3,293	0.716
30	Manufacture of office, accounting and computing machinery	-0.1037	[0.152]	320	0.670
31	Manufacture of electrical machinery and apparatus N.E.C.	-0.0480	[0.045]	1,713	0.731
32	Manufacture of radio, television and communication equipment and apparatus	0.1458**	[0.068]	1,165	0.716
33	Manufacture of medical, precision and optical instruments, watches and clocks	0.1107	[0.094]	469	0.800
34	Manufacture of motor vehicles, trailers and semi-trailers	0.0775***	[0.029]	2,326	0.707
35	Manufacture of other transport equipment	0.1869	[0.116]	475	0.812
36	Manufacture of furniture; manufacturing N.E.C.	0.1914	[0.121]	610	0.568

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Notes: The Table reports estimates of Equation (2) with the sample restricted to a given 2-digit NIC industry. Each row reports the coefficient, standard errors clustered at the firm level and the R^2 from one regression as well as the number of observations.

Table 1.A.4: Balancing Tests

		Mean				t-test	
Variable	**Sample**	*Treated*	*Control*	% bias	% bias ↓	t	p>|t|
Lag (Sales)	Unmatched	5.2063	5.6853	-26.6		-10.86	0.000
	Matched	5.2063	5.2287	-1.2	95.3	-0.46	0.687
Lag (Productivity)	Unmatched	-0.1919	-0.3138	9.0		3.88	0.000
	Matched	-0.1919	-0.0937	-7.2	19.3	-2.27	0.023

Table 1.A.5: First-Stage Results

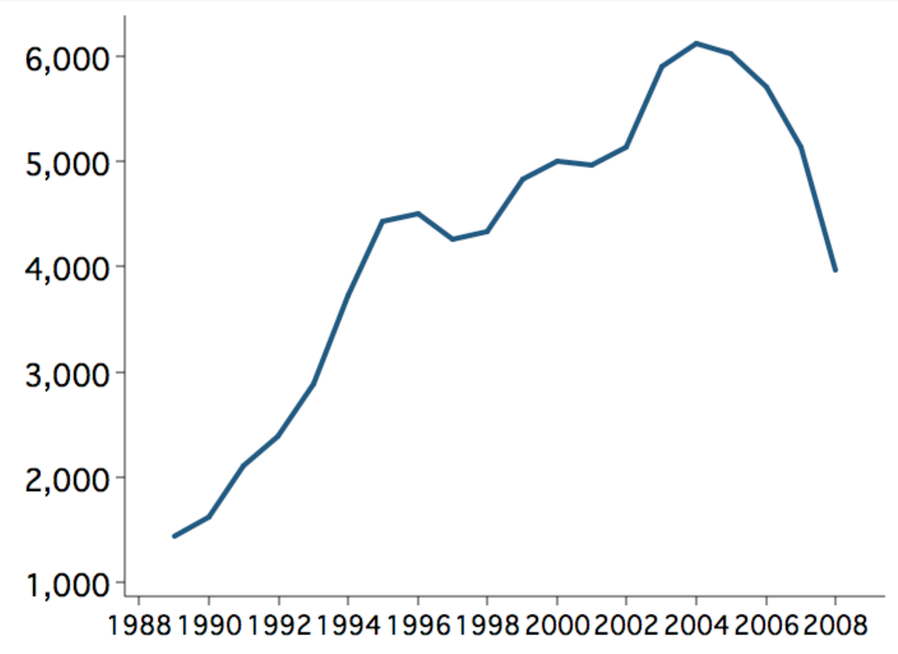
Predictors	**Start**	
Δ Weighted tariffs	-0.0009**	
	[0.000]	
Weighted tariffs (levels)		-0.0007*
		[0.000]
Age	-0.0052***	-0.0056***
	[0.001]	[0.001]
Size	0.0000	0.0000
	[0.000]	[0.000]
Firm & Year FE	*Yes*	*Yes*
#	16,442	15,899
R^2	0.496	0.499
Partial R^2	0.001	0.001
F-Statistic	10.17	5.61

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in square brackets (clustered at the firm level)

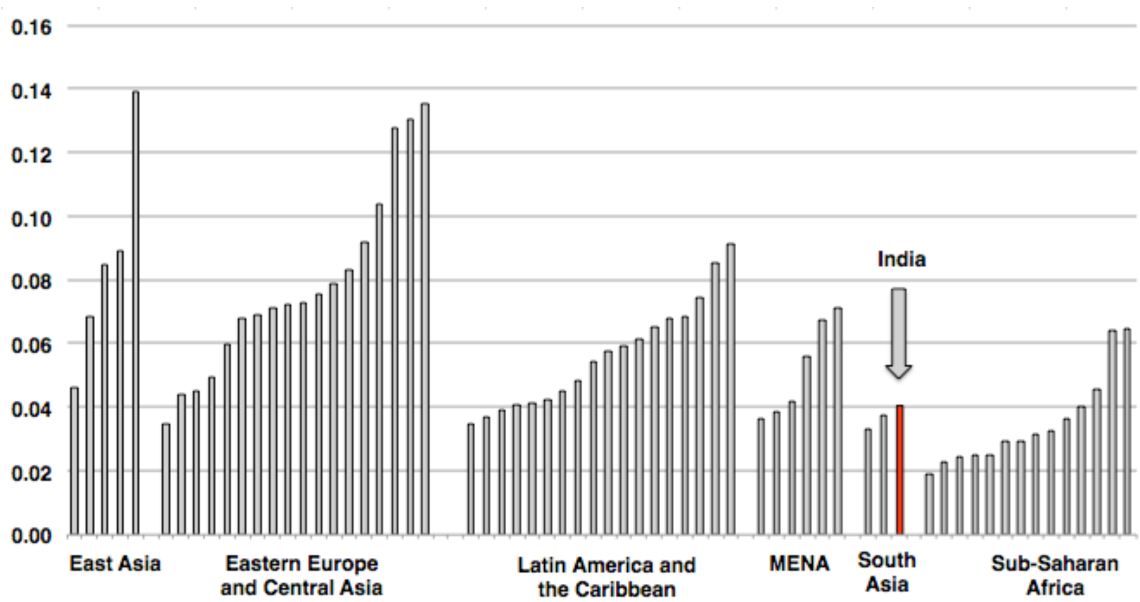
Figures

Figure 1.A.1: Number of exporters



Source: Based on Prowess, CMIE (Centre for Monitoring of the Indian Economy)

Figure 1.A.2: Technology Achievement Index



Source: Based on data from Global Economic Prospects (World Bank 2008)

Figure 1.A.3: Identification of the true impact

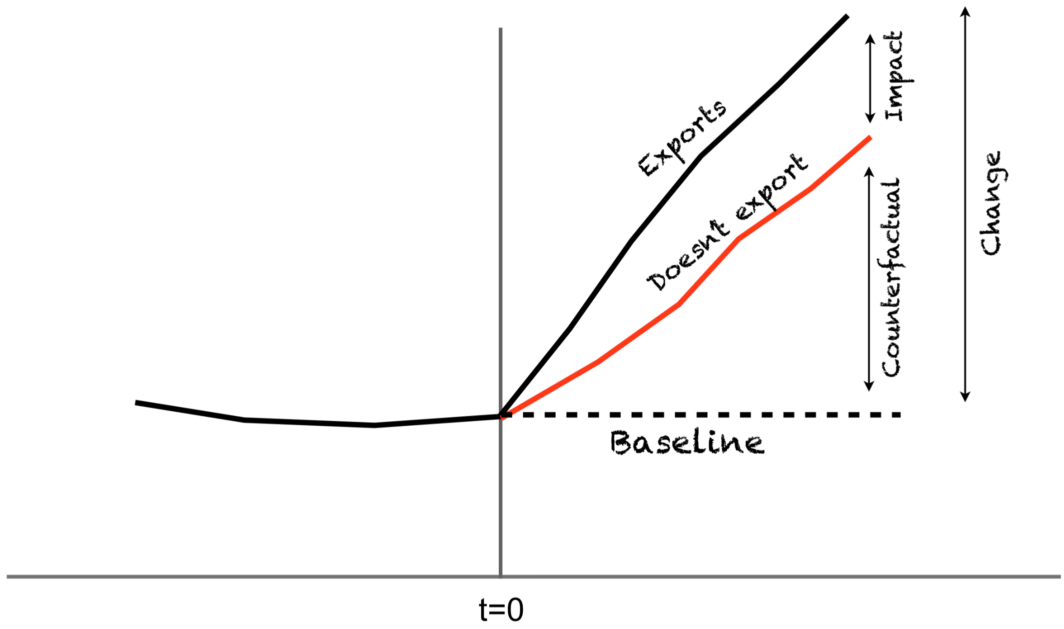


Figure 1.A.4: Region of Common Support (Unmatched and Matched samples)

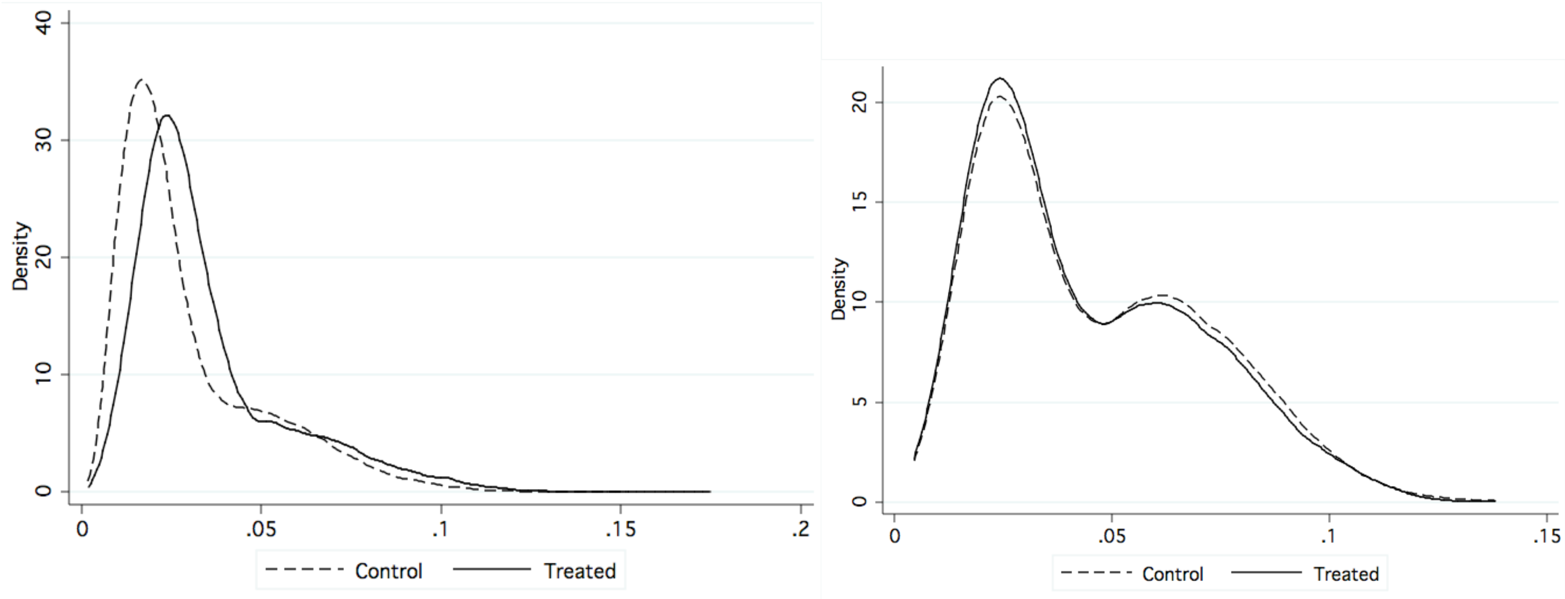
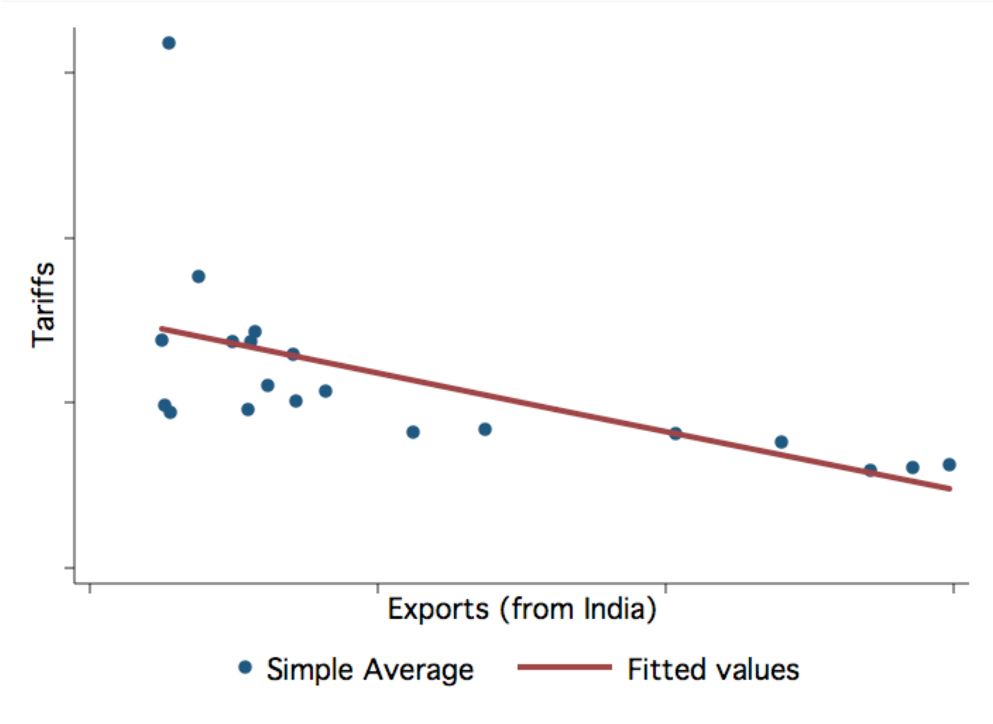
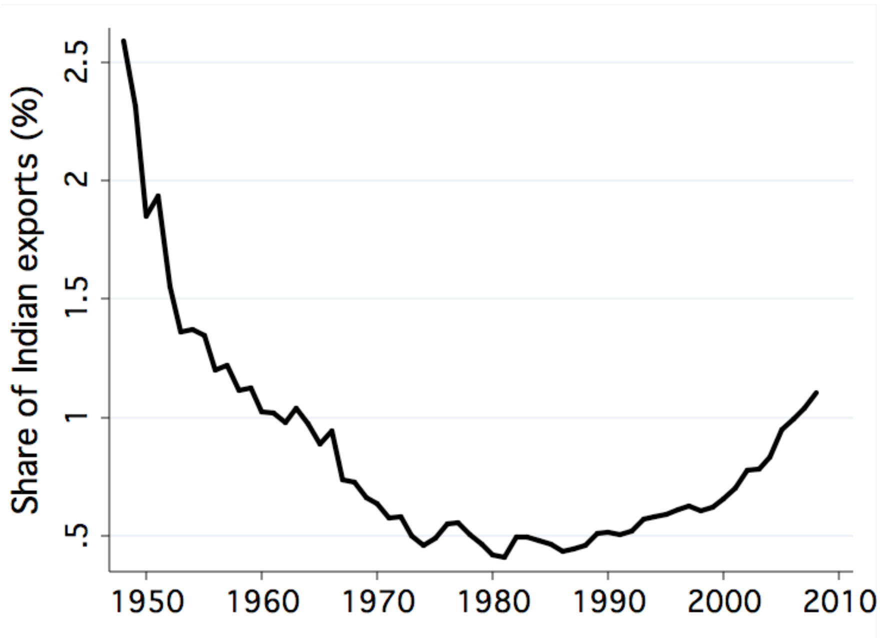


Figure 1.A.5: Tariffs and Exports



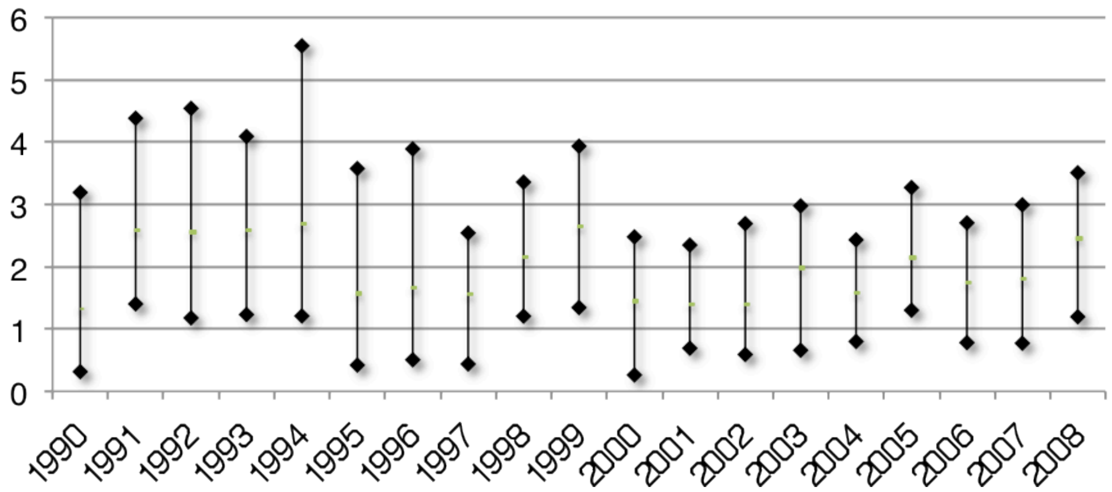
Source: World Integrated Trade Solution (WITS) and Trains

Figure 1.A.6: Indian Export share



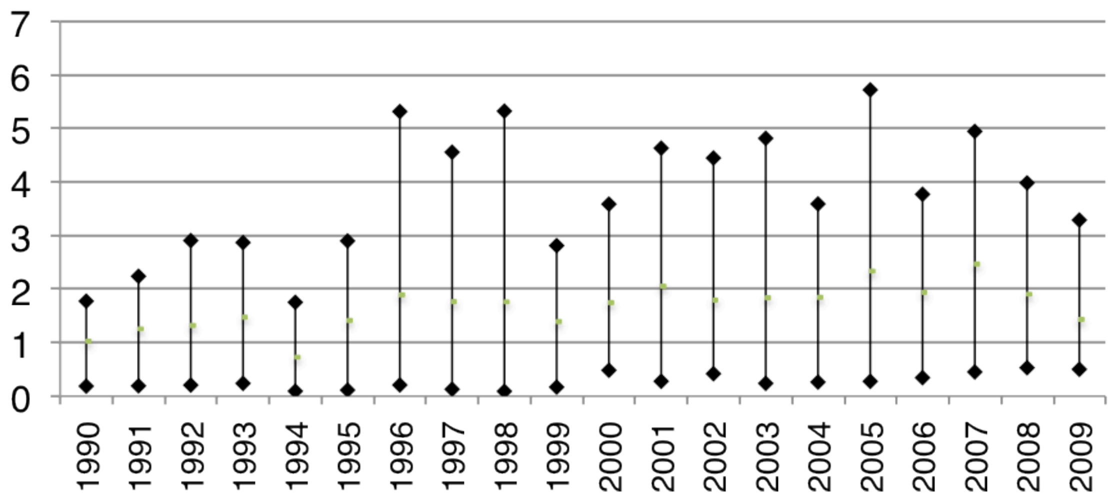
Source: World Integrated Trade Solution (WITS) and Comtrade

Figure 1.A.7: Range of export shares by industries



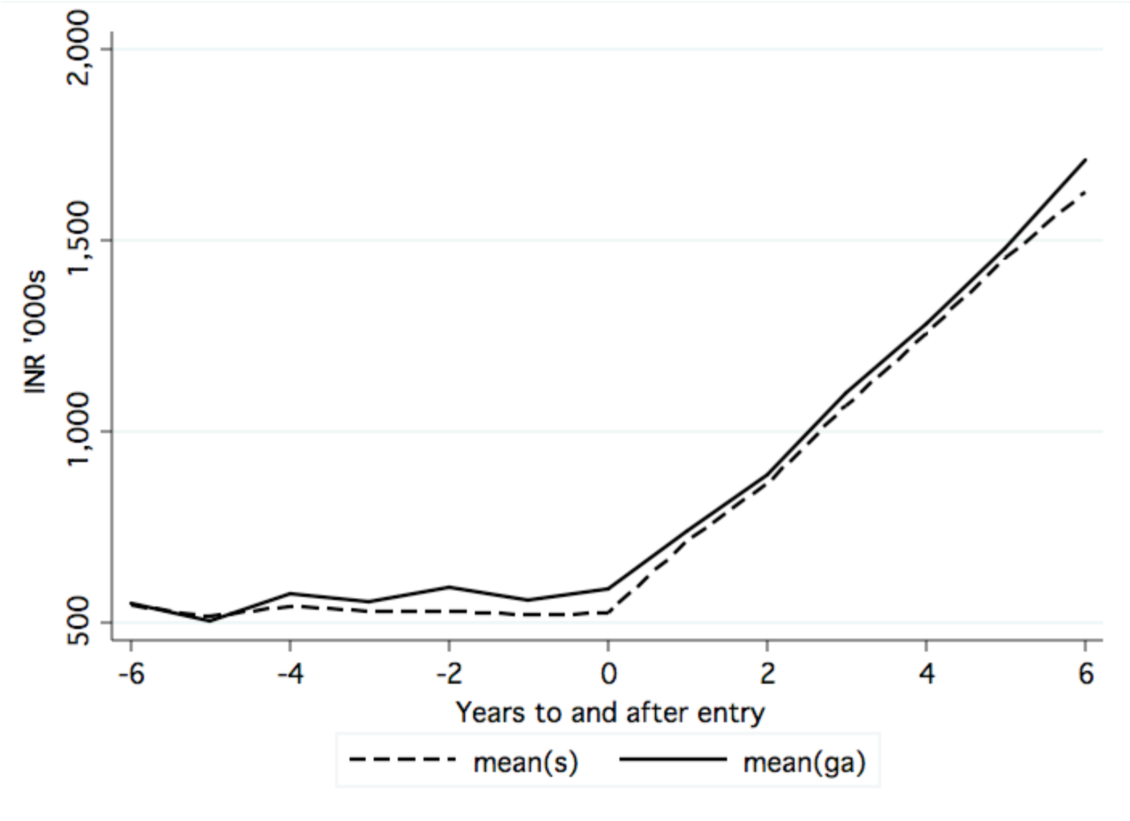
Source: World Integrated Trade Solution (WITS) and Comtrade

Figure 1.A.8: Range of export shares by top 20 export markets and industries



Source: World Integrated Trade Solution (WITS) and Comtrade

Figure 1.A.9: Sales, Capital and Entry into Export Markets



Notes: S = Sales, GA=Gross Assets (Capital)

CHAPTER 2

Industry and the Urge to Cluster: A Study of the Informal Sector in India[*]

[*] I am grateful to Diana Weinhold, Holger Görg, Henry Overman, 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.

Abstract

This paper studies the determinants of firm location choice at the district-level in India to gauge the relative importance of agglomeration economies vis-à-vis good business environment. A peculiar characteristic of the Indian economy is that within the non-farm sector, unorganised enterprises account for 43.2% of NDP and employ 71.6% of the workforce. I analyse National Sample Survey data that covers over 4.4 million enterprises in the unorganised sector, in both manufacturing and services industries. The empirical analysis is carried out using count models. Since the unorganised sector has little access to formal credit facilities and remains untouched by changes in regulations, in line with intuition I find that intra-industry clustering and linkages to buyers and suppliers are of paramount importance. In short, the general business environment seems to be largely irrelevant for the location decisions of unorganised sector establishments, whilst agglomeration economies seem to matter more. I conclude that public policy may be limited in its ability to encourage relocation of informal firms.

Keywords: Agglomeration, Informal Sector, Location Choice, India
JEL Classification: R1, R3, O1

2.1 Introduction

The unorganised, or informal sector[25] 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 often falls outside the scope for planned development efforts, and thus remains in the shadows with regard to productivity, social security and statistics.

This paper is a first attempt to understand the forces that drive the clustering of informal sector activities in India. I study how new firms within the Indian unorganised sector choose to locate themselves across districts[26] in the country. Using count models I carry out an empirical test of the decisions of individual firms. In the model, firms compare potential profitability as a function of observable location specific advantages, market access, agglomeration economies and a set of unobserved local attributes of the district. And so, to unpack the location decisions of unorganised sector firms, an econometric analysis of location patterns is carried out to identify the 'revealed preferences' of firms. Firm-level data for the unorganised sector is taken from surveys conducted by the National Sample Survey Organisation (NSSO), which includes information on the number and type of new firms within each district.

It is important to test whether individual firm's decisions are based on agglomeration economies, or on other factors, such as good business environment – the latter being more amenable to change by policy than the former. In theory, if government is interested in encouraging industrial growth in particular regions, it should have a clear understanding of what factors drive firm location decisions. Since earlier papers have analysed the case of location decisions made by the Indian formal sector (albeit only for manufacturing firms - see Lall et al 2004, Lall and Chakravorty 2005), this paper does not reinvent the wheel and concentrates mainly on an analysis of the informal sector. It does provide an overview of how the results differ between the formal and the informal sector as a sub-section in the robustness analysis, but the point of the paper is not to

[25] A number of countries, including India, often use the terms 'unorganised sector' and 'informal sector' interchangeably.

[26] India is a federal union of 28 states and 7 union territories, which are further sub-divided into 604 districts.

make a comparative study. Since there is little or no research that sheds light on what factors attract smaller, unorganised sector firms to a location, this paper is, first and foremost, an attempt to fill in this gap in the knowledge. In particular, one might expect informal firms to behave differently from those in the formal sector – they may be less encumbered by regulatory structures, they could be less sensitive to wage levels since they mostly rely on own or family-labour, and they could me more mobile, across locations and industries. The importance of networks of social interaction with their intermediate goods' suppliers or their final goods' buyers within a location might outweigh the importance of infrastructure or institutional factors. These links might be with other establishments in the informal sector, or with those in the formal sector, and the paper will make an attempt to disentangle the relationship between the two sectors, formal and informal, and across industries both manufacturing and services. Since the informal sector in India is a significant source of employment (32%) and economic growth (22.6%) (National Account Statistics 2005), there remains a yawning gap in the empirical understanding of how a large proportion of the country's economic sector makes location decisions.

While the results of the analysis provide an understanding of what drives clustering in informal industries in India, they also add to a rapidly growing body of empirical evidence that tests the theoretical implications of Krugman's economic geography. This paper finds that agglomeration economies have a significant effect on firms' location decisions, and that the ability of incremental policy reforms to counter the effects of geography may be limited. Indeed, clustering of firms within the same industry, and clustering close to buyers or suppliers has very different implications for the formal versus the informal sector. At the same time, informal firms seem unaffected by labour costs, while the same may not be true of formal firms. A similar result is obtained for different types of infrastructure. If public policy finds itself limited in its ability to affect industrial diversity or the clustering of buyers and suppliers, then in the case of the unorganised sector, geography could indeed be destiny.

The paper is organised as follows. The next section provides a descriptive overview of the clustering of informal sector activity, in both the manufacturing and services sectors. Section 2.3 starts with a theoretical explanation of the factors influencing the location of economic activity, and presents evidence of how these theories have been tested empirically in the literature. This section also provides an overview of how agglomeration economies may be different for sectors, formal compared to informal, and services compared to manufacturing. Section 2.4 lays out the estimation framework and discusses the main sources of data. Section 2.5 presents the results of the model. Section 2.6 describes the identification strategy employed, disaggregates the results by size and industry-type and studies co-agglomeration and input-output linkages between the organised and the unorganised sectors. Section 2.7 concludes and discusses the implications of the findings.

2.2 Descriptive Analysis

The unorganised sector in India refers to those enterprises whose activities or collection of data is not regulated under legal provision and/or which do not maintain regular accounts. These enterprises are not registered under the Factories Act of 1948. The Act 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. Thus, it can be reasonably assumed that all privately-owned manufacturing enterprises meeting these two criteria are said to be in the organised sector. All public sector enterprises are also automatically assumed to be in the organised sector. Services enterprises are not required to register under the Factories Act (unless they happen to also be engaged in manufacturing activities), and thus, most privately owned services firms are officially classified as being in the unorganised sector. This also applies to large banks, insurance and real-estate firms that maintain proper accounts but which are not registered under the Factories Act. Although the data on unorganised firms excludes firms in the finance and trade sectors, later in the paper, I analyse enterprises by size to try and control for this problem of definition of what constitutes as unorganised for services firms.

The terms ‘unorganised’ and ‘informal’ sector enterprises are used interchangeably in this paper; technically, however, the latter are a subset of the former. The informal sector comprises mainly of unincorporated proprietary or partnership enterprises, while the unorganised sector includes the same along with cooperative societies, trusts and private limited companies. The data used in the paper refers primarily to firms in the unorganised sector. The unorganised sector in India cuts across various well-defined industries and crafts, conglomerates like 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.

The difference between what constitutes as organised and unorganised extends far beyond adherence to a bureaucratic procedure of registration. The unorganised sector consists predominantly of very small-scale enterprises – between 90 to 93 per cent of the establishments in the sample employ less than 5 workers, and between 68 and 70 per cent employ no hired labour at all. Thus, these enterprises are tiny with regard to economic characteristics, and there is very little understanding of the sort of incentives that may drive their location behaviour in India owing to lack of good data.

The unorganised sector in India continues to occupy a substantial place in the country's economy. Its share in the country's Net Domestic Product (NDP) was 56.7% in 2002-03. The importance of the unorganised sector differs substantially across farm and non-farm activities. For instance, in the same year, its share of agricultural NDP was a whopping 96%, and its share of manufacturing and services NDP was 39.5% and 46.9% respectively. The unorganised sector's total NDP contribution can be broken down into its services (43.2%) and manufacturing (16.8%) components. Manufacturing enterprises are often registered because they require more licenses and need access to more infrastructure and capital. On the other hand, service activities can be undertaken without many of these pre-requisites.

The importance of the unorganised sector is even starker with regards to employment. In 2004-05, the unorganised sector was a source of livelihood to approximately 86.3% of the country's workforce. Although a large section of the unorganised sector works within agricultural activities, it is pertinent to note that 71.6% of the total employment in the non-farm sector was also unorganised. In other words, although the unorganised sector contributes just over half of the country's NDP, it employs almost 90% of its workforce.

The contribution of the unorganised sector to employment has also remained broadly stable over the last few decades, with that of the formal sector rising very slowly over time. Informal agricultural employment has barely budged around the 99.4 per cent mark. In fact the proportion of unorganised sector employment has risen for all these sectors, especially for services and manufacturing by a few percentage points over the period of study (1983-94 to 1999-2000). Sectors like electricity, gas and water supply, and transport and communication have also experienced rapid informalisation of their workforce. In other words, the dominance of unorganised employment in the country shows no signs of abating (see Table 2.A.3).

Having established that the unorganised sector is much too important too be ignored, and before studying the impact of various factors affecting the location of unorganised firms, I will now establish that both types of industry within the unorganised sector, manufacturing and services, show evidence of spatial clustering[27] across different districts in India. A study of what drives spatial concentration of economic activity can only be interesting if such patterns exist in the first place.

Over the last decade, there has been much interest in studying the location and the geographic concentration of economic activity. The clustering of economic activity has important implications for development, through its effect on employment and growth. The location of clusters of economic activity drive growth by increasing the productivity of firms and industries, by increasing the pace of innovation through the exchange of ideas and by stimulating the formation of new businesses. This creates jobs, which in turn attracts more people and activity, leading to a virtuous cycle of growth and employment.

The Government of India has focussed much attention on trying to encourage industrial activity in secondary cities or to areas where such activity has not previously clustered or even favoured. The United Progressive Alliance (UPA) government has used a number of tools targeting those within the informal sector. For instance, special health insurance schemes have been set up, social security measures were put into place and funds have been made available for technical, marketing and credit facilities to households, workers and firms in the informal sector. In addition, a number of states in India, for instance, Maharashtra, have provided a slew of incentive schemes focussing

[27] Clustering is a phenomenon in which events or artefacts are not randomly distributed over space, but tend to be organised into proximate groups.

primarily on small-scale enterprises. Such incentives include exemptions from central excise duties, sales tax reductions, product and power subsidies etc. Since enterprises can enjoy these incentives provided their production is below a certain threshold, this has to some extent helped flourish smaller enterprises.

However, even though the unorganised sector is of critical importance to the economy, there is little understanding of what attracts these activities to locations. As pointed out before, unorganised sector establishments look very different from those in the organised sector – they are smaller, have less capital, avoid taxes and regulatory burdens and probably rely more on informal networks and less on public goods such as power, water and transportation. It could be theorised that the economic incentives that drive their location decisions may also be equally different. Note here that I do not make any normative assessments of whether informality should be encouraged or not. There is evidence in other countries (see La porta and Shleifer 2011) to show that informal firms provide inexpensive goods and services of substantially lower quality, however this is also exactly what their customers are able to buy[28].

There are many methods to ascertain whether firms are uniformly distributed across various locations or if they show patterns of spatial concentration. Clustering in its simplest forms can be shown graphically, or through a bird's eye view of where industry is located by means of maps.

Figure 2.A.2 provides maps representing an actual representation of firm density for the country – the size of the circle is proportional to the number of new informal firm births within the district. If we were to assume, as I do later in the paper, that a given firm chooses a location based on the current characteristics of that location, we need to look at the location decisions made by new firms to study if these show any signs of geographical clustering. Studying the past location decisions of existing firms would be difficult in the face of inadequate data with regard to two aspects: data on when these firms made their decisions and data corresponding to the characteristics of the location when these decisions were taken.

[28] Indeed, La Porta and Shleifer (2011) conclude that since informal firms are so inefficient, bringing them into the formal sector by taxing them or requiring them to comply with government regulations would probably drive most of them out of business.

The total number of new informal manufacturing and services units exceeded 2 million respectively. It is clear that whilst some districts in the country host a lot of new unorganised economic activity, others are virtually empty. Also firm births tend to cluster in the same geographical districts, albeit with some differences depending on the type of sub-sector. There are 604 districts in the country, of which informal manufacturing firms are present in 578 districts, and of these around 39 districts account for 50% of all economic activity. On the other hand, informal services firms are present in 556 districts. Of these, around 60 districts account for 50% of all economic activity. In other words, new informal activity is highly concentrated within a few districts in the country.

Of course one could argue that clustering in these districts is simply a factor of the size of the district. And so, the next set of maps carries out the same exercise, but after controlling for the area of the district (in km^2), district population and coastal dummies (i.e. whether the district borders a coast) – and there remains evidence of concentration of economic activity in the country. After adding controls, clustering moves from particular districts to clusters of districts. In other words, the per capita rate remains high for the densely populated districts and for their neighbouring districts (see Figure A.2.3).

Although maps provide a convenient visual representation of the location of new economic activity, more detailed statistics are required to ascertain if there is any evidence of clustering. If economic activity of a particular industry is biased towards a subset of regions, then the industry is said to be 'concentrated'; and if economic activity of a particular region is biased towards a subset of industries, the region is said to be 'specialised'. I use the Theil index to study what regions are specialised, and the Ellison-Glaeser (EG) Index[29] to study concentration across industries (see Appendix A for construction of these indices).

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 NIC sector across districts. The index belongs to the family of generalised entropy inequality

[29] Duranton and Overman (2005) use a more distance-sensitive measure of concentration. I am unable to estimate their index owing to lack of micro-data on firm location.

measures wherein the values vary between 0 and ∞, with zero representing an equal distribution and higher values representing higher values of inequality[30]. I calculate the Theil Index for unorganised manufacturing and services, and for organised manufacturing and services[31]. I present the results in the form of maps for visual comparisons and also list the top ten districts by contribution to the Theil Index. The value of the index for unorganised manufacturing and services equals approximately 3009 and 9503 respectively. The corresponding values for organised manufacturing and services are 4827 and 417. It should be noted that whilst I have comparable data for unorganised manufacturing and services industries (578 and 556 districts respectively), the same is far from true for organised manufacturing and services (495 and 127 districts respectively). It is interesting to note that the maps depicting the contribution of the Theil Index for the unorganised sector correspond to some extent to the visual clustering presented in earlier maps. In other words, districts such as Mumbai, Delhi, Kolkata, Bangalore, Hyderabad, Ahmadabad, Thane, Pune etc show evidence of agglomeration even after using different descriptive techniques to control for district-specific characteristics and for the size and the distribution of firms across districts.

There is also limited concordance between the top ten most clustered districts (arranged by count of new firms, count of new firms controlling for size, and contributions to the Theil Index) across unorganised and organised sectors for both industries, manufacturing and services. In other words, not only does the unorganised sector, for both manufacturing and services industries, shows more evidence of clustering, but it also seems to be clustered in different districts. The differences in the patterns of clustering is interesting, since we may also be interested in exploring the linkages between organised and unorganised activity later in the paper to see if they are substitutes or complements.

The EG Index provides an indication of the district within which new unorganised sector activity is concentrated. 2.A.6 and Table 2.A.7 provide the EG Indices for the

---

[30] The value of the index increases in the inequality of the distribution of firm births by district with respect to total firm births: $T = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{x_j}{\bar{x}}.\ln\frac{x_j}{\bar{x}}\right)$, where x_j is the number of firm births in district j.

[31] As explained later in the text, I use the Annual Survey of Industries for data on organised manufacturing, and the Prowess database for organised services.

unorganised sector, for manufacturing and services industries, across districts. The Index has the property of controlling simultaneously for the employment distribution among firms and regions. In their paper, Ellison and Glaeser (1997) demonstrate that the index takes the value of zero under the null hypothesis of random location conditional on aggregate employment in that region. In other words, the no-agglomeration benchmark is when the value of the index is zero (i.e. $E(\gamma)=0$). In general, if the EG index is greater than 0.05, the industry is considered to be highly concentrated. I find that manufactures of office, accounting and computing equipment, transport and communications equipment, and that of leather products, among others is highly concentrated in a few districts. Services related to research and development, computers and supporting transport and other activities also shows evidence of much concentration.

Having established that there is overwhelming evidence of clustering in unorganised industry across different districts in India, this paper will examine the factors that drive such clustering. In particular it will focus on identifying the role of agglomeration economies in influencing the decision of firms to cluster, i.e. to locate close to one another. It will examine the nature and scale of agglomeration economies using district and NIC 2-digit-level data for unorganised firms in India.

2.3 Theoretical background and Literature

This section will provide a brief overview of the theoretical understanding of agglomeration economies and outline a few empirical studies of relevance.

For an excellent overview of the location theory, see Brulhart (1998) (Table 1, Page 778) that describes the different theoretical schools and lists their principal distinguishing features. Marshall (1919) was the first to identify the benefits from industrial clustering. Clusters of firms, predominantly in the same sector, could take advantage of localisation economies, such as the sharing of sector-specific inputs, skilled labour and knowledge. Thus, cost-saving externalities are maximised when a local industry is specialised. The Marshall-Arrow-Romer (Marshall 1890, Arrow, 1962, Romer 1986) models predict that such externalities predominantly occur within the same industry. Therefore, if an industry is subject to localisation externalities, firms are likely to locate in a few regions where other firms in that industry are already clustered.

The next level is that of inter-industry clustering[32], i.e. when firms in a given industry and those in related industries agglomerate in a particular location. The benefits of clustering would include inter-industry linkages, buyer-supplier networks, and opportunities for efficient sub-contracting. Venables (1996) demonstrates that agglomeration could occur through the combination of firm location decisions and buyer-supplier linkages, since the presence of local suppliers could reduce transaction costs and increase profitability. Inter-industry linkages can also serve as a channel for vital information transfers.

An overall large size of the urban agglomeration and its more diverse industry mix is also thought to provide external benefits beyond those realised within a single sector or due to a tight buyer-supplier network (Henderson 2003). Chinitiz (1961) and Jacobs (1969) proposed that important knowledge transfers primarily occur across industries and the diversity of local industry mix is important for these externality benefits. These benefits are typically called urbanisation economies and include access to specialised financial and professional services, availability of a large labour pool with multiple specialisations, inter-industry information transfers and the availability of less costly general infrastructure. Larger cities also provide a larger home market for end products, and make it easier to attract skilled employees. Other factors that make big cities more attractive are urban amenities not available in smaller towns and a large number of complementary service providers such as financial and legal advisers, advertising and real estate services etc.

Thus, industrial clustering could take place at different levels, which would have different implications for the associated agglomeration economies. A firm could gain from economies of agglomeration that arise from localisation economies, that occur as a result of concentration of firms within the same industry; inter-industry economies, that occur as a result of concentration of firms in related industries in a particular area; and

[32] As Deichmann et al (2005) points out, empirically the distinction between own-industry versus cross-industry is dependent on the level of sectoral aggregation.

urbanisation economies, that occur across all industries as a result of the scale of a city or region by means of its large markets and urban diversity. It is also pertinent to note that localisation, inter-industry and urbanisation economies are not mutually exclusive – they may occur individually or in combination.

In the empirical literature, there are two broad approaches to identify the determinants of firms' location decisions. One is survey-based or the 'stated preference' approach', for instance to ask firms directly, through an investment climate survey, for instance, about what location factors are important to them. The second approach is a modelling approach or an econometric analysis of empirical patterns used to identify 'revealed preferences' based on the characteristics of the region.

To the best of my knowledge, there are no empirical tests in the literature on factors that could drive the location decisions of informal activity[33]. The existing research looks mainly at the formal sector – whether for manufacturing, or services or both. For instance, with regards to formal manufacturing in India, Lall and Meningstae (2005) analyse the productivity of plants sampled from 40 of the country's largest industrial cities and found that differences in clustering across locations were explained by market access, labour regulation and the quality of power supply. With regards to foreign entrants into domestic manufacturing sectors, Head and Reis (1996) show that foreign firms in China preferred to locate in cities where other foreign firms are located. In their paper Head and Mayer (2004) show that downstream linkages made regions in Europe more attractive to Japanese investors, but the paper does not account for access to suppliers. Cheng and Kwan (2000), and Amiti and Javorcki (2005) also confirm that regional markets and buyer-supplier linkages were important factors affecting the location decisions of foreign firms.

Services firms are theorised to be different from manufacturing. For instance, in some services, product specialisation, rather than standardisation, may be more important in capturing markets (Enderwick 1989), and proximity to competitors, suppliers and

[33] There is however, a large and developed literature on workers' occupational choices, with their intellectual roots in the Harris-Todaro (1970) model in which workers compare expected incomes in a dual-sector setting. Within this setting, Tiglao and Tsutsumi (2005) model location choices of informal households in Manila, in conjunction with occupational sector choices.

markets may be significant determinants of location decisions (Bagchi-Sen 1995). With the introduction of new communication technologies and the ability to slice the service production chain more thinly, it could be argued that proximity would cease to be an important factor in explaining agglomeration economies. Earlier research conducted in North America (Kirn 1987 for the US, and Coffey and McRae 1989 for Canada) found that producer services did not necessarily follow population and manufacturing location patterns – they could locate in peripheral regions and develop an export base. However, more recent research (Dekle and Eaton 1999, Coffey and Shearmur 2002) found evidence that the agglomeration economies exerted a stronger influence in services than in manufacturing, in spite of advances in information and communications technology.

There are a number of reasons why informal activities are different from the formal economy. For instance, they are usually an extension of the household economy and start-ups that require little or no capital investment. Informal sector enterprises in India comprise of unregulated micro-enterprises, the bulk of which employ less than five workers, and all of which employ less than 50 workers. Examples of such enterprises are those that produce bidis (Indian cigarettes), small piece-rate suppliers to the textile, weaving or footwear sectors, small shopkeepers etc. The informal sector is also the largest employer of rural migrants in big cities like Mumbai, Kolkata and Delhi, and like in other countries, the sector serves as the only source of employment to those who are unable to find work in the formal economy. Thus, small enterprises have been viewed as an important means of promoting industrialisation and employment in poor countries.

McGee (1977) noted that the informal sector in South-East Asian cities tended to concentrate in areas of dense population such as nodes of transportation, or where there are adjacent activities such as entertainment complexes, public markets and also in those localities where they could benefit from product complementarities and mutual customer attraction. A priori, there is no reason to assume that informal sector activity remains unaffected by agglomeration economies. Indeed, it could be hypothesised that 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 them than to firms operating in the organised sector. The informal sector in India largely ignores labour regulations, officially recognised 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.

Production in the formal sector is also dependent on subcontracting among informal firms specialised in some aspect of the vertical production chain. Although parts of the unorganised sector pertain mostly to the production of non-tradables in the economy (think of street vendors and domestic help) they are also an important input to the production of intermediate goods, processed exports and import substitutes, supported by supply side contracts with the formal sector. For instance, informal carpet weavers in Agra operate alongside larger, more formal carpet designers and exporting firms in the city. And to the extent that the informal sector is linked to its formal counterpart, wages in the sector could be affected by structural changes in the formal industrial sector.

With the theoretical and empirical literatures in mind, this paper will concentrate on the extent to which agglomeration economies matter to informal firms' location decisions, and compare them to those in the formal sector. The next section will describe the estimation framework employed and them move on to discussing the results and possible endogeneity bias.

2.4 Estimation Framework

2.4.1 Econometric model

A popular model of location choice are conditional logits which assume that a firm evaluates alternative locations at each time period, and would consider relocation if its profitability in another place exceeded that at its current location[34]. The use of a discrete choice framework to model location behaviour goes back to the 1970s, when Carlton

[34] In reality, relocation can be costly and firms need to take account of sunk investments in production capacity, and other costs of moving. However, these relocation costs are not considered in the model.

(1979) adapted and applied McFadden's (1974) Random Utility Maximisation (RUM) Framework to firm location decisions.

Within such a discrete choice framework, a general profit function is used to explain how new firms choose a location. Following McFadden the model assumes a set $J = (1,2,.....j,.....n)$ of possible locations (districts) assuming that location j offers profitability level π_{ijk} to a firm i in industry k. The resulting profitability equation yielded by location j to a firm i in industry k is:

$$\pi_{ijk} = \beta Z_{ijk} + \varepsilon_{ijk} \tag{1}$$

where β is the vector of unknown coefficients to be estimated and ε_{ijk} is a random term. Thus, the profit equation is composed of a deterministic and a stochastic component. Under the assumption of independent and identically distributed error terms ε_{ijk}, with type I extreme-value distribution, then it can be assumed that the *i*th firm will choose district *j* if $\pi_j^i \geq \pi_l^i$ for all *l*, where *l* indexes all the possible location choices to the *i*th firm. Thus, the probability that any firm will choose to locate in a district *j* is given by:

$$p_{ijk}(\pi_{ij} \geq \pi_{il} \forall l \neq j) = \frac{e^{\beta Z_{ijk}}}{\sum_{m=1}^{J} e^{\beta Z_{ijk}}} \tag{2}$$

where p_{ijk} is the probability that firm i in industry k locates in district j. If we let $d_{ijk} = 1$ if firm i of industry k picks location j, and $d_{ijk} = 0$ otherwise, then we can write the log likelihood of the conditional logit model as follows:

$$\log L_{cl} = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} d_{ijk} \log p_{ijk} \tag{3}$$

In practice, however, the implementation of the conditional logit model in the face of a large set of spatial alternatives is very cumbersome[35]. The conditional logit model is also characterised by the assumption of Independence of Irrelevant Alternatives (IIA). Consequently, the ratio of the logit probabilities for any two alternatives does not depend on any alternatives other than the two considered. More formally, this implies that the ε_{ijk}s are independent across individual firms and choices; all locations would be symmetric substitutes after controlling for observables. This assumption would be violated if districts within particular states were closer substitutes than others outside of the state boundary. The addition of dummy variables for each individual choice would effectively control for choice specific unobservables, amounting to the following specification:

$$\pi_{ijk} = \delta_j + \beta Z_{ijk} + \varepsilon_{ijk} \tag{4}$$

where δ_js are the alternative specific constants introduced to absorb factors that are specific to each particular choice. In this case all explanatory variables (observable or unobservable) that only change across choices are absorbed by the alternative specific constants. In the presence of large datasets, such as the one I plan on using, this implementation would be impractical because of the large number of parameters to be estimated. And this would still leave the problem of the IIA unsolved.

To overcome the potential IIA problem and to remain computationally tractable in the face of numerous location alternatives caused by the use of disaggregate or micro level of geography, a number of studies investigated the applicability of count models such as Poisson and Negative Binomial (NB), to predict the number of firms located on each of a large number of alternative locations (Papke 1991, Becker and Henderson 2000, Coughlin and Segev 2000, Holl 2004a, 2004b, 2004c, Carod and Antolin 2004). Papke (1991) models the number of firm formation with a Poisson distribution and also controls for unobserved location heterogeneity using a fixed-effects framework. Wu (1999) compares the applicability of Poisson and NB in intra-metropolitan location behaviour of foreign direct investment in Guangzhou, China. Becker and Henderson (2000) investigate the impacts of environmental regulation on the formation of polluting

[35] Guimaraes et al. (2003) provide an overview of the problems and how different researchers have attempted to deal with them in the past.

plants, where they indirectly compare empirical results between fixed-effects Poisson and conditional logit. Coughlin and Segev (2000) apply NB to examine the location determinants of foreign-owned manufacturing plants at US county level. Using Portuguese municipality data, Holl (2004a) estimates a fixed-effects Poisson model, Holl (2004b) estimates a fixed-effects NB model, and Holl (2004c) estimates both a fixed-effects Poisson and NB when studying location behaviour of new or relocating firms. Carod and Antolin (2004) focus on the level of geographic aggregation and present three econometric results based on a multinomial logit, conditional logit and the Poisson model, but they do not directly address the comparability between the multinomial logit and Poisson.

In this way, count models gained popularity as the number of alternative locations increased, since what these lead to computational burdens in conditional logit models but in count models these are an advantage owing to the availability of more numerous observations. In addition, one might also think of the problem as that of explaining the firm births within a location as a function of the characteristics of that location – see Ghani et al (2011). If this were the case, then the functional form of a count model, wherein the dependent variable is the count of the new firms within a given industry within a given district, would also perform well. Unlike an ordinary least squares (OLS) specification, a count specification would have the added benefit of allowing for the possibility of zero counts. However, count models were at the time not understood to be as theoretically well founded as the conditional logit model, which is based on the RUM framework. This was until Guimaraes et al (2003, 2004) showed that count models can be specified in a way that is theoretically and empirically consistent with conditional logit models and thereby the RUM framework.

Guimaraes et al 2003 show that the implementation of conditional logit models yields identical log-likelihood functions to Poisson regression models when the regressors are not individual specific. They demonstrate how to control for the potential IIA violation by making use of an equivalence relation between the conditional logit and Poisson regression likelihood functions. In a separate paper, Guimaraes et al (2004) provide an empirical demonstration. In this model the alternative constant is a fixed-effect in a Poisson regression model, and coefficients of the model can be given an economic interpretation compatible with the Random Utility Maximisation framework.

Guimaraes et al (2003) demonstrate that Equation (3) is equivalent to that of a Poisson model that takes the number of new firms in a district, n_{jk}, as the dependent variable and includes a set of location-specific explanatory variables. The same results will be obtained if we assume that n_{jk} follows a Poisson distribution with expected value equal to $E(n_{jk}) = \lambda_{jk} = \exp(\alpha h_{jk} + \beta Z_{jk})$, where $[\alpha, \beta]$ is the vector of parameters to be estimated and h_{jk} is a vector of K dummy variables, each one assuming the value 1 if the observation belongs to industry k. Thus, the above problem can be modelled as a Poisson regression where the $[\alpha, \beta]$ vector can be estimated regardless of the number of δ_j parameters. Information on actual individual firm choices is grouped into vectors of counts without any loss of information. This occurs since there are groups of firms faced with the same choice set and the same choice characteristics. For instance, consider the problem of identification of the relevant regional factors that affect firm location. Typically, researchers view these individual location decisions as profit (utility) maximising actions. Firms from diverse sectors evaluate the regional characteristics of different regions (i.e. districts) and choose to locate in the region that maximises potential profits. In this case, it is common to assume that all firms face the same choice set, and the relevant characteristics of the regional choices are identical for firms belonging to the same industry. The available information consists of regional counts of firm births by industry and variables that reflect the characteristics of the regions. Despite the fact that the data consist of individual level choices, the true variation of the data is at the group level. Thus, data for the dependent variable may be summarised by vectors of counts.

I am interested in modelling the data using McFadden’s discrete choice Random Utility Maximisation (RUM) framework. This means that inference is based on the multinomial distribution because my interest lies in studying the impact that covariates have on choice probabilities, treating the number of firms in each industry group as given, wherein all firms share some common industry-level characteristics. This introduces the possibility that there exist some unobservable industry-specific effects that are likely to equally influence all firms belong to the same industry. If this happens, then the individual choices will be correlated and the vectors of counts will exhibit extra multinomial variation (i.e. overdispersion). Much like what happens with count models, in this circumstance the conditional logit model will remain consistent but will tend to underestimate the variance-covariance matrix.

Guimaraes and Lindrooth (2007) illustrate how the problem can be tackled by using Negative-Binomial count models, based on the Dirichlet-multinomial distribution. The Dirichlet-Multinomial regression is a natural extension of McFadden's conditional logit model, and their paper shows that this relationship is the same as that of the negative binomial regression to the Poisson regression. In other words, the NB model provides a parametric alternative to deal with the problem of overdispersed data, and the parameters of the model would be equivalent to the Dirichlet-multinomial regression, which is an extension of the conditional logit regression. In the same way that the likelihood for the conditional logit model is obtained by letting the n_{jk} (i.e. the number of new firms in a given industry and district) following a Poisson law and conditioning on the total sum for each group, a similar relationship can be derived for the Dirichet-multinomial model. In this case, n_{jk} is modelled directly as an overdispersed count variable, wherein the number of new firms is distributed according to the negative binomial law implying that the total sum of counts for each group also follows a negative binomial distribution. Thus, the approach by Guimarares et al (2003, 2004) and Guimaraes and Lindroth (2007) effectively controls for the IIA assumption, for conditional logit and for Dirichlet-logit models. Accordingly, in this I will use the equivalent Poisson and negative binomial regressions to generate coefficients. See Mukim and Nunnenkamp (2012) for a comprehensive list of empirical papers that use count models and those that use conditional logits.

To sum up, I test the importance of economic geography and locational factors by implementing a count model, wherein the count of new firms within a location is modelled as a function of factors common to the location and those common to particular sectors within a location. The original estimation framework is based on a location decision model in which individual firms compare profitability across different locations. Since a firm would choose a location depending on the characteristics of the location at that given point in time, the model studies the decisions made by new firms. In other words, the location decisions of firms are modelled at the point of time at which they begin operations at a particular location as a function of the characteristics of the chosen location at the time of the decision to locate. As explained later, I use the count of new firms for the years for which I have adequate data for the explanatory variables.

2.4.2 *Specification of variables*

The deterministic component of the function consists of the various attributes of the location that can influence the profitability of a firm in that particular location, the random component consists of the unobserved characteristics of the location, and measurement errors. The dependent variable in the model is the count of new informal firms at time t, whilst all the explanatory variables in the model are defined at time $t-1$. Section 4.3 below describes the sources of data and the cross-sectional time period for manufacturing and services firms in detail. To reduce clutter, I don't show the time subscript below.

The observables in this model are given by:

$$Z_{ijk} : \sigma_{jk}, \Lambda_{jk}, U_j, MA_j, Ed_j, X_j, W_j, WE_j$$

Where:

σ_{jk} represents localisation economies, represented by the share of employment in industry k found in district j

Λ_{jk} represents inter-industry trading relations measured by the strength of buyer-supplier linkages

U_j represents urbanisation economies in district j (measured by the Herfindahl Index to reflect industrial diversity)

MA_j summarises access to markets in neighbouring districts

Ed_j measures the level of human capital in district j

X_j captures the quality and availability of infrastructure (electricity and communications)

W_j a vector of factor input price variables in district j

WE_j captures the level of wealth in district j

ξ_j measures unobserved characteristics of the district which can affect the firm's profitability. Each firm considers these factors at the time it is making its location decision, but these are not captured in the data. The specifics of the endogeneity problem are dealt with in more detail in Section 2.6.

The economic geography variables in this model are represented by market access (MA_j), localisation economies (σ_{jk}), inter-industry economies (Λ_{jk}) and urbanisation economies (U_j). The variables representing business environment are Ed_j (educational attainment) X_j (quality and availability of power and communications' infrastructure) and WE_j (wealth). The remainder of this section provides a detailed description of each of the variables used in the model.

Localisation economies (σ_{jk}) can be measured by own industry employment in the region, own industry establishments in the region, or an index of concentration, which reflects disproportionately high concentration of the industry in the region in comparison to the nation. I measure localisation economies as the proportion of sector $k's$ employment in district j as a share of all of sector $k's$ total employment in the country. The variable takes a different value for each industry in a given district, across districts. The higher this value, the higher the expectation of intra-industry concentration benefits in the district.

$$\sigma_{jk} = \frac{E_{j,k}}{E_k}$$

There are several approaches for defining inter-industry linkages: input-output based, labour skill based and technology flow based. Although these approaches represent different aspects of industry linkages and the structure of a regional economy, the most common approach is to use the national level input-output accounts as templates for identifying strengths and weaknesses in regional buyer-supplier linkages (Feser and Bergman 2000). The strong presence or lack of nationally identified buyer-supplier linkages at the local level can be a good indicator of the probability that a firm is located in that region. To evaluate the strength of buyer (supplier) linkages for each industry, a summation of regional (here district) industry employment weighted by the industry's input (output) coefficient column (row) vector from the national input-output account is used:

$$\Lambda_{jk} = \sum_{k=1}^{n} w_k e_{jk}$$

where, Λ_{jk} is the strength of the buyer (supplier) linkage, $\square$ is industry *k's* national input (output) co-efficient column (row) vector and e_{jk} is total employment for industry *k* in district *j*. The measure examines local level inter-industry linkages based on national input-output accounts. The national I-O coefficient column vectors describe intermediate goods requirements for each industry, whilst the I-O coefficient row vectors describe final good sales for each industry. Assuming that local industries follow the national average in terms of their purchasing (selling) patterns of intermediate (final) goods, national level linkages can be imposed to the local level industry structure for examining whether district *j* has a right mix of buyer-supplier industries for industry *k*. By multiplying the national I-O coefficient vector for industry *k* and the employment size of each sector in district *j*, simple local employment numbers can be weighted based on what industry *k* purchases or sells nationally.

I use the Herfindal measure to examine the degree of economic diversity, as a measure of urbanisation (U_j) in each district. The Herfindal index of a district j (U_j) is the sum of squares of employment shares of all industries in district j:

$$U_j = \sum_k \left(\frac{E_{jk}}{E_j} \right)^2$$

Unlike measures of specialisation, which focus on one industry, the diversity index considers the industry mix of the entire regional economy. The largest value for U_j is one when the entire regional economy is dominated by a single industry. Thus a higher value signifies lower level of economic diversity.

In principle, improved access to consumer markets (including inter-industry buyers and suppliers) will increase the demand for a firm's products, thereby providing the incentive to increase scale and invest in cost-reducing technologies. The proposed model will use the formulation proposed initially by Hansen (1959), which states that the accessibility at point 1 to a particular type of activity at area 2 (say, employment) is directly proportional to the size of the activity at area 2 (say, number of jobs) and inversely proportional to some function of the distance separating point 1 from area 2. Accessibility is thus defined as the potential for opportunities for interactions with neighbouring districts and is defined as:

$$MA_j = \sum_m \frac{S_m}{d_{j-m}^b}$$

Where, MA_j is the accessibility indicator estimated for location j, S_m is a size indicator at destination *m* (in this case, district population), d_{jm} is a measure of distance between origin *j* and destination *m*, and *b* describes how increasing distance reduces the expected level of interaction[36]. The size of the district j is not included in the computation of market access – only that of neighbouring districts is taken into account[37]. The accessibility indicator is constructed using population (as the size indicator), distance (as a measure of separation) and is estimated with the exponent value set to 1. The market access measure has been constructed by allowing transport to occur along the orthodromic distance[38] connecting any two districts within a 500-kilometre radius.

I also use data on education to assess the role played by the human capital across different districts on the decisions of firms across different industries to situate themselves in a particular district. I include a measure of the effect of education, captured by the education variable - Ed_j. This is defined as the proportion of the population within the district with a high-school education.

I define X_j as a measure of 'natural advantage' through the embedded quality and availability of infrastructure in the district. I use the availability of power (proxied by the proportion of households with access to electricity) within a location as an indicator of the provision of infrastructure. In addition I also use the proportion of households within a district with a telephone connection as an indicator of communications' infrastructure.

---

36 In the original model proposed by Hanson (1959), b is an exponent describing the effect of the travel time between the zones.

37 The final specification includes population to control for the size of district j.

38 Also known as great-circle distance, it is the shortest distance between any two points on the surface of a sphere.

W_j is an indicator of labour costs in a district, and is given by nominal district-level wage rates (i.e. non-agricultural hourly wages). The expected effect of this variable is hard to pin down theoretically. On the one hand, if wages were a measure of input costs[39] then one would expect informal activity to be inversely related to wages, since high costs within a location would make it less attractive. However, it is also important to control for the skill set of the workers since a positive coefficient on wages could be proxying for more skilled-labour. In theory, workers with higher ability could demand a higher wage rate and in turn enjoy a higher level of consumption. And in line with Rauch's (1992) findings[40] one might expect that better-educated workers are able to capture the returns from agglomeration in the form of higher wages. If this were the case, higher wages would not dissuade firms that benefit from agglomeration economies from choosing a location, since the higher costs would be defrayed by higher productivity. In general one would expect informal firms to pay low wages consistent with some proportion of production being carried out by household enterprises.

Although I am unable to directly control for the ability of the worker, I include education as a proxy for the level of human capital within the district. I then use the proportion of high-income households (WE_j) within a district is an indicator of the general level of wealth, or more specifically, consumer expenditure 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 group[41].

In summary, the economic geography variables are supplemented with controls for infrastructure (education, electricity and telephone), market size (wealth) and input cost (wages). It is also standard to control for transport infrastructure, but owing to lack of data I have been unable to include road or rail transport controls at the district level.

[39] Since the vast majority of informal firms are small enterprises and have little or no access to capital or machinery, labour would be their primary input. Thus, wage rates would serve as a good proxy for general input costs.

[40] Using the 1980 US Census, Rauch finds that wages are higher in cities with higher average education and that the higher returns to education in denser markets could arise because higher levels of human capital could result in the transmission of better ideas.

[41] 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.

Table 2.1: Descriptive Statistics

Variable	Expected Sign	#		Mean	
		Manufacturing (2004-2005)	*Services (1999-2000)*	*Manufacturing (2004-2005)*	*Services (1999-2000)*
New firms		567	572	3,531	4,111
Localisation	+	557	469	0.003	0.002
Input	+	557	462	4213.2	3821.3
Output	+	557	462	2189.6	8237.7
Urbanisation	-	578	586	0.41	0.33
Market Access	+	574	582	869363	871313
Education	+	578	480	0.074	0.056
Electricity	+	578	486	0.633	0.559
Telephone	+	578	486	0.368	0.083
Wealth	+	578	486	0.051	0.054
Wages	-/+	574	483	100.94	93.47

Notes: # refers to the number of districts for which data is available. There are a total of 604 districts in the country.

2.4.3. Data Sources

The dependent variable, used in the reduced form estimation, is the count of new firms within the informal sector in India. I run two separate regressions, for manufacturing and services industries. The data is drawn from the Fifty-Seventh Round (July 2001-June 2002: Unorganised Service Sector) and the Sixty-Second Round (July 2005-June 2006: Unorganised Manufacturing Enterprises) of the National Sample Survey Organisation. The former household survey contains data on services enterprises in the informal sector (NIC division 38-97), and the latter on manufacturing enterprises in the informal sector (NIC division 15-37). Enterprises are divided into (1) own account enterprises, which are normally run by household labour and which do not hire outside labour on a regular basis, (2) non-directory establishments, which employ one to five workers (including household and hired taken together) and (3) directory establishments, which employ six or more workers (including household and hired taken together).

I extract data on new firms from the question that asks the enterprise its status over the last 3 years (expanding/stagnant/contracting/operated for less than 3 years). I select enterprises that respond in the positive to the latter option, in each of the two surveys. The surveys also contain data on the district within which the enterprise is located. The total number of new services firms within the 1999 survey equals 2,409,204 and the count of new manufacturing firms for the 2004 survey is 2,041,137. In short, I carry out two separate cross-sections, one for unorganised manufacturing firms and the other for unorganised services firms. Since the surveys sample different firm populations, I am not able to exploit changes between the two rounds. However, I am interested in looking at what factors drive unorganised manufacturing and/or services firms to a district.

The choice of years is dictated by the data. Whilst data on the dependent variable is drawn from the NSSO Rounds described above, I extract data from the Employment and Unemployment Surveys - Round 55.10 (July 1999 – June 2000) and Round 61.10 (July 2004 – June 2005). The former is the source of explanatory variables for the cross-sectional analysis for services, and the latter for manufacturing. This data, which is disaggregated by industry and district, allows me to construct my agglomeration variables. On average I have data on 24 and 16 two-digit industries for manufacturing and services respectively. It is important to keep in mind that since employment data is taken from household surveys, it includes employment within the economy as a whole, and does not differentiate between the formal and the informal sector. In other words, the construction of localisation, input-output and urbanisation economies already assumes linkages between the organised and unorganised sectors. Data on education, electricity and communications infrastructure, and on wages and wealth within the district are also drawn from the household surveys. I use population data from the 2001 Census to construct the market access variable.

Table 2.2: Predictor Variables

	Variable	Indicator	Source(s)	Availability 1999-2000	Availability 2004-2005
Economic Geography	Localisation	Intra-industry concentration	NSSO	√	√
	Input/Output economies	Buyer/Supplier linkages	NSSO	√	√
	Urbanisation	Industrial Diversity	NSSO	√	√
	Market Access	Neighbouring markets	Orthodromic distance calculations	√	√
Business Environment	Education	Persons with a High-School education	NSSO	√	√
	Electricity	Persons with access to electricity	NSSO	√	√
	Telephone	Households with a telephone connection	NSSO	√	√
	Wages	Non-agricultural hourly wages	NSSO	√	√
	Wealth	High-income households	NSSO	√	√

Notes: NSSO - National Sample Survey Organisation

2.5 Results and Discussion

I start with an illustration of the characteristics of the data to explain my modelling choices. The first observation is that the data is over-dispersed. The mean number of new firms per district is around 4,111 for the services sector, and 3,531 for the manufacturing sector (see Table 2.A.8). At the same time the respective standard deviations are around 1.6 to 2.3 times the mean. A Poisson model implies that the expected count, or mean value, is equal to the variance. This is a strong assumption and does not hold for my data. A frequent occurrence with count data is an excess of zeroes – in this case, however, this is not a significant problem. Only 29 districts (of a total of 586) have zero new services units, and 52 districts (of a total of 578) have zero new manufacturing units.

I also check the suitability of the different types of models with regards to their predictive power. 'Obs' refers to actual observations in the data, and Fit_p and Fit_nb refer to the predictions of the fitted Poisson and Negative Binomial (NB) models respectively. Of all the locations in the sample, 4.9% have no new services units, and 9% have no new manufacturing units. In both cases, the Poisson model (Fit_p) predicts that 0% of all districts would have no new units – clearly the model underestimates the probability of zero counts. The negative binomial (Fit_nb), which allows for greater variation in the variable than that of a true Poisson, predicts that 0.66% and 3.25% of all districts will have no new services or manufacturing units respectively.

The response variable is 'count', i.e. the number of new firms in a given industry in a given district. The Poisson regression models the log of the expected count as a function of the predictor variables. More formally, $\beta = \log(\mu_{x+1}) - \log(\mu_x)$, where β is the regression coefficient, μ is the expected count and the subscripts represent where the regressor, say x, is evaluated at x and at $x+1$ (here implying a unit percentage change in the regressor[42]). Since the difference of two logs is equal to the log of their quotient, i.e. $\log(\mu_{x+1}) - \log(\mu_x) = \log(\frac{\mu_{x+1}}{\mu_x})$, thus one could also interpret the parameter estimate

[42] This is because the regressors are in logarithms of the original independent variables.

as the log of the ratio of expected counts. In this case, the count refers to the 'rate' of new firms per district.

Table 2.3 presents the results for informal manufacturing and Table 2.4 for informal services. The impact of economic geography variables represented by localisation, input, output, urbanisation (industrial diversity) and market access is studied in model (1), while business environment variables represented by education, telephone, electricity, wages and wealth are introduced in model (2). In model (3) I introduce industry fixed-effects to control for differences across two-digit industries, and in model (4) I re-run the specification in model (3) within a Poisson model, i.e. assuming the data is not overdispersed. . I also control for the size of the district (population). The coefficients can be interpreted as follows – if localisation were to increase by one percent, the expected number of new informal manufacturing firms would increase by 28.98 percent (see specification 3 in Table 2.3).

Table 2.3: Informal Manufacturing

Variable	**Negative Binomial**			
	(1)	**(2)**	**(3)**	**(4)**
Localisation	-0.9739*	-0.8976	0.2898***	0.6951
Input	1.0012***	1.0098***	0.0787	0.0725
Output	-0.6382***	-0.6515***	0.2575**	0.2122
Urbanisation	-0.0508	0.0268	-0.1895	-0.2321
Market Access	0.0970	0.0403	0.0103	0.0377
Education		0.0034	0.0056	-0.0149
Telephone		0.0060**	0.0061**	-0.0009
Electricity		-0.0069**	-0.0039	-0.0040
Wages		-0.0006	-0.0022	-0.0020
Wealth		0.0191**	0.0179**	0.0361***
Population	0.6636***	0.6524***	0.7722***	0.9900***
Industry Fixed Effects	✗	✗	✗	✓
#	3,762	3,762	3,762	3,762
AIC	31409	39109	30927	29795
BIC	32460	39162	30005	30984

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

As the model selection criteria I examine and compare the Bayesian information criterion (BIC) and Akaike's information criterion (AIC). Since the models are used to fit the same data, the model with the smallest values of the information criteria is considered better. The results of the negative binomial models, for both manufacturing and services, have the best goodness-of-fit statistics.

After controlling for the two-digit industrial sector, localisation (σ_{jk}) has a strong positive and statistically significant effect – since localisation refers to the clustering of firms within the same industry within a location, this indicates that new firms are attracted to districts where there is strong within-industry clustering of firms. The opposite result is observed without the inclusion of industry dummies (i.e. in model 1 and model 2), indicating that when the coefficient on localisation is averaged across all industries, it has a strong negative effect. Output linkages, i.e. linkages to buyers, also have a positive and significant effect on the attractiveness of a district to new informal manufacturing activity. On the other hand, the presence of intermediate goods' suppliers has no discernible impact on firms' location decisions, once industry dummies are included. Market access (MA_j), i.e. being located close to larger, more populated districts again seems to have no effect on how attractive a district is to informal manufacturing activity.

With regard to business environment variables, the effect of education and power infrastructure seems to be insignificant. The proportion of the population with access to a telephone line seems to have a positive effect – a unit increase in the proportion raises the expected number of new informal manufacturing firms by 0.61 percent. Wages also seem to be unrelated to a location's attractiveness – as mentioned before I am unable to directly account for the skill set of the worker. However, I do include the proportion of wealthy households within the district as a control – this would allow me to control for the ability of some workers to demand higher wages, and also provide an indication of the demand within a district. The proportion of wealthy households within the district and the size of the district, both seem to have a significant and positive effect on the count of new informal manufacturing firms.

The results (Table 2.4), after controlling for industry, for informal service firm births show that the higher the intra-industry concentration, the more the attractiveness of the location. In fact if localisation were to increase by one percent, the expected count of new informal services firms would increase by 34.21 percent (see model 3). This is especially informative since informal services refer mainly to small shopkeepers and households providing services, implying that even for new small-scale units proximity to those within the same industry is an important factor. Just as for the case of informal manufacturers, the effect of output linkages is positive and significant implying that informal services tend to be attracted to those industries that they supply to, but not to those that they may buy intermediate products from. Industrial diversity within a district seems to have a negative and significant effect - recall that since a higher Herfindahl index implies lower industrial diversity, the direction of the sign of the coefficient is evidence of a positive association between more industrial diversity and more profits, or greater attractiveness of the district. Access to larger markets in neighbouring locations also increases the count of new informal services firms in a location.

Table 2.4: Informal Services

Variable	**Negative Binomial**			
	(1)	**(2)**	**(3)**	**(4)**
Localisation	-0.3389***	-0.4704***	0.3421**	-0.3099
Input	0.0780*	0.0838**	-0.3303*	-0.1322
Output	0.1906***	0.1788***	0.3770***	0.2862***
Urbanisation	-0.3002***	-0.2370*	-0.3977***	-0.6186***
Market Access	0.1520	0.2524**	0.2275**	0.0129
Education		-0.0305***	-0.0190*	-0.0192
Telephone		0.0191*	0.0194*	0.0001
Electricity		0.0086***	0.0103***	0.0082***
Wages		-0.0012	-0.0016	-0.0048***
Wealth		0.0058	0.0024	0.0034
Population	0.7385***	0.7935***	1.0458***	0.6972***
Industry Fixed Effects	✗	✗	✗	✓
#	2,655	2,655	2,655	2,655
AIC	34029	56743	36781	35117
BIC	33098	53569	40562	36263

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

With regard to the business environment, access to electricity and telephones has a positive and significant effect, whilst education seems to have a negative effect. This would imply that a lower educational attainment is associated with making the district more attractive to new services firms. A closer look at the unorganised sector by educational level reveals that almost 90% of those with less than a high-school degree found themselves working in the unorganised sector. In other words, districts with a larger proportion of the population with a lower level of education attract more unorganised services activity. The size of the district, i.e. population, strongly attracts informal services activity. This is intuitive since one would expect clustering from personal consumer services (such as hairdressers, or rickshaw drivers) that supply the final demands for consumers and thus need to be located close to urban populations.

What is interesting to note, is that although the effect of most variables is consistent whether averaged across industries or within, that of the three economic geography variables, i.e. localisation, input and output linkages, changes sign dramatically with the introduction of industry fixed-effects in Model 3 in some of the specifications. This would imply that within-industry variation has a very different effect from that of across industries. Controlling for industry group is important since the results should reflect that agglomeration economies might be very different for different types of industries. Indeed, I explore the industry-specific effects in more detail in Tables 2.A.9 and 2.A.10, for manufacturing and services industries separately. The effect of localisation is positive and significant for manufacturing industries such as those related to leather treatment and production, communication equipment, other transport material (essentially tyres, car covers etc), and furniture. Industrial diversity is important for informal firms in wearing apparel, rubber and plastic products and most informal services activities.

In general, the effects of localisation and output-linkages, and the absence of the effects of wages, are broadly stable across different count models employed for both the manufacturing and the services sector. Access to telecommunications seems to matter positively for manufacturing and services firms. Economic geography variables such as industrial diversity and market access also affect the count of new informal services firms positively, but have no significant effect on that for informal manufacturing firms. The size (i.e. population) of the district also makes a location more attractive to informal activity, more so for services than for manufacturing.

2.6 Endogeneity Issues, Robustness and Other Exercises

Although all the regressors have been lagged, there could remain endogeneity concerns that would 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 activities, firms will be attracted to that location. 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. How to isolate the effect that runs from agglomeration to performance thus represents a considerable challenge. With regard to the proposed analysis, the presence of these unobservable sources of a location's natural advantage complicates the estimation procedure, particularly in identifying the contribution of production externalities to the location decision of firms.

Ellison and Glaeser (1997) point out that the effects of unobservable sources of 'natural advantage' will not be separately identified from those of production externalities between firms that arise simply from firms locating near one another. Simply including the number of firms or employment in a particular industry, which is a commonly used indicator in empirical studies evaluating localisation economies, will not be able 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. As it is impossible to get data on all the factors relevant to a firm's location decision, it would be helpful to control for any unobservable factors.

As I am unable to find an instrument that predicts the decision of the firm only through its effect on agglomeration and that is exogenous, I introduce district fixed-effects in my estimation. Location fixed effects successfully controls 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 geography of the district. Since I am mainly interested in the effect of agglomeration on the decision of firms, these specifications do not include variables pertaining to the business environment. Additionally, as the business environment variables don't vary by both industry and district, these will not be identified with the introduction of both sets of fixed effects.

The results for both manufacturing and services sectors are presented in Table 2.5. Columns (1) and (4) carry out the regressions without any controls, columns (2) and (5) introduce district fixed-effects, and columns (3) and (6) include district fixed-effects, and finally, columns (4) and (8) include both district and NIC 2-digit industry fixed effects.

It is interesting to observe that as the controls get more stringent the coefficients for the agglomeration variables behave differently from one specification to the next. Just as earlier in the paper, the introduction of industry fixed-effects underlies the reason behind the change. Controlling for district-level unobservables affects the magnitude of localisation and input-output only marginally, whilst the direction remains unchanged. Similar results are observed for the accessibility indicator. This provides evidence that the unobserved characteristics of the location might not have been driving the results reported earlier.

In summary, after controlling for the effect of unobservable characteristics of the location, and for the industry, intra-industry clustering seems to have a strong positive effect on attracting informal firms in both the manufacturing and services sectors. Both, informal services and manufacturing are drawn to those they sell to, while informal manufacturing seems to be drawn to those they buy from as well. For both sectors, wherever the coefficients are identified, market access, industrial diversity and the size of the district continue to be important. I also re-run these regressions within an ordinary least squares model, with log of non-zero counts of firm births as the dependent variable. The results are presented in Table 2.A.11.

Table 2.5: Controlling for Unobservables (Negative Binomial)

	Manufacturing				Services			
Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Localisation	-0.3739*	0.3320***	-0.4531***	0.2080**	-0.3389***	-0.4116*	-0.2609***	0.3247*
	[0.544]	[0.542]	[0.026]	[0.089]	[0.010]	[0.150]	[0.090]	[0.199]
Input	1.0012***	0.0572	0.4533***	0.4852***	0.0780*	-0.3299***	0.1018***	-0.4177***
	[0.070]	[0.118]	[0.085]	[0.167]	[0.040]	[0.061]	[0.031]	[0.052]
Output	-0.6382***	0.1259	-0.9292***	0.2716***	0.1906***	0.4007***	0.1933***	0.4334***
	[0.065]	[0.098]	[0.075]	[0.139]	[0.033]	[0.044]	[0.023]	[0.040]
Urbanisation	-0.0508	-0.2639***	0.0000	0.0000	-0.3002***	-0.5321***	0.0000	0.0000
	[0.099]	[0.093]	[0.000]	[0.000]	[0.086]	[0.087]	[0.000]	[0.000]
Market Access	0.0970	0.0309	0.0000	0.0000	0.1520	0.0875	0.0000	0.0000
	[0.081]	[0.084]	[0.000]	[0.000]	[0.106]	[0.119]	[0.000]	[0.000]
Population	0.6636***	0.7967***	0.0000	0.0000	0.7385***	0.9699***	0.0000	0.0000
	[0.097]	[0.090]	[0.000]	[0.000]	[0.071]	[0.082]	[0.000]	[0.000]
District FEs	×	×	√	√	×	×	√	√
Industry FEs	×	√	×	√	×	√	×	√
#	3,762	3,762	3,762	3,762	3,796	3,796	3,796	3,796
AIC	29409	29254	28368	28163	33027	32683	31508	31025
BIC	29454	29411	28390	28303	33073	32774	31531	31117

Non-Exponentiated coefficients
Standard errors in square brackets clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

2.6.1 Robustness check

As a robustness check, I carry out the same exercise by differentiating between firms of different sizes. I divide the sample of enterprises into those that are small (i.e. employ less than 5 workers) and large (i.e. they employ more than 5 workers). In the case of unorganised manufacturing, almost 90 per cent of the firms in the sample, thus defined, are small-scale enterprises. For informal services, small-scale enterprises account for 93 per cent of the sample. The sample could also be divided into own-account enterprises (OAE) and establishments. Own-account enterprises do not employ any hired workers on a regular basis, whilst establishment enterprises employ one or more workers on a regular basis. Around 68 per cent of all informal manufacturing, and approximately 70 per cent of all informal services enterprises are own-account enterprises.

When I compare manufacturing firms by their sizes, I make a few interesting observations (see Table 2.6). Localisation economies continue to have a strong positive effect on small-scale, large-scale enterprises and establishments. However, while small-scale and large-scale enterprises seem to be attracted to those they sell to, but not who they buy from, own-account enterprises seem to be attracted to neither. Large-scale manufacturers, on the other hand are drawn to their suppliers. Most importantly, the size of the district, i.e. the population explains an important part of what makes a location attractive to small-scale, large-scale and OAE enterprises.

Some of these results also hold for small-scale or OAE informal services enterprises (see Table 2.7). Localisation has a very strong positive and statistically significant effect small-scale and own account enterprises – implying that new births tend to take place in locations with more existing firms within the same two-digit industry. Similarly, enterprises of all sizes seem to be co-located with those they supply to, while large-scale firms are also drawn to those they buy from. The level of industrial diversity has a positive impact on all establishments, except for large-scale firms, in which case it has a negative but statistically insignificant effect. Access to larger neighbouring markets seems to matter only to smaller enterprises. Lower education is invariably associated with more numerous small-scale enterprises. Wages have no impact, except a negative effect on the largest firms, and the size of the district makes a location more attractive to all kinds of informal services firms, no matter what the size. The results for smaller enterprises remain relatively stable compared to those for that for the full sample, and a few differences emerge for large-scale enterprises only.

Table 2.6: Manufacturing enterprises by size

	Negative Binomial			
	Small-scale	*Large-scale*	*OAE*	*Establishments*
Localisation	0.5844*	0.2102***	0.4459	0.9406*
Input	-0.1907	-0.3800*	-0.9355	0.4475***
Output	0.3469***	0.6748***	0.0596	0.2262**
Urbanisation	-0.0750	-0.2835	-0.0603	-0.2468*
Market Access	0.0596	0.6713**	0.0391	0.0186
Education	0.0084	0.0068	-0.0020	0.0067
Telephone	0.0072**	0.0053	0.0043	0.0153***
Electricity	-0.0050*	0.0092**	-0.0053	-0.0047
Wages	-0.0029	-0.0038	-0.0019	-0.0029
Wealth	0.0189**	0.0087	0.0256	0.0046
Population	0.7252***	0.3339***	0.7588	0.5448***
Industry FEs	✓	✓	✓	✓
#	6,150	2,024	3,557	4,617

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

Table 2.7: Services enterprises by size

	Negative Binomial			
	Small-scale	*Large-scale*	*OAE*	*Establishments*
Localisation	0.4749***	0.8225	0.5452***	-0.4040
Input	-0.2961***	-0.1593	-0.2201***	-0.4544***
Output	0.3513***	0.2674**	0.2901***	0.4202***
Urbanisation	-0.3762***	0.2362	-0.3014**	-0.4337***
Market Access	0.2294**	0.1704	0.2949***	0.0996
Education	-0.0181*	0.0479*	-0.0191	-0.0104
Telephone	0.0189*	0.0321**	0.0194*	0.0160*
Electricity	0.0099***	0.0050	0.0103***	0.0089***
Wages	-0.0015	-0.0002	-0.0014	-0.0018**
Wealth	0.0030	-0.0297*	0.0021	0.0130
Population	1.0062***	0.6142***	1.0074***	0.8849***
Industry FEs	✓	✓	✓	✓
#	5,071	1,515	2,778	3,808

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

In summary, the results are broadly similar to those obtained before, except that the impact of certain factors seems to be stronger for small-scale firms than for larger establishments in the data. In their analysis of Italian firms Lafourcade and Mion (2003) also find that small firms are more spatially concentrated than large ones and are more sensitive to input-output linkages. Additionally, as the data is unable to differentiate between formal and informal services, controlling for the size of the firm provides a reasonable approximation of informality, and excludes large services enterprises that are not formally registered under the Factories Act, but which in all other ways might be run like formal-sector enterprises.

2.6.2 Unorganised versus organised

I also carry out the same exercise for the organised manufacturing and services sector in India, to check how the results differ. I use data for manufacturing firms from the Annual Survey of Industries (ASI) and on services firms from the Prowess database. Prowess is a corporate database that contains normalised data built on a sound understanding of disclosures of over 20,000 companies in India. ASI contains data on over 140,000 manufacturing firms in India. I then re-run the regressions for new firms for the two cross-sections – 1999-2000 for services and 2004-2005 for manufacturing. I also differentiate between firms by size. Since both Prowess and ASI tend to sample mostly large firms, my definition of what constitutes a small firm is one with less than 100 employees. The results are reported in Table 2.8.

Owing to the much smaller sample sizes, the count of new firms by district, industry and size tend to be quite low and there is a preponderance of zeroes, which makes running count models difficult. This is especially the case when industry dummies are being used. As explained by Silva and Tenreyro (2010), count models are less likely to converge in cases when the regressor is zero and the dependent variable is positive, otherwise being non-negative with at least one positive observation. Dummy variables often fit these characteristics, especially when the dummy equals zero for all observations with a positive dependent variable, and having positive value when the dependent variable equals zero. This is very much the case in this dataset when industry-level dummies are introduced into the specification. In a recent paper, Silva and Tenreyro (2011) define a new code within Stata to deal with this problem. The alternatives suggested, however, do not deal very effectively with the problem caused by the dummy variable. This is because their procedure would involve dropping some of the dummies and the zero observations, which would dimply an arbitrary redefinition of the reference category. This strategy, as pointed out by the authors themselves, is unlikely to be a sensible one. Another alternative might be to use generalised linear models that reweight the least squares algorithm to make it more stable[43]. This option does not solve the problems associated with the industry-level dummy, as all the dummies are dropped from the procedure, effectively reducing the specification to one without any industry fixed-effects. In the absence of other options to reach convergence within the count models, I use ordinary least squares (OLS) estimations. The OLS method ignores observations where the count of patents is zero, but on the flip side it is able to account for the unobservables at the level of the industry and the direction and magnitude of the coefficients would help provide some level of confidence in the estimates.

Since I have data on much fewer firms when using the Prowess or the ASI dataset for the organised sector, most of the predictor variables are no longer significant. It is also instructive to compare the results with those obtained by differentiating the unorganised sector by size. For instance, localisation has a positive and significant effect on the rate of new firm births for all formal firms, whether in manufacturing or services, and whether classified as small or large. For the informal manufacturing sector (see Table 2.6), this was also the case, except for very small own-account manufacturing units. The effect of localisation was positive and significant only for small-scale firms and own-account enterprises for informal services (Table 2.7), while the effect for informal large-scale firms was insignificant. This is in contrast to the positive and significant effect of localisation for formal services, whether large or small. This could be evidence that large-scale unorganised sector firms are not the same as those in the formal sector, which consist mostly of finance, insurance, IT firms etc. Such firms also seem to benefit from spillovers when in proximity to one another.

There are some similarities however; for instance, large-scale informal services units are attracted to those they sell to, but not those they buy from, and this result is also true

[43] The authors provide a new command 'ppml' in Stata to carry out this estimation. But this continues to (correctly) drop the district dummy.

for formal services firms. Telecommunications infrastructure attracts both large-scale informal services and manufacturing units, and formal units in both sectors, more so for services than manufacturing. The effect of population is positive for formal services, but negative (albeit statistically insignificant) for manufacturing units. The latter could be explained by urban regulations that prevent heavy industries from clustering near large population settlements in cities and towns. For informal units, the size of the district tended to have a positive effect. And interestingly, education has a positive effect on formal services firms, and large-scale informal services firms, but it is negatively associated with small-scale informal services firm births.

Table 2.8: Organised Manufacturing and Services (OLS)

	Services (Prowess)		**Manufacturing (ASI)**	
Variable	*Small*	*Large*	*Small*	*Large*
Localisation	0.2881***	0.3374**	0.5219***	0.3201***
Input	-0.0012	-0.0986	-0.2248	-0.2883***
Output	0.0088	0.1778	0.0495	0.1099**
Urbanisation	0.0056	0.0411	-0.1833	-0.3760
Market Access	-0.0038	0.1366	-0.0794	0.2804*
Education	-0.0017	0.0232	0.0360*	0.0348**
Telephone	0.0008*	0.0046	0.0513**	0.0354**
Electricity	0.0001	-0.0049	-0.0003	-0.0012
Wages	-0.0001	0.0006	0.0007	-0.0006
Wealth	0.0026	-0.0102	-0.0397	-0.0360
Population	-0.0018	-0.0631	0.4608**	0.2306*
Industry Fixed Effects	✓	✓	✓	✓
#	3,584	4,849	2,742	2,742
R^2	0.109	0.190	0.177	0.176

Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

One might also expect that the rate of informal activity would be higher in places where there are barriers to entry to formal activity. Dutta et al (2011) find evidence of increased informal activity in states in India that suffer from higher levels of corruption. In other words, it may be possible that informal activity serves as a substitute to the formal sector. If this were the case, one might expect to see a negative correlation between the informal and formal firm births. In the data, I find that both the count and the rate of new firm activity are positively correlated at the geographical level of the state and that of the district. Indeed, Mitra (2009) finds that the incidence of informality is high in more industrialised Indian states, suggesting that informal activity could be complementary to the formal sector. Arimah (2001) provides evidence of linkages between the formal and informal sector in Nigeria in the form of sub-contracting and the flow of consumer goods and raw materials. Thus, I investigate the inter-linkages between the types of sectors that could be driving these correlations.

2.6.3 Measures of co-agglomeration

While the data treats formal and informal manufacturing and services as separate units, in reality these firms are inter-linked in a number of ways. The agglomeration variables (localisation, input, output and industrial diversity) have been constructed taking total employment, i.e. across the formal and informal sector, into account. However, this does not tell us anything about the linkages between and across formal and informal, manufacturing and services firms. Following Ellison and Glaeser (1997, 2010) I compute pair-wise coagglomeration measures for all 2-digit industries for manufacturing and services, across the organised and the unorganised sector (see Appendix A for construction of the Index). I have at my disposal data from four different sources: organised manufacturing data comes from the Annual Survey of Industries, unorganised manufacturing and services data comes from two different surveys of the National Sample Survey Organisation, and organised services data comes from the Prowess database.

Clearly, the Prowess database contains very few observations as compared to data from the NSSO and the ASI. I use the Annual Survey of Industries instead of Prowess for manufacturing firms, as the former is a richer source of data, even though the latter also contains data on manufacturing units. Since Prowess accounts for such a small proportion of firms in the sample, using this database gives an inflated value of coagglomeration. In other words, owing to the small size of these sectors when data for total employment is pooled the small number of firms in the dataset causes the coagglomeration index to be very volatile. Thus, I drop data from Prowess, and construct coagglomeration measures using the remaining databases. Subsequently, I am unable to construct coagglomeration measures for formal services.

Table 2.A.13 lists the 20 most coagglomerated sectors. Similar to the EG agglomeration index, the no-coagglomeration benchmark is when the value of the index is zero (i.e.

$E(\gamma)=0$). In general, if the EG coagglomeration index is greater than 0.05, the industries are considered to be highly concentrated.

Certain coaggomerations, such as office and computing maintenance and market research activities with education, i.e. primary, secondary, distance learning education activities, seem intuitive – one might expect these industries to use similar labour pools. However, others, such as the coagglomeration of manufactures of apparel with education, or that of recreational and entertainment activities with recycling, is not clear.

Earlier results found that linkages to buyers explained a large proportion of new informal activity within a district. I will now verify to what extent these linkages are correlated with the final coagglomeration indices observed in my data. Whilst the earlier analysis made no distinction between organised and unorganised industries, this analysis teases out the importance of each type of activity (i.e. formal or informal) for each type of industry (i.e. manufacturing and services). To relate the measure of coagglomeration to a single measure of linkages between a pair of industries, I follow Ellison et al (2010) and construct an input-output index (see Appendix A for construction of the Index). I then relate this single measure for each pair of industry to the coagglomeration measure also constructed for each pair on industry – except that the latter are also constructed separately for formal and informal manufacturing. The table below provides the correlation values for each pair of coagglomerated industries with the standard input-output index.

Since I do not have data on labour market pooling and knowledge spillovers, in this section I try to discern the effect of input-output spillovers only. A major limitation of the EG index is that it does not distinguish between spillovers and natural advantages to explain the coagglomeration of firms. Thus, I will be unable to single out the effect of buyer-supplier linkages from that of natural advantage. A high correlation may be an indication that the pair of industries are coagglomerated owing to input-output linkages, while a low correlation may be an indication that other factors, such as say, labour market pooling or technological spillovers underlie the observed coagglomeration.

I find that although coagglomeration and input-output linkages are positively associated, the level of correlation is quite low. Coagglomeration between formal manufacturing and formal and informal manufacturing and services does seem to have some correlation with the standard input-output measures, perhaps indicating that these buyer-supplier linkages may explain the coagglomeration to some extent. Interestingly, the standard input-output measure is negatively associated with the coagglomeration of formal services with itself – implying that other linkages may be more important. The same outcome is true for coagglomeration of informal services.

It could also be argued that input-output linkages and coagglomeration are endogenous – in other words, firms may use the outputs of (or sell to) particular sectors simply because these sectors are coagglomerated. If it is assumed that input-output linkages are determined by given production technologies and that the national input-output vectors are representative at the local scale, then I can rule out scenarios in which firms would adjust their inputs or outputs according to what was locally available. If this were true, I would also expect to find a higher correlation between my measures of input-output linkages and coagglomeration.

The results for input and output linkages to explain the attractiveness of a location to informal manufacturing activity was significant and positive – in other words, being located closer to buyers or suppliers made a location more attractive to new units. The coagglomeration exercise conducted above shows that input-output linkages are in fact positively correlated with the EG measure of coagglomeration, which is what I would expect in light of my earlier results. Similarly, with regards to informal services, although output linkages made a location more attractive to new informal services units, input linkages had a negative effect. The standard input-output measure in the above analysis is an un-directional measure of the input and output variables and thus it could be capturing the negative effect of output linkages found in the earlier regression analysis.

2.7 Conclusion

This paper seeks answers to the following question: What factors influence the spatial distribution of informal economic activity within India? The main aim of the paper is to understand what drives the process of spatial variations in informal industrial activity, i.e. in identifying the factors that determine location decisions. It is important to understand why economic activity tends to concentrate geographically because if one can explain geographic concentration, then one can go some way towards explaining important aspects of international trade and economic growth. The importance of this research is underscored by two inter-related factors – that the clustering of economic activity has important implications for economic development and that the contribution of the informal sector to economic growth and employment makes it a potent tool in influencing regional economic policy.

The empirical analysis finds that economic geography factors have an important effect on informal firms' performance, and thus their decision to locate in a particular area. In the case of formal manufacturing in India, Lall and Mengistae (2005) find that there is a pattern in the data whereby geographically disadvantaged cities seem to compensate partially for their natural disadvantage by having a better business environment than more geographically advantaged locations. The findings in this paper are that economic geography factors, such as localisation and input-output economies, do in fact positively impact the attractiveness of a district to new informal activity, whilst industrial diversity and access to markets is more important for services. The analysis finds that the presence of telecommunications infrastructure seems to matter, especially to services units and larger manufacturing units, whilst the size of the district is also important. This is an indication that governments may be somewhat limited in their ability to narrow regional disparities in hosting of informal economic activity, which is an important source of growth and employment.

This research also makes an important contribution to the empirical literature on industrial development and economic geography. There is little or no research on the factors driving the location of informal industry, although a handful of papers study the effects of agglomeration economies and business environment on the spatial concentration of manufacturing in emerging countries. In large developing countries the informal sector accounts for an important proportion of domestic product and employment, and any study that does not account for the sector is scarcely representative. In addition, whilst the theoretical development of new economic geography has received much attention in the literature, there is still much scarcity of empirical tests for developing countries. The available evidence on the functioning of small and poor enterprises is incomplete in many important ways – using recent data complied by national sample surveys this paper provides a first glimpse into the extent to which locational attributes drive their incentives.

In addition I also use district and industry fixed effects to rule out omitted variables bias by controlling for the difference between first and second nature economic geography. In summary, this paper provides evidence of the validity of the forces emphasised by new economic geography and location theory approaches. The study does not attempt to perfect the theory of economic geography, but it does attempt to confront the existing tenets with data on unorganised industry in India.

The policy implications of the research and its findings are of significant importance – policy-makers need to have an understanding of the relative importance of existing agglomeration economies and business environment if they are interested in influencing the decisions of informal activity. With the importance of this sector and its potential effect on employment and economic growth, such an understanding could provide a powerful tool for spreading growth and employment to geographically less-advantaged regions. This analysis finds that governments may find it an uphill task to encourage informal economic activity to locate to regions that it has not previously favoured.

References

Amiti, M. and Javorcki, B.S. (2005) ‘Trade costs and location of foreign firms in China’, Working Paper No. 55, IMF.

Arimah, B.C. (2001) ‘Nature and determinants of linkages between informal and formal sector enterprises in Nigeria’, African Development Review, 13(1): 114-144.

Arrow, K.J. (1962) ‘The economic implications of learning by doing’, Review of Economic Studies, 29: 155-173.

Bagchi-Sen, S. (1995) 'FDI in US Producer Services: A Temporal Analysis of Foreign Direct Investment in the Finance, Insurance and Real Estate Sectors', Regional Studies, 29:2, 159 – 170.

Brakman, S., Garretson, H., Gorter, J., van der Horst, A. and Schramm, M. (2005) ‘New Economic Geography, empirics and regional policy’, Working Paper No. 56, CPB Netherlands Bureau for Economic Policy Analysis.

Brulhart, M. (1998) ‘Economic Geography, industry location and trade: The evidence’, The World Economy, 21 (6): 775-801.

Carlton, D. (1979) ‘Why new firms locate where they do: an econometric model’, in W. Wheaton (ed.), Interregional movements and regional growth, Washington DC.

Carlton, D. (1983) ‘The location and employment choices of new firms: an econometric model with discrete and continuous endogenous variables’, Review of Economics and Statistics, 65 (3): 440-449.

Carod, J.M.A. and Antolín, M.C.M. (2004) ‘Firm size and geographical aggregation: An empirical appraisal in industrial location’, Small Business Economics 22 (3-4): 299-312.

Cheng, L.K. and Kwan, Y.K. (2000) ‘What are the determinants of the location of foreign direct investment? The Chinese experience’, Journal of International Economics, 52: 379-400.

Chinitz, B. (1961) ‘Contrasts in Agglomeration: New York and Pittsburgh’, American Economic Review, 51: 279-289.

Coffey. W.J. and Shearmur, R.G. (2002) ‘Agglomeration and dispersion of high-order service employment in the Montreal metropolitan region 1981-96’, Urban Studies, 39(3): 359-378.

Coughlin, C.C. and Segev, E. (2000) ‘Location determinants of new foreign-owned manufacturing plants’, Journal of Regional Science 40 (2): 323-351.

Deichmann, U., Kaiser, K., Lall, S.V. and Shalizi, Z. (2005) 'Agglomeration, transport and regional development in Indonesia', Policy Research Working Paper No. 3477, World Bank

Dekle, R. and Eaton, J. (1999) 'Agglomeration and land rents: evidence from the prefectures', Journal of Urban Economics, 46: 2001-2014.

Duranton, G. and Overman, H.G. (2005) 'Testing for localization using micro-geographic data', Review of Economic Studies, 72(4): 1077-1006.

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

Ellison, G. and Glaeser, G.L. (1997) 'Geographic Concentration of U.S. Manufacturing Industries: A Dartboard Approach', Journal of Political Economy, 105 (5): 889–927.

Enderwick, P. (1989) Multinational Service Firms, Routledge, London

Feser, E.J. and Bergman, E.M. (2000) 'National industry cluster templates: a framework for applied regional cluster analysis', Regional Studies, 34: 1-20.

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

Guimaraes, P., Figueiredo, O. and Woodward, D. (2000) 'Agglomeration and the Location of Foreign Direct Investment in Portugal', Journal of Urban Economics, 47 (1): 115-135.

Guimaraes, P., Figueiredo, O. and Woodward, D. (2003) 'A Tractable Approach to the Firm Location Decision Problem', Review of Economics and Statistics, 85 (1): 201-204.

Guimaraes, P., Figueiredo, O. and Woodward, D. (2004) 'Industrial Location Modeling: Extending the Random Utility Framework', Journal of Regional Science, 44 (1): 1-20.

Guimaraes, P. and Lindrooth, R.C. (2007) 'Controlling for overdispersion in grouped conditional logit models: A computationally simple application of Dirichlet-multinomial regression', The Econometrics Journal 10 (2): 439-452.

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

Harris, J.R. and Todaro, M.P. (1970) 'Migration, unemployment and development: a two-sector analysis', American Economic Review, 61(1): 126-142.

Head, K. and Mayer, T. (2004) 'Market potential and the location of Japanese investment in the European Union', Review of Economics and Statistics, 86(4): 959-972.

Head, H. and Reis, J. (1996) 'Inter-city competition for foreign investment: Static and dynamic effects of China's Incentive', Journal of Urban Economics, 40: 38-60.

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

Holl, A. (2004a) 'Manufacturing location and impacts of road transport infrastructure: empirical evidence from Spain', Regional Science and Urban Economics 34 (3): 341-363.

Holl, A. (2004b) 'Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from Portugal', Journal of Regional Science 44 (4): 693-712.

Holl, A. (2004c) 'Start-ups and relocations: Manufacturing plant location in Portugal', Papers in Regional Science 83 (4): 649-668.

Jacobs, J. (1969) The Economy of Cities. Cambridge: MIT Press.

Kirn, T.J. (1987) 'Growth and change in the service sector of the United States: a spatial perspective', Annals of the Association of American Geography, 77: 353-372.

Krugman, P. (1991) 'Increasing returns and economic geography', Journal of Political Economy, 99: 483-499.

La Porta, R. and Shleifer, A. (2011) 'The unofficial economy in Africa', NBER Working Paper No. 16821.

Lafourcade, M. and 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. and Chakravorty, S. (2003) 'Diversity Matters: The Economic Geography of Industrial Location in India', Policy Research Working Paper 3072, World Bank

Lall, S.V. and Chakravorty, S. (2005) 'Industrial Location and spatial inequality: Theory and Evidence from India', Review of Development Economics, 9 (1): 47-68.

Lall, S.V., Shalizi, Z. and Deichmann, U. (2004) 'Agglomeration economies and productivity in Indian industry', Journal of Development Economics, 73 (2): 643-673.

Lall, S.V. and 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. and Kar, S. (2009) 'A contemporary perspective on the informal labour market: theory, policy and the Indian experience', Economic and Political Weekly, Vol. XLIV, No. 14.

Marshall, A. (1890) Principles of Economics. London: Macmillan

Marshall, A. (1919) Industry and Trade. London: Macmillan

McFadden, D. (1974) 'Conditional logit analysis of qualitative behaviour', in P. Zarembka (ed.) Frontiers in Econometrics, Press: New York (Academic Press): 105-142

McGee, T.G. (1977) 'Hawkers in South East Asia Cities: Planning for a Bazaar Economy', International Development Research Centre, Ottawa.

Mitra, A. (2009) 'Impact of trade on service sector employment in India', Institute of Economic Growth, Working Paper Series No. E/298.

Mukim, M. and Nunnenkamp, P. (2012) 'The location choices of foreign investors: a district-level analysis in India', The World Economy (Forthcoming). Also available as Kiel Working Paper No. 1628.

Papke, L.E. (1991) 'Interstate Business Tax Differentials and New Firm Location - Evidence from Panel Data', Journal of Public Economics 45 (1): 47-68.

Rauch, J. E. (1992) 'Productivity gains from geographic concentration of human capital: Evidence from the cities', Journal of Urban Economics, 34, 380–400.

Romer, P.M. (1986) 'Increasing returns and long-run growth', Journal of Political Economy, 94 (5): 1002-1037.

Sakhthivel, S. and Joddar, P. (2006) 'Unorganised sector workforce in India: trends, patterns and social security coverage', Economic and Political Weekly, May 27.

Tiglao, N.C. and Tsutsumi, M. (2005) 'Modeling households and location choices in metro Manila using spatial microsimulation approach', Proceedings of the Easter Asia Society for Transportation Studies, 5: 2719-2194.

Venables, A.J. (1996) 'Equilibrium locations of vertically linked industries', International Economic Review, 49: 341-359.

Wu, F.L. (1999) 'Intrametropolitan FDI firm location in Guangzhou, China - A Poisson and negative binomial analysis', Annals of Regional Science 33 (4): 535-555.

Appendix 2.A: Indices

Ellison and Glaeser (1997) Index:

The EG Index for industry k is equal to:

$$\gamma_k = \frac{G - \left(1 - \sum_j x_j^2\right) H_k}{\left(1 - \sum_j x_j^2\right)\left(1 - H_k\right)}$$

where G for industry k is defined as:

$$G = \left(s_j^k - x_j\right)^2$$

and s refers to the share of total employment of district j for industry k, x refers to the share of district j in total employment, and H is the plant employment Herfindahl index, corresponding to the sum of the squares of the share of employment of each plant over the total employment of the industry.

Theil Index:

The Theil index for specialisation here measures the extent of over or under representation of a district with regards to employment across a set of industries. The value of the index is[44]:

$$T_r = \sum_k \frac{x_{jk}}{x_j}\left(\log\frac{x_{jk}}{x_j} - \log\frac{x_k}{x}\right)$$

where:

x_{jk} refers to employment in industry k in district j
x_j refers to total employment in district j
x_k refers to total employment in industry k
x refers to total employment

Ellison and Glaeser (2010) Coagglomeration Index:

The EG coagglomeration index applies to industry pairs, and for industries i and j it is defined as:

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

[44] See Brakman et al (2005) for more on the calculation of the index for concentration.

where m indexes geographic areas (here, districts), s_{mi} is the share of industry i's employment contained in area m, x_m measures the aggregate size of area m (which is modelled as the mean employment share in the district across manufacturing/services industries).

Ellison and Glaeser (2010) Input-Output Index:

$Input_{i \leftarrow j}$ is defined at the share of industry i's inputs that come from industry j.
$Output_{i \rightarrow j}$ is defined as the share of industry i's outputs that are sold to industry j.
To construct a proxy for the linkages between a pair of industries, I follow Ellison et al (2010) and define unidirectional versions of the input and output variables by:
$Input_{ij} = \max\left\{Input_{i \leftarrow j}, Input_{j \leftarrow i}\right\}$ and
$Output_{ij} = \max\left\{Output_{i \rightarrow j}, Output_{j \rightarrow i}\right\}$

The combined variable is then defined as:

$$InputOutput_{ij} = \max\left\{Input_{ij}, Output_{ij}\right\}$$

Tables

Table 2.A.1: Share of unorganised activity (2002-03)

Industry	Organised (% of NDP)	Unorganised (% of NDP)	Total
Agriculture, forestry, fishing	4.1	95.9	100
Mining, manufacturing, electricity and construction	60.5	39.5	100
Services	53.1	46.9	100
Total	**43.3**	**56.7**	**100**

Source: National Account Statistics 2005

Table 2.A.2: Distribution of Employment (2004-2005)

		Number of workers (millions)	**Distribution of workers (%)**
Agriculture	*Organised*	6.1	2.4
	Unorganised	252.8	97.6
		258.9	***100***
Non-Agriculture	*Organised*	56.5	28.4
	Unorganised	142.1	71.6
		198.5	***100***
Total	*Organised*	62.6	13.7
	Unorganised	394.9	86.3
		457.5	***100***

Source: NSSO Sample Survey 2004-2005

Table 2.A.3: Employment by sector (%)

Industry	1983-84		1987-88		1993-94		1999-2000	
	Org	*Unorg*	*Org*	*Unorg*	*Org*	*Unorg*	*Org*	*Unorg*
Agriculture, forestry and fishing	0.6	99.4	0.7	99.3	0.6	99.4	0.6	99.4
Mining and quarrying	55.5	44.5	44.2	55.8	40.7	59.3	43.2	56.8
Manufacturing	19.7	80.3	17.3	82.7	16.1	83.9	14.9	85.1
Electricity, gas and water	90.7	9.3	71.3	28.7	69.7	30.3	79.0	21.0
Construction	17.7	82.3	10.1	89.9	10	90	6.5	93.5
Trade, hotels and restaurants	2.1	97.9	1.8	98.2	1.6	98.4	1.2	98.8
Transport, storage and communication	38.8	61.2	34.8	65.2	29.7	70.3	21.5	78.5
Services	40.3	59.7	36.8	63.2	31.7	68.3	34.8	65.2

Source: Sakhtivel and Joddar 2006[45]

[45] Organised employment figures are obtained from annual reports (1983 and 1988) and Quarterly Employment Review (1994 and 2000).

Table 2.A.4: Contributions to the Theil Index (Unorganised sector)

District	Manu	District	Serv
Mumbai	255.43	Kolkata	984.42
Ludhiana	146.34	Mumbai	958.80
South Tripura	100.84	Delhi	361.93
Kolkata	80.03	Purba Champaran	248.53
Delhi	52.53	Medinipur	226.19
Ahmadabad	47.11	Ernakulam	175.90
Jaipur	44.08	Pune	169.70
South 24 Parganas	43.08	Thane	161.71
Coimbatore	42.63	Bangalore	139.19
West Tripura	42.19	Hyderabad	137.65
Surat	39.93	Lucknow	131.88
Thane	39.70	Kanpur Nagar	128.59
North 24 Parganas	39.52	West Tripura	104.66
Haora	37.08	South 24 Parganas	99.99
Murshidabad	36.44	Jammu	96.08
Srinagar	34.17	Thiruvananthapuram	95.27
Hyderabad	34.00	Madurai	92.67
Varanasi	32.53	West Godavari	90.62
Virudhunagar	31.18	North 24 Parganas	90.12
Vellore	29.69	Barddhaman	86.76

Table 2.A.5: Contributions to the Theil Index (Organised sector)

District	Manu	District	Serv
Bangalore Urban	549.39	Mumbai city	147.16
Mumbai city	433.85	Delhi	65.41
Coimbatore	334.99	Kolkata	35.57
Vellore	246.75	Bangalore Urban	28.74
Ludhiana	180.36	Chennai	22.01
Pune	163.48	Hyderabad	21.49
Thane	143.64	Pune	14.28
Pudukkottai	129.05	Gandhinagar	7.20
Delhi	117.97	Kottayam	4.76
Surat	110.00	Vadodara	4.26
Ahmadabad	96.51	Dharmapuri	4.10
Chennai	93.01	Ludhiana	3.75
Thiruvallur	84.87	Kanpur	3.59
Guntur	83.19	Bhilwara	3.18
Kollam	81.97	South 24 Parganas	3.02
Nizamabad	78.29	Faridabad	2.76
Gurgaon	76.82	Ranchi	2.44
Gautam Buddha Nagar	75.07	Ahmadabad	2.26
Daman	68.29	Coimbatore	2.22
Rangareddi	68.27	Thanjavur	2.10

Table 2.A.6: Ellison-Glaeser Index (Unorganised Manufacturing)

NIC	Description	EG Index
30	Office, accounting and computing machinery	0.204
35	Other transport equipment	0.105
32	Radio, television and communications equipment	0.069
33	Medical, precision and optical instruments, watches and clocks	0.045
19	Tanning and dressing of leather; manufacture of luggage, handbags saddlery, harness and footwear	0.023
31	Electrical machinery and apparatus	0.021
34	Motor vehicles, trailers and semi-trailers	0.017
23	Coke, refined petroleum and nuclear fuel	0.016
27	Basic metals	0.013
16	Tobacco Products	0.012
29	Machinery and equipment	0.010
24	Chemical and chemical products	0.010
25	Rubber and plastic products	0.009
21	Paper and Paper products	0.008
22	Publishing, printing and reproduction of recorded media	0.008
17	Textiles	0.007
26	Other non-metallic mineral products	0.006
36	Furniture	0.004
20	Wood and cork products (except furniture)	0.003
28	Fabricated metal products (except machinery and equipments)	0.003
18	Wearing apparel; Dressing and dyeing of fur	0.002
15	Food products and Beverages	-0.007

Table 2.A.7: Ellison-Glaeser Index (Unorganised Services)

NIC	Description	EG Index
73	Research and development	0.287
61	Water transport	0.206
72	Computer and related activities	0.099
63	Supporting and auxilliary transport activities; activities of travel agencies	0.015
90	Sewage and refuse disposal, sanitation and similar activities	0.013
70	Real estate activities	0.005
91	Activities of membership organisations	0.004
71	Renting of machinery and equipment without operator and of personal and household goods	0.003
74	Other business activities	0.003
60	Land transport; transport via pipelines	0.002
80	Education	0.002
93	Other service activities	0.002
85	Health and social work	0.001
92	Recreational, cultural and sporting activities	0.001
55	Hotels and restaurants	0.000
64	Post and communications	0.000

Table 2.A.8: Characteristics of the Data

	Services			Manufacturing		
Variable	#	Mean	Std. Dev.	#	Mean	Std. Dev.
count	586	4111.27	6749.53	578	3531.38	8207.68
count>0	557	4325.32	6856.00	526	3880.49	8525.32
Obs	586	0.0495	0.2171	578	0.0900	0.2864
Fit_p	480	0.0000	0.0000	570	0.0000	0.0000
Fit_nb	480	0.0066	0.0025	570	0.0325	0.0227

Table 2.A.9: Manufacturing (Industry-Level Results)

NIC	NIC Sector Name	*Localisation*	*Input*	*Output*	*Urbanisation*	#
15	Manufacture of food products and beverages	0.3112	0.2826***		0.0693	488
16	Manufacture of tobacco products	-0.7287	0.5256*		-0.4798	123
17	Manufacture of textiles	-0.0018		0.7411***	0.3764	270
18	Manufacture of wearing apparel; dressing and dyeing of fur	-0.2071		0.5759***	-0.3690***	505
19	Tanning and dressing of leather; manufacture of luggage, handbags and footwear	0.2212**		-0.1546	-1.6314***	75
20	Manufacture of wood and of products of wood and cork, except furniture	0.8834		0.1563	-0.5190**	469
21	Manufacture of paper and paper products	-0.0618		0.4962	0.0163	65
22	Publishing, printing and reproduction of recorded media	0.9168		0.4342**	-0.5608*	146
24	Manufacture of chemicals and chemical products	-0.3427	0.6917***		-0.0423	114
25	Manufacture of rubber and plastic products	0.9815	0.1277		-0.0682***	80
26	Manufacture of other non-metallic mineral products	0.2504*		0.1588	-0.2350	353
27	Manufacture of basic metals	0.2532		-0.1331	-0.0777*	52
28	Manufacture of fabricated metal products, except machinery and equipments	0.1397	0.3423**		-0.7306***	344
29	Manufacture of machinery and equipment N.E.C.	0.9254	0.1116		-0.4328	146
30	Manufacture of office, accounting and computing machinery	-0.6519***	2.0266***		0.0687***	3
31	Manufacture of electrical machinery and apparatus N.E.C.	0.0249		0.1957	-0.5699*	98
32	Manufacture of radio, television and communication equipment and apparatus	-0.7742***	0.6258***		-0.0679*	12
33	Manufacture of medical, precision and optical instruments, watches and clocks	-0.3538***	1.1495		0.7871	11
35	Manufacture of other transport equipment	0.1419**		-1.0813*	-0.2271***	29
36	Manufacture of furniture; manufacturing N.E.C.	0.0241**	0.3259	-0.0835	-0.7527***	378

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.A.10: Services (Industry-Level Results)

NIC	NIC Sector Name	*Localisation*	*Input*	*Output*	*Urbanisation*	#
55	Hotels and restaurants	0.0310	0.3197***		-0.1743	406
60	Land transport; transport via pipelines	-0.4672		0.6307***	0.0072	453
61	Water transport	-0.7715		1.5657*	-2.6247	14
63	Supporting and auxilliary transport activities; activities of travel agencies	0.1716	0.0452		-1.3250***	114
64	Post and communications	0.5077		0.3738***	-0.5119***	284
70	Real estate activities	-0.7186		0.1103	-1.8636***	68
71	Renting of machinery, equipment without operator and of personal, household goods	0.5128		-0.2383	-0.1267	138
74	Other business activities	0.2924**	0.1865**		-0.5220***	322
85	Health and social work	-0.9580		0.3553**	-0.5517***	394
93	Other service activities	-0.7313		0.3594***	-0.6486***	421

Non-Exponentiated coefficients
Standard errors (not reported) clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

Table 2.A.11: Controlling for Unobservables (OLS)

	Manufacturing				Services			
Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Localisation	-0.9610**	0.7426***	-0.4041**	0.6241**	-0.1488**	-0.9123	-0.7828***	0.6804
	[3.563]	[3.809]	[4.927]	[5.212]	[11.731]	[13.529]	[8.691]	[4.041]
Input	0.7269***	0.1860***	0.7895***	1.0023***	0.0585*	-0.2908***	0.0965***	-0.4193***
	[0.051]	[0.072]	[0.064]	[0.082]	[0.032]	[0.070]	[0.033]	[0.051]
Output	-0.4012***	-0.0167	-0.4234***	-0.8504***	0.1976***	0.4226***	0.1886***	0.4574***
	[0.048]	[0.053]	[0.060]	[0.054]	[0.021]	[0.041]	[0.023]	[0.037]
Urbanisation	0.0614	-0.1582**	0.0000	0.0000	-0.3970***	-0.5253***	0.0000	0.0000
	[0.075]	[0.077]	[0.000]	[0.000]	[0.104]	[0.125]	[0.000]	[0.000]
Market Access	-0.0441	-0.0945	0.0000	0.0000	-0.0141	-0.0519	0.0000	0.0000
	[0.071]	[0.076]	[0.000]	[0.000]	[0.108]	[0.117]	[0.000]	[0.000]
Population	0.5781***	0.7765***	0.0000	0.0000	0.6224***	0.7585***	0.0000	0.0000
	[0.075]	[0.070]	[0.000]	[0.000]	[0.095]	[0.123]	[0.000]	[0.000]
District FEs	×	×	✓	✓	×	×	✓	✓
Industry FEs	×	✓	×	✓	×	✓	×	✓
#	2,078	2,078	2,098	2,098	2,299	2,299	2,319	2,319
R^2	0.185	0.308	0.446	0.558	0.228	0.310	0.559	0.665

Non-Exponentiated coefficients
Standard errors in square brackets clustered at the district level
*** p<0.01, ** p<0.05, * p<0.1

Table 2.A.12: Industry Data Sources

Type	Source	Frequency	Percent	Cumulative
Organised	ASI	40694	8.42	8.42
Organised	Prowess (manufacturing)	684	0.14	8.56
Organised	Prowess (services)	367	0.08	8.64
Unorganised	NSSO (manufacturing)	80591	16.67	25.31
Unorganised	NSSO (Services)	361040	74.69	100
Total		483376	100	

Table 2.A.13: Most Coagglomerated Industries

Industry1	Type*	Industry2	Type	Coagg index
Apparel and fur	Or	Education	Unor	**0.2321**
Repair/Maintenance of office and computing equipment	Unor	Education	Unor	**0.1715**
Education	Unor	Market research, consulting, bookkeeping etc	Or	**0.1429**
Recreation, motion picture, TV, radio activities	Or	Recycling	Or	**0.1274**
Medical, precision and optical instruments	Unor	Repair/Maintenance of office and computing equipment	Or	**0.1009**
Apparel and fur	Or	Repair/Maintenance of office and computing equipment	Unor	**0.0791**
Apparel and fur	Or	Market research, consulting, bookkeeping etc	Or	**0.0669**
R&D	Unor	Market research, consulting, bookkeeping etc	Or	**0.0611**
Sewage and refuse disposal, sanitation	Or	Leather	Or	**0.0574**
Office, accounting and computing equipment	Unor	Market research, consulting, bookkeeping etc	Or	**0.0523**
Coke and refined petroleum	Or	Collection, purification distribution of water	Or	**0.0511**
Repair/Maintenance of office and computing equipment	Unor	Market research, consulting, bookkeeping etc	Or	**0.0509**
Radio, TV, Communication Equipment	Or	Market research, consulting, bookkeeping etc	Or	**0.0464**
Auxiliary transport, storage and warehousing	Unor	Auxilliary transport, storage and warehousing	Or	**0.0445**
Furniture, jewellary, musical instruments etc	Or	Market research, consulting, bookkeeping etc	Or	**0.0427**
Repair/Maintenance of office and computing equipment	Unor	Market research, consulting, bookkeeping etc	Or	**0.0417**
Sea, coastal, inland water transport	Unor	Leather	Unor	**0.0390**
Furniture, jewellary, musical instruments etc	Or	Radio, TV, Communication Equipment	Unor	**0.0387**
Market research, consulting, bookkeeping etc	Or	Office, accounting and computing equipment	Or	**0.0380**
Market research, consulting, bookkeeping etc	Or	Medical, precision and optical instruments	Or	**0.0377**

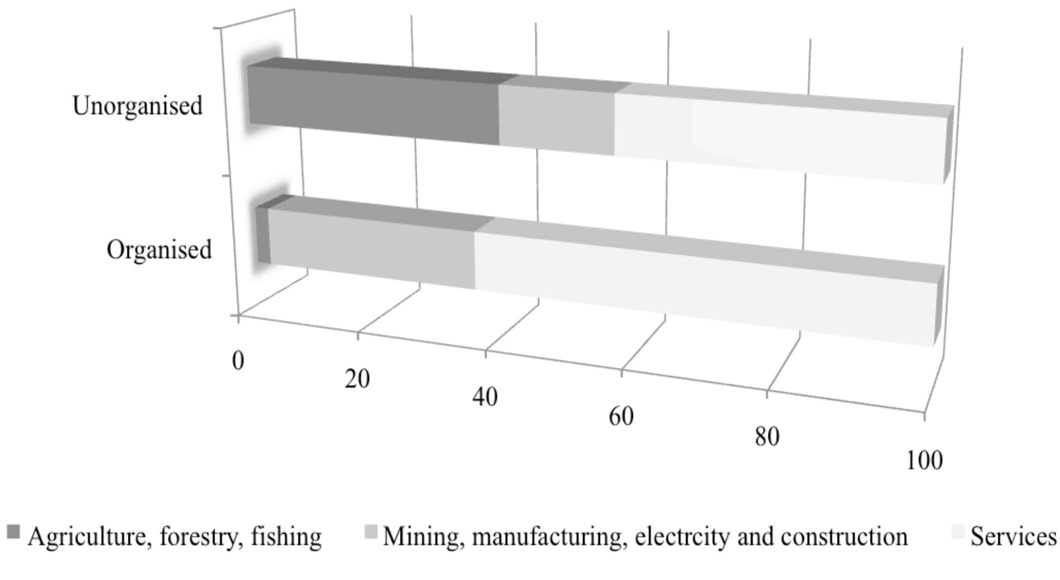
*Type refers to the organised (Or) or unorganised (Unor) sector

Table 2.A.14: Coagglomeration and input-output correlations

Industry 1	Industry 2	Correlation Index
Formal Manufacturing	Formal Manufacturing	0.0531
Formal Manufacturing	Formal Services	0.0688
Formal Manufacturing	Informal Manufacturing	0.0536
Formal Manufacturing	Informal Services	0.0529
Formal Services	Formal Services	-0.0382
Formal Services	Informal Manufacturing	0.0771
Formal Services	Informal Services	0.0175
Informal Manufacturing	Informal Manufacturing	0.0314
Informal Services	Informal Services	-0.0271
Informal Manufacturing	Informal Services	0.0502

Figures

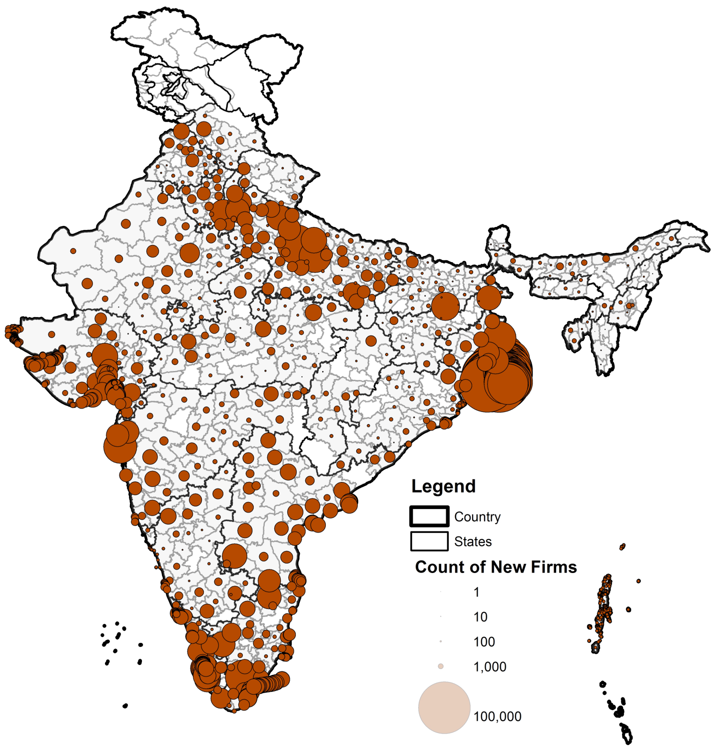
Figure 2.A.1: Share of activity as a % of sectoral NDP (2002-03)

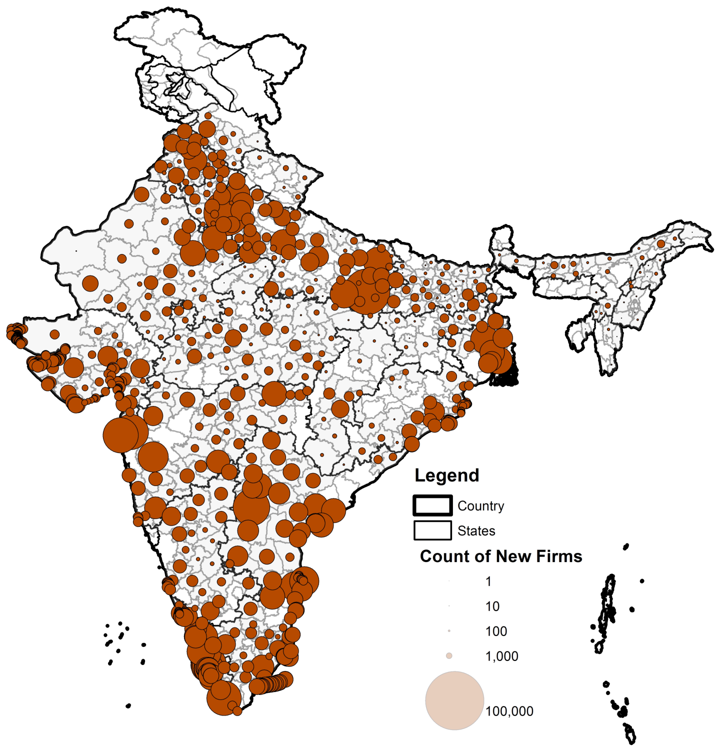


Source: National Account Statistics 2005

Figure 2.A.2: Distribution of Informal Activity

Manufacturing **Services**

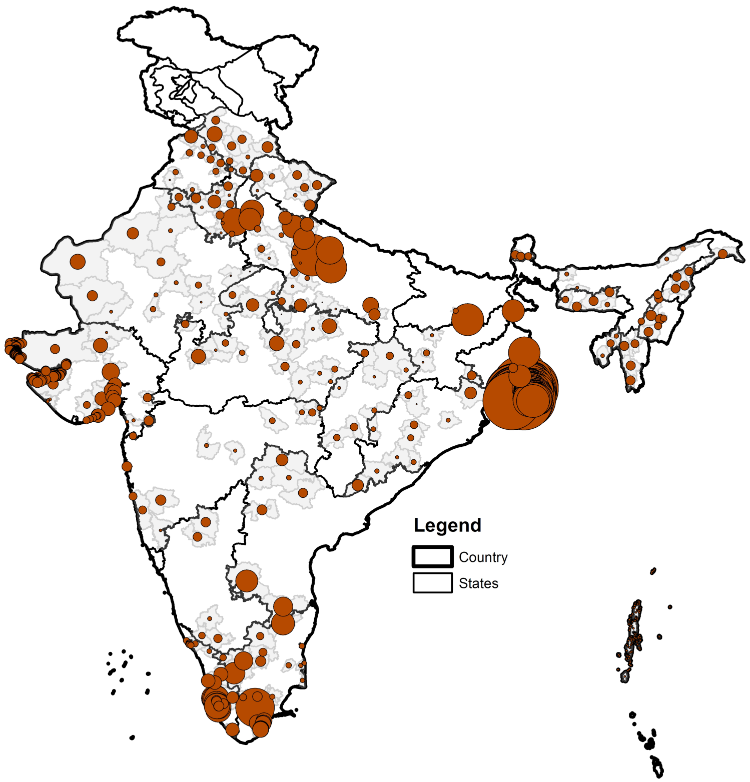


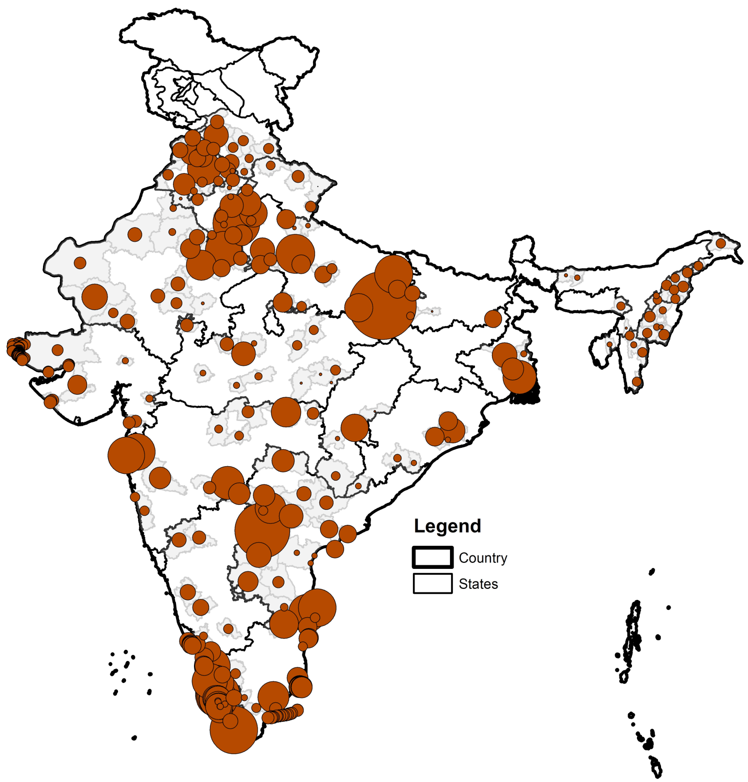


Source: Food and Agricultural Organisation (GAUL), NSSO and Census

Figure 2.A.3: Distribution of Informal Activity (with controls)

Manufacturing **Services**

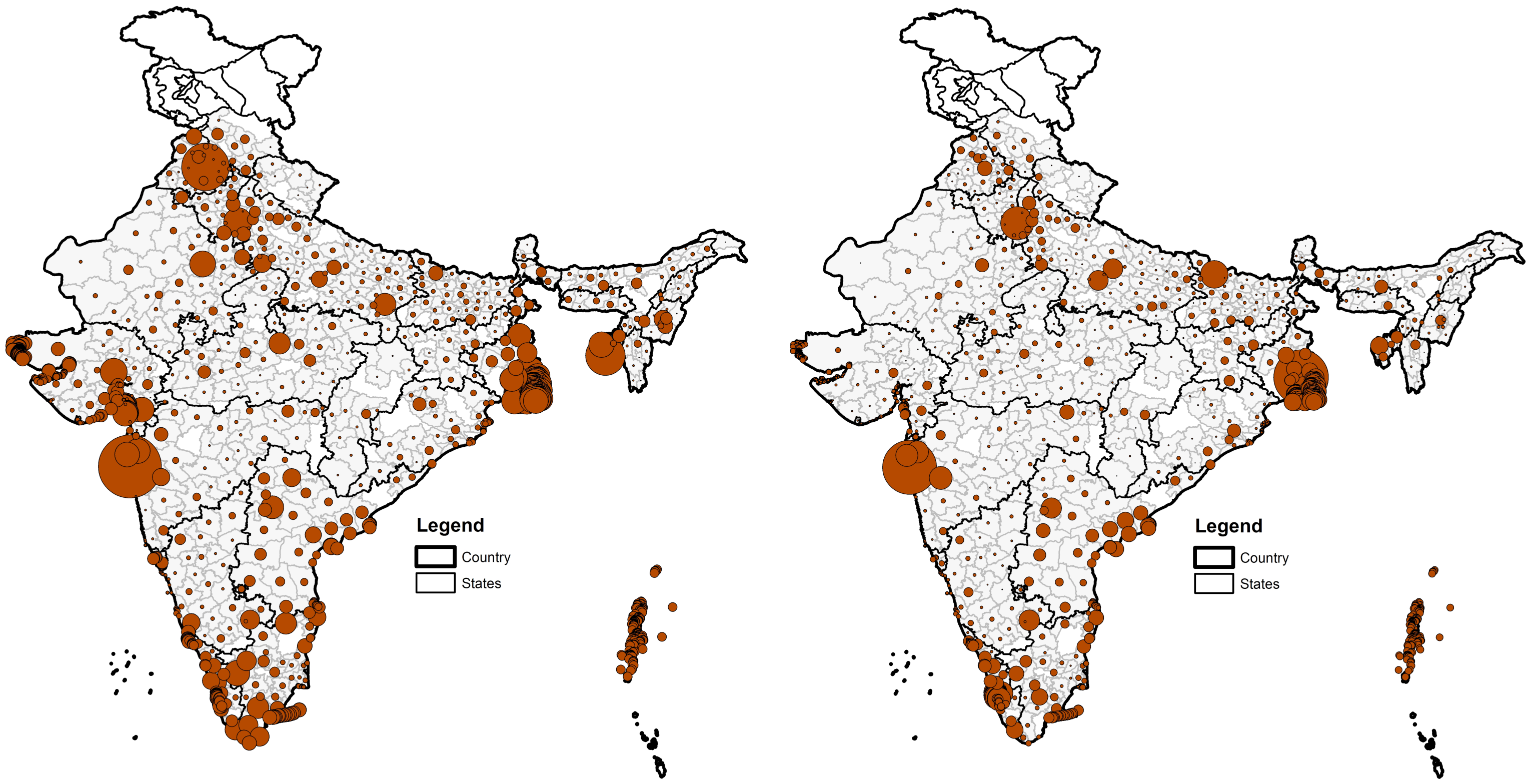




Source: Food and Agricultural Organisation (GAUL), NSSO, Census 2001

Figure 2.A.4: Contribution to the Theil Index (Unorganised Sector)

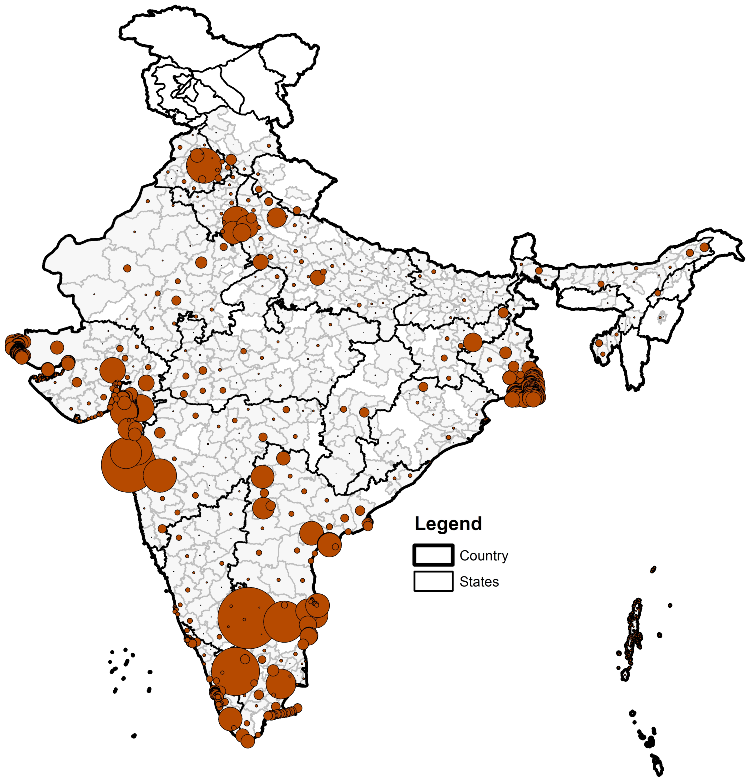
Manufacturing **Services**



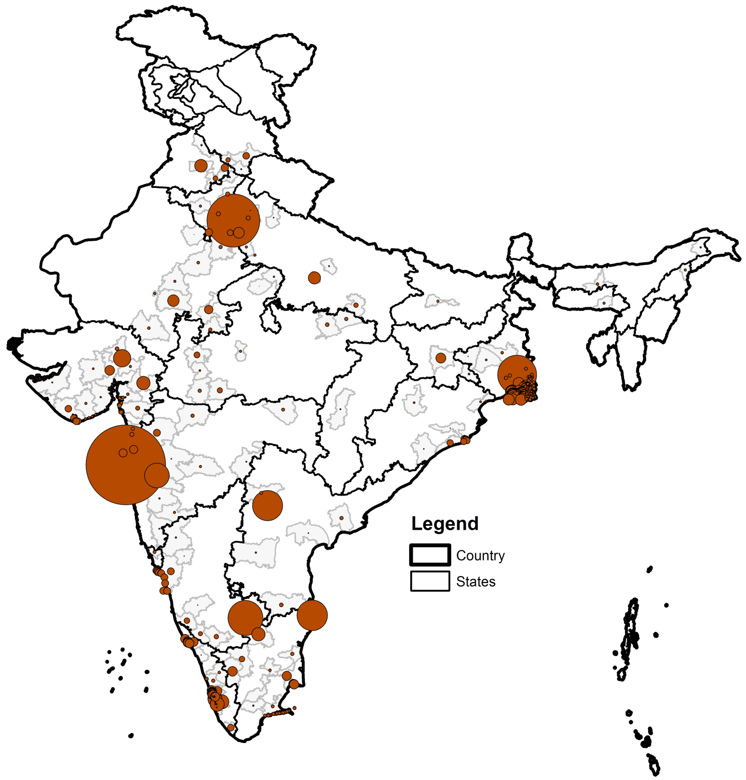
Source: Food and Agricultural Organisation (GAUL), NSSO, Census 2001

Figure 2.A.5: Contribution to the Theil Index (Organised Sector)

Manufacturing



Services



Source: Food and Agricultural Organisation (GAUL), ASI, Prowess, NSSO

CHAPTER 3

The Location Choices of Foreign Investors: A District-level Analysis in India*

(Joint with Peter Nunnenkamp)@

* Forthcoming in The World Economy (2011).
@ Peter Nunnenkamp came up with the initial idea for this paper, contributed partly to the data collection effort and to the drafting of the paper. My contribution to the paper is limited to the remainder – namely, data cleaning, literature review, model selection and econometric analysis, and help with data collection and drafting.

Abstract

This paper analyses the determinants of the location choices made by foreign investors at the district level in India to gauge the relative importance of economic geography factors, local business conditions, institutional conditions, and the presence of previous foreign investors. We employ a discrete-choice model and Poisson regressions to control for the potential violation of the assumption of Independence of Irrelevant Alternatives. Our sample includes about 19,500 foreign investment projects approved in 447 districts from 1991-2005. We find that foreign investors strongly prefer locations where other foreign investors are. This effect is significantly positive and robust across different years, sectors and different types of FDI. Moreover, path dependence remains significantly positive when controlling for institutional conditions at the state and district level. Foreign investors tend to follow previous investors from the same country of origin, but also investors from other countries of origin. They are also attracted to industrially diverse locations and to districts with better infrastructure and institutional conditions, though these findings are less robust. Surprisingly, districts in the neighbourhood of large metro areas do not benefit, in terms of attracting more FDI, from having easier access to these markets than remote Indian districts. On the contrary, our results suggest that large metro areas divert FDI projects away from neighbouring districts, thereby perpetuating or even widening the urban-rural divide.

Keywords: FDI, economic geography, location choice, infrastructure, institutions, path dependence

JEL classification: F23; R12

3.1 Introduction

The stock of foreign direct investment (FDI) in India soared from less than US$ 2 billion in 1991, when the country opened up to world markets, to US$ 123 billion in 2008 (UNCTAD, 2009). Policymakers in India as well as external observers attach high expectations to FDI. According to the (former) Minister of Finance, P. Chidambaram, "FDI worked wonders in China and can do so in India" (*Indian Express*, 11 November, 2005). Bajpai and Sachs (2000: 1) claim that FDI brings "huge advantages with little or no downside." However, the Chinese evidence also suggests that FDI contributed to widening income gaps between prospering coastal regions and provinces in the hinterland (e.g., Fujita and Hu, 2001; Zhang and Zhang, 2003).

Sachs, Bajpai and Ramiah (2002) argue that the reform-mindedness of Indian states has rendered them more attractive to FDI. However, the concentration of FDI in a few relatively advanced regions may prevent the effects of FDI from spreading across the whole economy. To the extent that greater openness to FDI leads to further agglomeration, FDI may fuel regional divergence, rather than promoting convergence. According to the Schumpeterian growth model of Aghion et al. (2005), more FDI promotes growth in relatively advanced regions, while leaving growth almost unaffected in poorer regions. Indeed, FDI is clearly concentrated at the level of Indian states (e.g., Purfield, 2006). Maharashtra accounted for more than a quarter of the amount of approved FDI in all-India in 2001-2005, followed by Delhi and Karnataka, which together contributed another quarter. Preliminary evidence also points to strong FDI clustering within large Indian states (Nunnenkamp and Stracke, 2008).

For less advanced regions to share the benefits of FDI, it is thus important to gain insights into the location choices of foreign investors. We estimate count and discrete choice models using project-specific FDI data to assess the determinants of location choices at the level of Indian districts. The focus is on the post-reform period of 1991-2005. In addition to various factors reflecting the local business environment, we account for economic geography factors, including distance-weighted market potential, as well as institutional conditions and previous location choices by foreign investors.

In the next section, we discuss how the present analysis relates to the previous literature and we derive our hypotheses for the case of post-reform India. We describe the data and introduce the estimation approach in Section 3.3. The empirical findings are presented in Section 3.4. Section 3.5 concludes with a discussion of major contributions and limitations.

3.2 Hypothesis and Related Literature

A fairly strong concentration of FDI in relatively few locations can be observed both across and within host countries. A small group of developed countries persistently absorbed more than two-thirds of worldwide FDI stocks.[46] Among developing countries, the 20 top performers account for more than 80 per cent of total FDI stocks. At the level of particular host countries, FDI in the United States has been shown repeatedly to be located primarily in a few large and relatively advanced states.[47] At the finer level of US economic areas, almost one third of the FDI transactions used by Chung and Alcácer (2002) fall into just four major metropolitan areas (out of 170 economic areas). Coastal areas in China absorbed about 90 per cent of overall FDI inflows during the period 1986-1998 (Zhang and Zhang, 2003). Likewise, the spatial distribution of new greenfield FDI in Portugal in 1982-1992 was biased heavily towards urban and coastal locations, especially around the largest cities of Lisbon and Porto (Guimaraes, Figueiredo and Woodward, 2000). FDI in India, too, is strongly concentrated both across and within states (see Section 3).

Models of location choice by foreign investors have addressed various factors that may help explain the concentration of FDI across and within host countries. The theoretical starting point typically is that foreign firms decide on a particular location based on expected profitability. Consequently, location choices depend on how the characteristics of one particular spatial unit and its geographic environment affect firms' profits relative to the characteristics of other spatial units. Major factors shaping these choices include expected demand for a firm's products, the supply of required inputs, factor costs, the quality of infrastructure and institutional conditions. In addition, previous

[46] See: http://stats.unctad.org/FDI.
[47] For instance, just two US states – California and New York – account for a quarter of the sample of manufacturing firms underlying the analysis of Coughlin, Terza and Arromdee (1991). See Coughlin and Segev (2000b) for an analysis at the level of US counties.

location choices by peers and competitors figure prominently on the list of FDI determinants and have received particular attention in the recent empirical literature.

A priori considerations on more specific hypotheses involve considerable ambiguity depending on the level of regional disaggregation and the type of FDI (Alegría, 2006; Blonigen et al., 2007). For instance, it may seem obvious that the size and purchasing power of local markets induce more foreign investors to enter a location. This hypothesis is most plausible in a cross-country context and as long as FDI is purely horizontal.[48] The motive of market access may also shape the distribution of horizontal FDI across fairly large spatial units in major host countries such as US or Indian states. Local demand should matter less, however, when location choices relate to smaller spatial units such as Indian districts, or when FDI is motivated by vertical specialization. Conversely, the *surrounding* market potential might become more important with smaller spatial units being analyzed.

Similar ambiguity prevails with regard to the costs of production. The relevance of wage costs, on which previous literature focuses, is "highly sensitive to small alterations in the conditioning information set" in cross-country studies according to the Extreme Bounds Analysis of Chakrabarti (2001). But even if higher wages discourage (vertical) FDI flows at the host country level, location choices by foreign investors within low-wage countries such as India are less likely to be affected. Regional wage disparity is small compared to average wage gaps between the host and source countries.[49] As a result, the concentration of FDI within low-wage countries is unlikely to be reversed by wage increases in its economic centres.

Availability of sufficiently skilled labour, which is the input of major interest in various studies, is likely to be a local pull factor accounting for the concentration of FDI within host countries such as India. According to the World Bank's Investment Climate

[48] This type of FDI essentially duplicates the parent company's production at home in the host countries of FDI. Market access motivations dominate over cost considerations. By contrast, vertical FDI provides a means to allocate specific steps of the production process to where the relevant cost advantages can be realized.

[49] See Alegría (2006) for a similar line of reasoning. In the case of India, average labour costs (per worker and day worked in 2003-04) differed by a factor of less than two between relatively rich states such as Maharashtra (Rs. 438) and relatively poor states such as Bihar (Rs. 237) (Government of India, 2006).

Assessment, survey respondents[50] complained about serious skill shortages in various Indian states (World Bank, 2004). Majumder (2008) stresses persistent regional disparities with respect to education. Majumder also provides extensive evidence on substantial variation in the regional quality of infrastructure, particularly with regard to financial infrastructure. Regional disparities in the quality of infrastructure appear to have widened in the post-reform era. The same seems to apply to institutional conditions. Bhaumik, Gangopadhyay and Krishnan (2009) argue that the degree of economic federalism in India has increased considerably since the early 1990s.

Local skills and efficient infrastructure can be expected to be important regional pull factors of FDI even though FDI-related outsourcing may primarily involve labour that is relatively low skilled from the country of origin's point of view. Feenstra and Hanson (1997) have clearly demonstrated that the corresponding labour demand of foreign investors qualifies as relatively high skilled in lower-income host countries such as in India. Even as early as the second half of the 1990s, UNCTAD had argued that foreign investors were increasingly pursuing so-called complex integration strategies. Accordingly, host countries would have to offer "an adequate combination of the principal locational determinants …. important for global corporate competitiveness" (UNCTAD, 1998: 112), including sufficiently skilled labour, adequate infrastructure facilities and specialized support services. Specifically related to India, UNCTAD (2004: 172-3) expects that services outsourced to India are moving towards higher value-added levels, thereby giving rise to fiercer competition for skilled local labour.

Regional disparities in India, in combination with increasingly complex integration strategies of firms, may strengthen the incentives of foreign investors to cluster in economic centres.[51] It is thus of particular interest to assess the self-reinforcing effects of FDI on current location choices. Existing clusters of FDI may attract subsequent FDI by allowing for knowledge spillovers as well as offering a wider range of intermediate inputs. According to Bobonis and Shatz (2007), an additional one per cent of FDI stock from a particular source country in a particular US state boosts the value of subsequent FDI from that source country in that state by 0.11 to 0.15 per cent. Head, Ries and

[50] Admittedly, these respondents had already chosen to locate themselves in these regions.
[51] Several studies suggest that regional inequality has increased in post-reform India, including Sachs, Bajpai and Ramiah (2002) and Kochhar et al. (2006). According to Lall and Chakravorty (2005), this also holds at the level of Indian districts.

Swenson (1995; 1999) use count data on the location choices of Japanese FDI in manufacturing industries of US states. The likelihood of a state being chosen by a subsequent investor in a particular industry increases by five to six per cent for states where the count of previous Japanese investments in this industry is ten per cent higher. By contrast, Guimaraes, Figueiredo and Woodward (2000) find the self-reinforcing effects of previous location choices by foreign investors to be rather weak in Portugal. Compared to the aforementioned studies on FDI at the level of US states, Guimaraes et al. analyze location choices at a much finer regional level, namely the 275 (fairly small) Portuguese *conselhos*, similar to our focus below on Indian districts.

Among developing host countries, China has received most attention with regard to the self-reinforcing effects of FDI. Head and Ries (1996) estimate a model of self-reinforcing FDI using data on the distribution of 931 foreign ventures across 54 Chinese cities in 1984-1991. Cheng and Kwan (2000: 379) consider FDI in 29 Chinese provinces in 1985-1995, finding "a strong self-reinforcing effect of FDI on itself."[52]

Recent contributions to the literature have refined the tools of accounting for economic geography and self-reinforcing FDI effects. Until recently, it was common to apply a simple form of geographic relationship among spatial units, i.e., setting a dummy variable equal to one for adjacent countries or regions.[53] By contrast, distance-related weighting schemes have been used by Blonigen et al. (2007) as well as Baltagi, Egger and Pfaffermayr (2007) to model more complex spatial effects, notably the surrounding market potential and the self-reinforcing effects of existing FDI clusters. A few studies have employed these tools so far at the regional level to assess location choices of foreign investors within particular host countries, or a group of host countries. For example, Alegría (2006) includes the external market potential, weighted by inverse distances, as an economic geography factor driving 4,800 instances of intra-EU FDI in the period 1998-2005. Ledyaeva (2009) assesses FDI determinants in Russian regions, accounting for external market potential and the spatially lagged dependent FDI

[52] Coughlin and Segev (2000a) also use provincial FDI data for addressing the dependence among Chinese provinces by estimating a spatial error (autocorrelation) model. Increased FDI in a province has positive effects on FDI in neighbouring provinces.
[53] Studies applying this concept of binary contiguity include: Head, Ries and Swenson (1995), Coughlin and Segev (2000a), and Bobonis and Shatz (2007).

variable.[54] Likewise, Crozet, Mayer and Mucchielli (2004) include external market potential and spatially lagged dependent variables, both weighted according to inverse distances in their study on FDI in France. The latter study resembles the present analysis in two respects: (i) the number of foreign investors deciding on where to locate is relatively large (almost 4,000 observations over ten years), and (ii) location choices relate to narrowly defined spatial units (92 French *départements*), rather than large regions such as US states.

3.3 Data and Estimation

3.3.1 FDI Data

We draw on a detailed account of FDI approvals in India during the period 1991-2005. The unpublished data were kindly made available by the Department of Industrial Promotion and Policy (DIPP) of the Ministry of Commerce and Industry. The dataset covers about 19,500 FDI projects, providing project-specific information on approved amounts, the home country of the foreign investor, as well as the state and district in India where the project is located. Non-resident Indians are included as a distinct source of FDI. It is also possible to distinguish FDI projects by foreign equity shares, making it possible to assess whether FDI determinants differ between minority and majority owned subsidiaries in India. Moreover, information on planned activities allows for a classification of FDI projects into broad sectors, notably a distinction between FDI in manufacturing and services.[55]

Approved FDI amounts may deviate considerably from realized FDI. However, it does not seriously constrain the subsequent analysis that the regional distribution of realised FDI in India is not available.[56] We focus on the counts of FDI projects, rather than

[54] As we do in the subsequent analysis, Ledyaeva (2009) performs several cross-section estimations for sub-periods of the whole period under consideration (1996-2005). Ledyaeva is mainly interested in whether FDI determinants and the type of FDI changed after the 1998 financial crisis in Russia.

[55] The sector structure of FDI in India has changed considerably since the early 1990s. FDI in services accounted for about 60 per cent of all approved FDI projects in recent years. In sharp contrast, FDI in manufacturing clearly dominated in the first half of the 1990s. FDI in the primary sector remained marginal throughout the period of observation. Chakraborty and Nunnenkamp (2008) observed similar shifts for realized FDI stocks.

[56] It may also be noted that aggregate data on realized FDI in India is not perfect either. It is only since 2000 that the Reserve Bank of India reports a revised series of realized FDI inflows that

approved amounts. While it cannot be ruled out that some approved FDI projects are not carried out at all, the count measure is unaffected by the typical gap between approved and realised amounts for particular projects. Changes in approval procedures after India's reform program of 1991 should not pose a major problem either. So-called automatic route approvals are included in the database until October 2004, according to information received from the Ministry of Commerce and Industry. Hence, our estimations are not distorted by the progressive extension of the list of FDI projects subject to the automatic approval route.

FDI in India is strongly concentrated at the state level. Maharashtra, Delhi and Karnataka accounted for more than half of the amount of approved FDI in all-India in 2001-2005 (Nunnenkamp and Stracke, 2008). Figure 1 shows that FDI is also spatially concentrated within states, i.e. at the district level. The maps reveal the density of FDI project applications; the size of the circles is proportional to the number of applications within the district. The left-hand side map illustrates that whilst some districts in the country potentially attract a lot of FDI activity, others are virtually empty. Of the possible 604 districts, FDI seems to be attracted to only 320 districts over the period of 1991-2005. Of these, 50 per cent of all FDI is drawn to only six districts. The right-hand side map replicates the same exercise, but after controlling for district population. FDI applications increase in districts in the southern and western parts of the country, and activity in districts around Delhi and Mumbai is better highlighted.

3.3.2 Econometric Model

The two most popular models of location choice are conditional logits (or nested logits) and Poisson regressions.[57] The use of a discrete choice framework to model location behaviour stretches back to the 1970s, when Carlton (1979) adapted and applied McFadden's (1974) Random Utility Maximisation framework to firm location decisions.

In line with the discrete choice framework, we start with positing a general profit function to explain the location behaviour of foreign investors choosing a location (here

---

includes reinvested earnings and debt transactions between related entities (to be counted as FDI according to international standards).

[57] See Appendix table 1 for a summary of studies employing these models.

district) in India. Following McFadden (1974) it is assumed that an investor i choosing to locate in district j will derive a profit of π_{ij}:

$$\pi_{ij} = U_{ij} + \varepsilon_{ij} \quad (1)$$

where U_{ij} is the deterministic part and ε_{ij} represents the random variable. District j will be preferred by the investor i if

$$\pi_{ij} > \pi_{ik} \text{ , } \forall k, k \neq j$$

The stochastic nature of the profit function implies that the probability that location j is selected by the investor i equals:

$$P_{ij} = \Pr ob(\pi_{ij} > \pi_{ik}) \text{ , } \forall k, k \neq j$$

It is assumed that the ith firm will choose district j if $\pi_{ij} > \pi_{ik}$ for all k where k indexes all the possible location choices to the ith firm. Under the assumption of independent and identically distributed error terms ε, with type I extreme-value distribution, the probability of choosing district j becomes:

$$P_{ij} = \frac{\exp(U_{ij})}{\sum_{k=1}^{n} \exp(U_{ik})} \quad (2)$$

The above equation expresses the conditional logit formulation. If we further assume that the systematic part of profit is affected by a set of m regressors, we can estimate the effects these have on location decisions. Typically, it is assumed that U_{ij} is a linear combination of the explanatory variables:

$$U_{ij} = \beta_1 X_{ij}^1 + \beta_2 X_{ij}^2 + ... + \beta_m X_{ij}^m$$

The simplicity of the conditional logit model (CLM) allows first insights into the behaviour of foreign investors across different districts within the country. For instance, it is possible that a foreign investor may consider large urban agglomerations in different states as possible alternatives. In other words, an investor may consider Kota in the state of West Bengal and Pune in Maharashtra as possible alternatives since they serve as satellite towns to larger cities (Kolkata and Mumbai, respectively).

In practice, however, the implementation of the conditional logit model in the face of a large set of spatial alternatives is very cumbersome.[58] The CLM is also characterised by the assumption of Independence of Irrelevant Alternatives (IIA). Consequently the ratio of the logit probabilities for any two alternatives j and k does not depend on any alternatives other than j and k. More formally this implies that the ε_{ij} are independent across individuals and choices; all locations would be symmetric substitutes after controlling for observables. This assumption could be violated if districts within particular states are closer substitutes than others outside of the state boundary. To effectively control for the IIA assumption, one would need to introduce a dummy variable for each individual choice. This would amount to a specification of the following type:

$$\pi_{ij} = U_{ij} + \varepsilon_{ij} = \delta_j + \beta' z_{ij} + \varepsilon_{ij} \qquad (3)$$

where δ_j s are the alternative specific constants introduced to absorb factors that are specific to each particular choice. In this case all explanatory variables (observable or unobservable) that only change across choices are absorbed by the alternative specific constants. However, in the presence of a large dataset this implementation would be impractical because of the large number of parameters to be estimated.

Count models gained popularity as the number of alternative locations increased, since what these lead to computational burdens in conditional logit models but in count models these are an advantage owing to the availability of more numerous observations. However, count models were at the time not understood to be as theoretically well founded as the conditional logit model, which is based on the Random Utility Maximisation (RUM) framework. This was until Guimaraes et al (2003, 2004) showed that count models can be specified in a way that is theoretically and empirically consistent with conditional logit models and thereby the RUM framework.

---

[58] Guimaraes, Figueiredo and Woodward (2003) provide an overview of the problems and how different researchers have attempted to deal with them in the past.

As an econometric alternative, Guimaraes et al (2003) show that the implementation of conditional logit models yields identical log likelihood functions to Poisson regression models when the regressors are not individual specific. They demonstrate how to control for the potential IIA violation by making use of an equivalence relation between the CLM and Poisson regression likelihood functions. In a separate paper, Guimaraes et al (2004) provide an empirical demonstration. In this model the alternative constant is a fixed-effect in a Poisson regression model, and coefficients of the model can be given an economic interpretation compatible with the Random Utility Maximisation framework.

Let n_{ij} be the number of investments in region j. Based on the profit function in Equation (3), the probability of investor i selecting location j would then become:

$$P_{ij} = \frac{\exp(\delta_j + \beta' z_{ij})}{\sum_{j=1}^{J} \exp(\delta_j + \beta' z_{ij})} \tag{4}$$

The parameters of equation (4) can then be estimated by maximising the following log-likelihood:

$$\ln L_{cl} = \sum_{j=1}^{J} n_j \log P_{ij} \tag{5}$$

Guimaraes, Figueiredo and Woodward (2003) show that Equation (5) is equivalent to that of a Poisson model that takes n_j as the dependent variable and includes a set of location-specific explanatory variables. The same results will be obtained if we assume that n_j follows a Poisson distribution with expected value equal to

$$E(n_j) = \lambda_j = \exp(\alpha + \beta' z_j)$$

That is, the above problem can be modelled as a Poisson regression where z_j varies across locations and n_j is the number of investments in j.

Information on actual individual firm choices is grouped into vectors of counts without any loss of information. This occurs since there are groups of firms faced with the same choice set and the same choice characteristics. For instance, consider the problem of identification of the relevant regional factors that affect firm location. Typically, researchers view these individual location decisions as profit (utility) maximising actions. Firms from diverse sectors evaluate the regional characteristics of different regions (i.e. districts) and choose to locate in the region that maximises potential profits. In this case, it is common to assume that all firms face the same choice set, and the relevant characteristics of the regional choices are identical for firms belonging to the same industry. The available information consists of regional counts of firm births by district and variables that reflect the characteristics of the regions. Despite the fact that the data consist of individual level choices, the true variation of the data is at the group level. Thus, data for the dependent variable may be summarised by vectors of counts. We are interested in modelling the data using McFadden's discrete choice Random Utility Maximisation (RUM) framework. This means that inference is based on the multinomial distribution because my interest lies in studying the impact that covariates have on choice probabilities, treating the number of firms in each group as given, wherein all firms share some common group-level characteristics. This introduces the possibility that there exist some unobservable group-specific effects that are likely to equally influence all firms belong to the same industry. If this happens, then the individual choices will be correlated and the vectors of counts will exhibit extra multinomial variation (i.e. overdispersion). Much like what happens with count models, in this circumstance the conditional logit model will remain consistent but will tend to underestimate the variance-covariance matrix.

Guimaraes and Lindrooth (2007) illustrate how the problem can be tackled by using Negative-Binomial count models, based on the Dirichlet-multinomial distribution. The Dirichlet-multinomial regression is a natural extension of McFadden's conditional logit model, and their paper shows that this relationship is the same as that of the negative binomial regression to the Poisson regression. In other words, the NB model provides a parametric alternative to deal with the problem of overdispersed data, and the parameters of the model would be equivalent to the Dirichlet-multinomial regression, which is an extension of the conditional logit regression. In the same way that the likelihood for the conditional logit model is obtained by letting the n_j (i.e. the number of new firms in a given district) following a Poisson law and conditioning on the total sum for each group, a similar relationship can be derived for the Dirichet-multinomial model. In this case, n_j is modelled directly as an overdispersed count variable, wherein the number of new firms is distributed according to the negative binomial law implying that the total sum of counts for each group also follows a negative binomial distribution.

More recently, Schmidheiny and Brulhart (2009) have shown that the elasticities of the conditional logit and Poisson models establish the boundary values at polar ends. In other words, they observe that the Poisson model implies more elastic responses by the dependent variable (in this case, investment counts) to given changes in own-region characteristics than the conditional logit model. Also, unlike in the conditional logit model, in the Poisson model, one region's change in locational attractiveness has no impact on the number of investments located among any of the other regions. We will exploit the features of all, conditional logits, Poisson and Negative Binomial, models to compute the two extremes for the elasticities, i.e. the percentage change in the expected number of firms in a region (or a neighbouring region) with respect to a unit change in the locational characteristics of the region.[59]

3.3.3 Specification of Variables

In the conditional logit model the dependent variable is a binary variable taking the value of one if the investor chooses to invest in a district and zero otherwise. In the count model the dependent variable is the count of new foreign investment projects approved in a district. As noted above, the overall sample includes about 19,500 foreign investment projects approved in 447 districts belonging to 35 states and union territories. Whilst we have annual FDI observations for the period 1991-2005, data on district-level location variables are available for only a few years. Our focus is on the fully specified model for 2001. The more limited estimations for two earlier years – 1996 and 1991 – serve as robustness tests. Furthermore, we perform panel estimations

[59] The conditional logit model implies a zero-sum allocation process of a fixed number of investments over the J locations. In the Poisson model, by contrast, new investments are non-rivalrous, in the sense that we are in a positive-sum economy and one region's gain is not another region's loss.

by pooling a limited set of variables for the three years 1991, 1996 and 2001. This helps mitigate endogeneity concerns (see Section 4.b for details).

The independent variables include characteristics of the district that can affect the profits of the investor. We classify these variables into those with an economic geography dimension, those that reflect business conditions in the district (notably, the availability of complementary factors of production and the quality of infrastructure), and those that relate to the behaviour of previous investors within the district as well as in other districts of the same state. In extended specifications, we also take account of institutional conditions at the state and district level. A brief description of major variables follows.

The economic geography variables in our model are represented by the Herfindahl index, market access, and population to indicate the size of the local market. We use the Herfindahl index to measure the degree of economic diversity in each region. HI_j is the sum of squares of employment shares of all industries in region j:

$$HI_j = \sum_k \left(\frac{E_{jk}}{E_j} \right)^2$$

Unlike measures of specialisation, which focus on one industry, the diversity index considers the industry mix of the entire regional economy. The largest value for HI_j is one when the entire regional economy is dominated by a single industry. Thus higher values signify lower levels of economic diversity.

Access to larger markets should provide a stronger incentive for investors to pick particular locations. For instance, investors in satellite towns, for instance Gurgaon, would also have access to larger neighbouring markets, say Delhi. The classic gravity model, which is commonly used in the analysis of trade between regions and countries, states that the interaction between two places is proportional to the size of the two places (as measured by population, employment or some other index of social or economic activity), and inversely proportional to some measure of separation such as distance. We use the formulation proposed initially by Hansen (1959) that states that the accessibility at point 1 to a particular type of activity at area 2 (say, employment) is directly proportional to the size of the activity at area 2 (say, number of jobs) and inversely proportional to some function of the distance separating point 1 from area 2. Accessibility is thus defined as the potential for opportunities for interaction. Thus, market accessibility is defined as:

$$MA_j = \sum_m \frac{S_m}{d_{j-m}^b}$$

where MA_j is the accessibility indicator estimated for location j, S_m is a size indicator at destination m, d_{jm} is a measure of distance between origin j and destination m, and b describes how increasing distance reduces the expected level of interaction.[60] The accessibility measure is constructed using population as the size indicator and distance as a measure of separation; it is estimated with the exponent value set to 1. The market access measure is constructed by allowing transport to occur along the straight line connecting any two districts. Instead of calculating the distance between any pair of districts across the country, we restrict the links to districts within a 500-kilometre radius.

Turning to district characteristics reflecting business conditions, we use non-agricultural hourly wage rates as an indicator of labour costs. We also account for education (higher-secondary education or middle-higher schools) as a proxy of the qualification of the workforce, which is traditionally considered to be an important complementary factor of production. In addition, we employ various proxies for the availability and quality of district-level infrastructure, including power (electricity), communications (telephone), transport (access to buses or roads), financial (bank branches) and health (access to health centres) infrastructure.

Although these variables might affect the business environment directly, we take a step beyond standard economic geography variables in extended specifications of our empirical models. The recent literature points to institutional conditions and the quality of governance as an additional dimension of location choice. For instance, Klapper, Laeven and Rajan (2006) find that costly regulations hamper the market entry of new firms, using firm-level data for various European countries. Bhaumik, Gangopadhyay

[60] We are grateful to Eckhardt Bode for providing us with the syntax for computing the great circle (orthodormic) distance calculations. In the original model proposed by Hanson (1959), b is an exponent describing the effect of the travel time between the zones.

and Krishnan (2009) focus on institutional factors at the level of Indian states and show that variations in entry rates in manufacturing industries were increasingly related with these factors in the era of economic federalism after major reforms in 1991. Institutional factors and the quality of governance might also have an important effect on making a more narrowly defined region such as Indian districts attractive to FDI.[61] However, data on institutional conditions are typically scarce for narrowly defined regions.

We attempt to overcome this bottleneck in two ways. First, we control for labour market institutions at the level of Indian states. Specifically, we include a dummy variable in our extended specifications which is set equal to one for FDI decisions involving all districts located in states with labour laws rated as pro-business by Besley and Burgess (2004). While labour regulations are mainly legislated and enforced by state governments, they are supposed to shape an important aspect of the cost of formal and informal contracting at the district level, too. Second, we also include a district-level variable on the frequency of riots and social unrest as a proxy for the quality of local institutions. This information is drawn from Marshall and Marshall (2008). Even though this district-level variable is not directly related to the cost of contract enforcement, it captures investor uncertainty and resembles rule-of-law indicators that are widely used in the cross-country literature on the determinants of FDI.

Finally, following Crozet, Mayer and Mucchielli (2004), we include a variable to account for previous investment choices by foreign investors. In particular, we assess whether foreign investors are attracted to locations that attracted other foreign investors before, and whether this effect is stronger for investors from the same country of origin. We include a count variable to take account of all foreign firms within a district and in all districts within the state, weighted by their distance. Formally stated, we include:

$$FA_j = Count_j + \sum_{j \in s} \frac{Count_m}{d_{j-m}} \quad \text{and} \quad FS_j = Count_j + \sum_{j \in s} \frac{Count_m}{d_{j-m}}$$

where FA_j and FS_j refer, respectively, to all foreign and all same-country foreign firms in location j. The values of *FA* and *FS* are computed for all years leading up to the year for which the cross-section is carried out.

[61] We thank an anonymous reviewer for having alerted us to this issue.

Table 3.1: Explanatory Variables – Description and Sources

	Variable	Indicator	Source(s)	Availability 1991	1996	2001
Economic geography	HI	Economic diversity	NSSO			√
	MA	Market access	Census/ Orthodromic distance calculations	√	√	√
	Population	Total population	Census data	√	√	√
	Wages	Non-agricultural hourly wage rates	NSSO			√
Business environment/ Infrastructure	Electricity*	Proportion of villages with access to electricity	NSSO/CMIE	√	√	√
	Telephone*	Proportion of villages with access to telephone connections	NSSO/CMIE	√	√	√
	Education*	Middle-higher schools per 1 lakh population	CMIE/NSSO	√	√	√
	Buses	Proportion of villages with bus services	Census data			√
	Roads	Road length per 100 square kilometre	CMIE	√	√	
	Banks	Banking branches per 1 lakh population	CMIE	√	√	√
	Health	Primary health centres per 1 lakh population	CMIE	√		
Institutional variables	Labour Regulations	Labour Law Flexibility	Besley and Burgess (2004)	√	√	√
	Riots	Number of riots per capita	Marshall and Marshall (2008)	√	√	√
Previous FDI	FA	Clustering of previous FDI	Ministry of Commerce and Industry	NA	√	√
	Cumulative	Total FDI projects in (t-1)	Ministry of Commerce and Industry	NA	√	√

Notes:*For 2001: Electricity, Telephone and Education refer to the proportion of population with access to electricity, with a telephone connection and with a higher-secondary education (Source: NSSO);
1 Lakh = 100,000; NSSO: National Sample Survey Organisation; CMIE: Centre for Monitoring of the Indian Economy

3.4 Results and Discussion

3.4.1 Basic Results

We illustrate the key characteristics of the data and the subsequent modelling choices, by using the 2001 cross-section as an example. One of the key characteristics of the data is that it is over-dispersed. In Table 3.A.3 the mean number of investments per district is around 11, while the standard deviation is over 90, i.e. over eight times the mean. A Poisson model implies that the expected count, or mean value, is equal to the variance. This is a strong assumption, and does not hold for our data.

A frequent occurrence with count data is an excess of zeroes compared to what would be expected under a Poisson model. This is indeed a problem faced by our data – the mean number of investments is about 35 when excluding zeros and the standard deviation is 161, i.e. around 4.6 times the mean. Also note that 369 out of 533 districts did not receive any investments in 2001-2003.[62] This implies that we would need to take into account, both, over-dispersion and the excess of zeroes in the data, when selecting a model to fit the data.

Another way to reiterate the unsuitability of the Poisson model in this case is to show that such a model is unable to predict the excess zeroes found in our data. In Table 3.A.3, "obs" refers to actual observations in the data, and fitp and fitnb refer to the predictions of the fitted Poisson and negative binomial models respectively. While 69.23 per cent of the locations in the sample received no investments, the Poisson model predicts that only 58.35 per cent would get no investments. Clearly the Poisson model underestimates the probability of zero counts. The negative binomial model, which allows for greater variation in the count variable than that of a true Poisson, predicts that 63.85 per cent of all districts will receive no investments, much closer to the observed value.

Against this backdrop, Table 3.2 reports the results of the 2001 cross-section, step-by-step and based on alternative models. In this way, we check whether the above noted path dependence of FDI decisions is sensitive to the introduction of economic

[62] Although there are a total of 604 districts in India, we exclude all districts for which we do not have data for the regressors.

geography (column 2) and infrastructure (column 3) variables. At this stage, the estimations are performed without institutional variables to retain as many observations as possible. The response variable is 'count', i.e. the number of investments received by a district. The count regression models the log of the expected count as a function of the predictor variables. More formally, $\beta = \log(\mu_{x+1}) - \log(\mu_x)$, where β is the regression coefficient, μ is the expected count and the subscripts represent where the regressor, say x, is evaluated at x and x+1 (here implying a unit percentage change in the regressor).[63] Since the difference of two logs is equal to the log of their quotient, i.e. $\log(\mu_{x+1}) - \log(\mu_x) = \log(\frac{\mu_{x+1}}{\mu_x})$, we could also interpret the parameter estimate as the log of the ratio of expected counts. In our case, the count refers to the 'rate' of investments per district. The coefficients are reported in non-exponentiated form and can be interpreted as follows: if the population were to increase by a percent, then the expected count of investments would increase by 43 percent.

Table 3.2: Model Specifications (2001 cross-section)

Model *Variables*	NB (1)	NB (2)	NB (3)	Poisson (4)	CL (5)
Population	0.4337***	0.2702**	0.4490***	0.6490**	0.6727***
FA	0.5737***	0.5285***	0.4423***	0.3623***	0.3069***
HI		-0.6165***	-0.1943	0.2351	-0.1153*
MA		-0.6309***	-0.7692***	-0.6115***	-0.6280***
Electricity			0.0008	0.0075	0.0018
Telephone			0.0006	-0.0052	-0.0051**
Education			0.0590**	0.0592*	0.0651***
Buses			0.1281	0.0318	0.0479***
Banks			0.0817***	0.0762**	0.0685***
Wages			-0.0045***	-0.0050*	-0.0054***
#	531	454	402	402	309,914
AIC	3,340	3,946	3,790	4,229	4,099
BIC	3,353	3,967	3,814	4,276	4,150

Notes: NB – Negative Binomial, CL – Conditional Logit.
Robust standard errors clustered at district level (not shown). Non-Exponentiated coefficients
*** p<0.01, ** p<0.05, * p<0.1

We also present the results of the conditional logit estimation in the last column of Table 3.2 commonly referred to as 'odds ratios'. The odds ratio can be interpreted as

[63] This is because the regressors are in logarithms of the original independent variables.

follows: A unit percentage increase in education would be associated with a 6.51 percent increase in the odds of receiving an investment in a district. The number of observations is much higher since the dependent variable is the choice of foreign investors to select a district, i.e. for each investment it equals 1 for the district chosen and 0 for all other districts not selected. In order to select the preferred model, the Bayesian information criterion (BIC) and Akaike's information criterion (AIC) are also provided. Since the models are used to fit the same data, the model with the smallest values of the information criteria is considered superior. By these criteria the negative binominal models generally perform better than the Poisson models.

For the time being, these results focus on the 2001 cross-section as the data situation is clearly superior compared to earlier years. Apart from the health-related variable on infrastructure, the full set of explanatory variables is available for 2001 (see Table 3.1 above). By contrast, two variables of major interest – the Herfindahl index on economic diversity and labour costs – are lacking for earlier years. Moreover, the data is available for most Indian districts in 2001, whereas coverage of districts is limited in 1991 and 1996. We will use data from the two additional years for which explanatory variables are available, 1991 and 1996, to carry out panel regressions later in the paper.

The dependent count variable used for the 2001 cross-section actually includes FDI projects approved during the three-year period 2001-2003. In this way we make use of a larger part of the FDI database introduced in Section 3 above. At the same time, the consideration of three years smoothes cyclical FDI fluctuations.[64] As a first result, it should be noted that we find a strong tendency of foreign investors to go where other foreign investors are already present. The preferred binominal models reveal a particularly strong clustering of FDI; but even the Poisson and the conditional logit (CL) models suggest that a percentage increase in the value of *FA* (which includes investors within the same district and in neighbouring districts) would result in a 30 to 44 per cent increase in the expected rate of FDI counts. Whenever available the coefficient is statistically significant at the 0.1 per cent level.[65] In other words, insofar as

[64] Similarly, we use FDI approvals in 1996-1998 for the 1996 cross-section and, respectively, approvals in 1991-1993 for the 1991 cross-section.
[65] Since convergence could not be achieved with the full set of predictor variables within zero-truncated models, the coefficients for the dropped variables could not be computed.

India's reform program initiated in the early 1990s resulted in more FDI, the latter seems to have concentrated in particular regions.

Turning to the economic geography variables, population consistently has a positive effect and, typically, is highly significant at the one per cent level. In the negative binominal models, the quantitative effect is about 44 per cent if population increases by one percentage point. Recalling the discussion on horizontal FDI in Section 2, population is surprisingly robust as a relevant driving force of FDI at the district level. This finding contradicts the popular view held in various source countries that the boom of FDI in post-reform India is mainly associated with vertical, i.e. cost-cutting FDI; it rather appears that FDI is horizontal, i.e. closely associated with the size of the local market.

In contrast to population in the district where FDI locates, distance-weighted population in other districts - representing our proxy of market access (*MA*) - does not appear to positively impact FDI. The preferred negative binominal models in Table 3.2 suggest a negative effect. This is in conflict with the hypothesis that districts in the neighbourhood of large metro areas are likely to benefit, in terms of attracting more FDI, from having easier access to these markets than remote Indian districts. Rather, it appears that large metro areas divert FDI projects away from neighbouring districts, thereby perpetuating or even widening the urban-rural divide.[66] Conversely, the sharp urban-rural divide in India implies limited market potential surrounding the metro areas, which further weakens any positive effects *MA* may have on FDI.

HI seems to have a negative relationship with FDI projects. Recall that this variable is a measure of the level of industrial diversity within the district. A higher *HI* implies higher employment concentration by one industry and lower industrial diversity. Thus, the negative coefficient for *HI* is evidence of a positive association between more industrial diversity and more FDI. This could be since foreign investors are increasingly pursuing complex integration strategies, as noted by UNCTAD (1998). Consequently, they rely on a diverse set of intermediate inputs from various industries, giving locations a competitive edge where these inputs are easily available. Our finding on *HI*

[66] It should be noted in this context that, for instance, almost 90 per cent of approved FDI projects in Karnataka went to Bangalore; Kolkata accounted for 70 per cent of projects approved in West Bengal (Nunnenkamp and Stracke, 2008).

is in line with Kathuria (2002), according to whom the degree of vertical integration of FDI projects has declined in post-reform India. Unfortunately, owing to lack of data we are unable to test this relationship further for other cross-sections.

It is not only greater industrial diversity that attracts FDI to Indian districts. The same applies to districts with a better-qualified workforce, as reflected in the positive and statistically significant coefficients. This may indicate that the availability of sufficiently qualified labour in Indian districts is important as a complementary factor of production.

While better-educated workers attract FDI, higher labour costs could be expected to discourage FDI. The negative effect is significant across the different models – a percent increase in wages seems to lower the percent of investments by anything between 0.45 and 0.54. As will be shown below, the impact of wages on FDI differs considerably across sectors.

Finally, Table 3.2 reveals somewhat ambiguous findings concerning the relationship between infrastructure and FDI at the district level. On the one hand, electricity seems to have no effect irrespective of the choice of model. On the other hand, telephone connections – our proxy of communication infrastructure – typically have a negative effect. The evidence varies across models with respect to transport infrastructure (proxied by bus services). Although it seems that the effect of financial infrastructure is positive and significant across different model specifications. As a matter of fact, the role of infrastructure as a determinant of FDI in India has remained disputed. While the World Bank (2004) claims that deficient infrastructure represents an important bottleneck to investment even in relatively advanced states such as Maharashtra and Gujarat, Chakravorty (2003) finds that infrastructure had little influence in determining the location or quantity of new industrial investment. Nunnenkamp and Stracke (2008) show that the impact of infrastructure on state-level FDI depends on the specific indicator chosen.

3.4.2 Extensions and Sub-samples

In the basic specification of our empirical models, we did not account for the quality of institutions and governance as an additional dimension of location choice. While cross-country studies can draw on a wealth of mostly survey-based data on institutions and governance (e.g., the World Bank's Worldwide Governance Indicators), comparable data are typically lacking for regional units within particular countries. Neglecting institutions and governance as possible determinants of FDI does not necessarily bias results in country-specific studies. The development of institutions and the quality of governance may reveal little variation within countries, especially in relatively small countries with centralized governance structures. In large countries with a federal structure such as India, institutions and governance are likely to vary at the state level, though probably much less so at lower levels of regional disaggregation. Indian districts belonging to the same state may not differ considerably with respect to formal institutions. All the same, they may differ with respect to the implementation and enforcement of good governance.

Against this backdrop, we extend the basic specification of our empirical models in two ways in the following: We include a dummy variable which is set equal to one for FDI decisions involving all districts located in states with labour laws rated as pro-business by Besley and Burgess (2004), and we include a district-level variable on the frequency of riots and social unrest as a proxy for the effective enforcement of rules that may reduce investor uncertainty.

Table 3.3: 2001 cross-section (Including institutional variables)

Model *Variables*	NB (1)	NB (2)	Poisson (3)	CL (4)
Population	0.5939***	0.4853***	0.9673***	-0.4444***
FA	0.4243***	0.4170***	0.5071***	0.4709***
HI	-0.1664	-0.1965	0.2770	0.1406
MA	-0.7727***	-0.7463***	-0.6871***	-1.2664***
Electricity	-0.0021	0.0002	0.0047	-0.0177***
Telephone	-0.0006	-0.0002	-0.0226*	-0.0215***
Education	0.0601**	0.0506*	0.0428	0.0918***
Buses	0.5223**	0.4138*	0.2416	0.3997***
Banks	0.1012***	0.0960***	0.1140***	0.2143***
Wages	-0.0043***	-0.0039***	-0.0052**	-0.0091***
Labour Regulations	0.1634	0.1942*	0.2838*	0.3966***
Riots		0.0303	-0.0731	-0.2375***
FS				0.2981***
#	368	342	342	158,177

Notes: NB – Negative Binomial, CL – Conditional Logit.
Robust standard errors clustered at district level (not shown). Non-Exponentiated coefficients
*** p<0.01, ** p<0.05, * p<0.1

In columns (1) and (2) of Table 3.3, we introduce the standard and the additional variables into the NB Model for the 2001 cross-section. It appears that path dependence (i.e. the coefficient on *FA*) weakens only slightly when we account for institutions. The coefficients of most of our standard variables are broadly stable and similar to those presented in Table 3.2. Interestingly, pro-business labour regulations at the state level have a positive and statistically significant effect on making a region more attractive to FDI. Additionally, the frequency of riots and unrest as a measure of local governance is strongly and negatively related to FDI within a district, under the conditional logit (CL) model presented in column (4).

The evidence on the importance of infrastructure strengthens in some respects when also accounting for institutions. For instance, transport infrastructure (*Buses*) as well as financial infrastructure (*Banks*) now has a strongly positive and statistically significant effect on FDI in the negative binominal and poisson models. These coefficients are also broadly stable with regards to the results for the conditional logit estimation presented as odds ratios. On the other hand, the effect of the two institutional variables is sensitive to the choice of the model. Pro-business labour regulations have a positive effect within the Poisson and the CL models, and social unrest (i.e. riots) has a negative and statistically significant effect in the CL models only. All in all, the estimations suggest that the basic results would not be grossly biased due to the omission of institutions and governance at the local level.

In Column (4), we re-run the CL model by separately accounting for previous investment choices by foreign investors of the same country of origin (*FS*).[67] We find that foreign investors tend to locate where investors from the same country of origin located before; the corresponding odds ratio is 0.29, at the 0.1 per cent level. However, the tendency to follow investors from the same country of origin does not appear to be stronger than the tendency to follow other foreign investors. This is broadly in line with the findings of Crozet, Mayer and Mucchielli (2004) for FDI in French *départements*.[68]

---

[67] This estimation can only be performed for the CML as the choices of foreign investors need to be matched on a one-to-one basis with those belonging to the same country. This matching exercise is not possible to carry out when investments are grouped as a count variable.

[68] The inclusion of *FS* affects the size and significance of some other variables, including labour regulations at the state level, when comparing the last columns in Tables 3 and 5. .

In the cross-section analysis reported so far, it is impossible to control for all possible sources of endogeneity.69 In particular, location choices by previous investors may be jointly determined with our dependent variable resulting in an omitted variable bias.70 This also prevents us from employing an instrumental variable (IV) approach. It is hard to conceive of an additional variable with available information at the district level that is correlated with location choices of previous investors, but uncorrelated with location choices of current investors. Moreover, the data situation does not allow for a panel analysis covering a continuous set of years. All the same, we address potential endogeneity concerns in the following by running panel regressions for the three years 1991, 1996 and 2001.

We only include those variables for which we have reliable data across the three years. The set of variables includes the institutional indicators introduced in the extended specifications above. On the other hand, the Herfindahl index and wages have to be excluded as the required data are not available for the two earlier years. We include year fixed effects to control for omitted variable bias[71]. As a result, we can test for the robustness of our major findings even though the panel specification is based on a limited set of determinants, compared to the cross-section analysis for the year 2001.

We present the empirical results for the negative binomial model in Table 3.4. Not surprisingly, the inclusion of year fixed effects in the panel regression affects the magnitudes of the coefficients to some extent. Nevertheless, several findings from the previous cross-section analysis prove to be fairly robust. For example, the coefficient on population is affected only marginally. The same applies to financial infrastructure (*Banks*). The coefficient on *FA* still reveals strong path dependence, even though its size is considerably smaller than in the cross-section analysis. Among the standard variables, the panel regression results on education deviate most pronouncedly from the results for the 2001 cross-section. Education, which is proxied by the number of middle-higher

[69] As noted before, we also estimated separate cross-sections for 1991 and 1996 to fully exploit the available data. Although this does not resolve endogeneity concerns, it reduces the possibility of bias.

[70] By contrast, there is little reason to be concerned about reverse causality running from our regressors to firm-specific location choices. Note also that we lag our regressors by assessing their impact on location choices in the concurrent and the two subsequent years, in order to mitigate possible endogeneity problems.

[71] Ideally, we would like to introduce location fixed effects to control for any omitted variables bias. However, concordance is not reached with the introduction of either state and/or district fixed effects.

schools per 100,000 inhabitants in the panel regressions, now seems to have a mostly statistically insignificant effect on FDI. State-level institutions as reflected in pro-business labour regulations seem to have no discernible effect. By contrast, governance at the district level (*Riots*) is insignificant for the full specification shown in columns (3) and (4) of Table 3.4.

Table 3.4: Panel Estimations: 1991, 1996 and 2001 (NB model)

Variables	(1)	(2)	(3)	(4)
Population	0.2918***	0.6671***	0.6268***	0.5259***
MA	0.4822	0.7199	-0.6608*	-0.5023***
Education	0.5966**	0.3938*	0.3728*	0.0100
Banks	0.7425***	0.8922***	0.0729***	0.0732***
Labour Regulations		0.0176	0.0425	0.1644
Riots			0.0109	0.0429
FA				0.2242***
Year Fixed Effects	*Yes*	*Yes*	*Yes*	*Yes*
Observations	814	791	759	438

Robust standard errors clustered at district level (not shown). Exponentiated coefficients
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

In the final step of our analysis, we perform estimations for sub-samples of FDI decisions, based on the full specification with institutional variables for the 2001 cross-section. First, we differentiate between FDI in the secondary and the tertiary sector and re-run the cross-section regressions for the year 2001 to observe if the effect of the predictor variables varies across these two sectors. Although we carried out the regressions for other models as well, we only report the results of the negative binomial specifications.[72] This is to facilitate comparison, but more importantly because the negative binomial model exhibited the best goodness-of-fit statistics.

In several respects, the results for the manufacturing and services sub-samples in Table 3.A.4 closely resemble the corresponding negative binominal model results for the overall FDI sample in Table 3.3. For both sub-samples, there appears to be a strong tendency to locate where other foreign investors have chosen to locate.[73] We find, as before, that population has a positive effect on FDI projects. While the tendency to

[72] Results from the other models are available on request.
[73] We estimated *FA* for the total of projects within the same district and in neighbouring districts; i.e., this variable is not further disaggregated by the type of sector. This is because we would like to capture the effect of FDI drawn to locations for reasons of intra-industry advantages, but also for buyer-supplier linkages across sectors.

follow previous investors appears to be slightly stronger in the manufacturing sector, population has a somewhat stronger impact of FDI projects in the services sector. This is in line with economic intuition, in which services industries usually benefit more than manufacturing industries from being close to where people are situated. As before for the overall sample, our measure of industrial diversity (*HI*) is not significantly different from one if the negative binominal model is estimated for sector-specific FDI.[74] The same applies to the two institutional variables when comparing the results in Table 3.A.4 with those in column (6) of Table 3.3.

Yet we find some interesting differences between FDI in manufacturing and FDI in services. Most notably, the costs of local labour discourage FDI only in the services sector. Here, the effect of wages is significant at the five per cent level, with a one percentage point increase in wages resulting in a decline by more than 30 per cent in the expected count of FDI projects. The contrast in wage effects between sectors invites the conclusion that vertical FDI, which is mainly motivated by cost considerations, is largely restricted to India's services sector, whereas FDI in its manufacturing sector continues to be horizontal, i.e., local market seeking.[75] This could also be because certain services sectors rely on access to cheap and abundant labour (for instance, call centres) and they would then be theorised to be more sensitive to wage increases. However, since we are unable to disaggregate the data down to the two-digit industry level, we cannot be certain which specific industries within these sectors maybe driving the results. Moreover, the sector-specific perspective adds to the ambiguity concerning infrastructure. The effect of electricity proves to be strongly significant for FDI in manufacturing only, while the effect of banking facilities at the district level on FDI is particularly strong in the services sector at the one per cent level of significance.

The next step distinguishes between majority and minority foreign owned joint ventures (JVs). We use foreign equity shares as presented in the DIPP database to group all FDI projects into these two categories.[76] This distinction may be relevant as higher foreign equity shares tend to be associated with a relatively strong bargaining position of

[74] The IRR for our variable on education appears to be of similar size for FDI in manufacturing and services. In contrast to the corresponding estimation for the overall sample, however, this variable turns statistically insignificant at the ten per cent level for sector-specific FDI.
[75] Agarwal (2001) suspects FDI in India continues to be domestic market seeking.
[76] Information on the foreign equity share is missing for various FDI projects, which results in a considerably reduced number of observations.

foreign investors (e.g., Asiedu and Esfahani, 2001) which, in turn, may imply that location choices are more strongly determined by the preferences of foreign investors than those of the host government. Indeed, there are some significant differences between majority and minority owned JVs with respect to the determinants of location choices. Majority owned JVs seem to rely more strongly on better-educated workers than minority JVs, possibly because the former involve transfers of relatively advanced technologies. More surprisingly, economic diversity (reflected in higher values of *HI*) appears to discourage majority owned JVs. Possibly, these JVs are associated with a higher degree of vertical integration and draw mainly on specialised inputs from within the same industry. On the other hand, infrastructure plays a more important role for minority owned JVs. This is most plausible for financial infrastructure given that majority owned JVs may turn primarily to international capital markets for financing.

Similar to majority owned JVs, larger FDI projects appear to be more reliant on better-educated workers. This is revealed by the third sample split when considering the median of the value of foreign equity as the dividing line between small and large FDI projects. By contrast, the effect of local financing is somewhat stronger for smaller FDI projects. At the same time, smaller FDI projects tend to be discouraged by higher wage costs. In other respects, however, our results are barely affected by distinguishing between small and large projects. Notably, local institutions and governance remain insignificant independent of the size of projects.

3.5 Concluding Remarks

This paper contributes to the empirical literature relating to the geography of foreign direct investment in a number of ways. Although there is some previous research on the behaviour of foreign investors in emerging countries like China, to our knowledge this is the first paper that analyses location decisions for over 19,500 FDI projects in India. We differentiate between economic geography factors, such as the presence of existing FDI, access to neighbouring markets and industrial diversity, and factors relating to the local business conditions including the quality of infrastructure at the level of districts. The difference between these two sets of factors is of obvious policy relevance: Whilst public policy may be in a position to influence the level and quality of infrastructure within a lagging region, its ability to affect economic clustering is limited.

Indeed, we find that path dependence tends to constrain the influence of regional policymakers. Foreign investors strongly prefer locations that already host other foreign investors. This effect is significantly positive and robust across different years, sectors and different types of FDI. Moreover, path dependence remains significantly positive when controlling for institutional conditions at the state and district level. Foreign investors tend to follow previous investors from the same country of origin, but also investors from other countries of origin. We also find that the degree of economic diversity within a location attracts FDI, but this result is less robust and we are unable to test it for different years. More surprisingly, districts in the neighbourhood of large metro areas do not benefit, in terms of attracting more FDI, from having easier access to these markets than remote Indian districts. On the contrary, our results suggest that large metro areas divert FDI projects away from neighbouring districts, thereby perpetuating or even widening the urban-rural divide.

However, geography is not destiny. In several respects local business conditions matter for the location choices of foreign investors at the level of districts. For instance, the presence of an educated population has recently become a significant factor drawing FDI projects to a location. This reveals that FDI in post-reform India is attracted not only by lower labour costs but also by the availability of sufficiently skilled labour as an important complementary factor of production. Providing adequate schooling and training thus appears to be an important policy tool for regional policymakers.

Investing in infrastructure represents another option to attract FDI. Access to power, transport and financial infrastructure – factors that the World Bank often classifies as ‘investment climate’ – clearly matters for the location decisions of foreign investors. However, the evidence is more ambiguous when it comes to the question of which aspect of infrastructure is particularly relevant for FDI in particular sectors. For instance, the presence of bank services within a location seems to be an important factor driving location choices for FDI in services, but less so for FDI in manufacturing. The opposite pattern prevails for power supply. This ambiguity may render it difficult for policymakers to decide on investment priorities in the area of infrastructure.

Institutional factors also seem to matter, both at the state and at the district level. Pro-business labour regulations tend to make a state more attractive to FDI, and just as one would expect, less social unrest within a district attracts more FDI. However, the latter result is not robust to our panel estimations, whilst the effect of labour law flexibility is sensitive to the choice of models. The policy options at the local level may be constrained to the extent that institutional conditions such as labour laws in India are a state subject. However it is difficult to know how implementation and enforcement might differ across districts within states. Furthermore, it remains open to question how foreign investors react to other critical governance issues such as the enforcement of property rights and the credibility and accountability of local authorities.

Future research may also help overcome a few other shortcomings once additional data are made available by Indian authorities. While we conduct panel regressions at the district level and use year fixed effects to mitigate endogeneity concerns, such concerns could be better addressed with data for more continuous years. Additional insights might also be gained if all FDI projects were differentiated between industrial sectors at a disaggregated level. Industry-specific estimations could reveal whether the location choices of foreign investors and the relative importance of economic geography and local business conditions differ across industries. For instance, computer or financial services would probably require different levels and types of labour skills than retail or transport services.

Finally, even though the FDI data offer various details on the characteristics of joint ventures in India, it would be desirable to match this dataset with data on the foreign parent company. Its size, age, productivity and technological sophistication may shape location decisions, in addition to district characteristics. This is especially true when such firm characteristics influence the relative bargaining position of foreign investors vis-à-vis regional authorities competing for FDI.

References

Agarwal, D.R. (2001), 'Foreign Direct Investment and Economic Development: A Comparative Case Study of China, Mexico and India', in R.K. Sen (ed.), Socio-Economic Development in the 21st Century (New Delhi: Deep & Deep).

Aghion, P., R. Burgess, S.J. Redding, and F. Zilibotti (2005), 'Entry Liberalization and Inequality in Industrial Performance', Journal of the European Economic Association, 3, 2-3, 291-302.

Alegría, R. (2006), 'Countries, Regions and Multinational Firms: Location Determinants in the European Union', London School of Economics, mimeo.

Asiedu, E. and H.S. Esfahani (2001), 'Ownership Structure in Foreign Direct Investment Projects', Review of Economics and Statistics, 84(4): 647-62.

Arzaghi, M and V. Henderson (2008), 'Networking off Madison Avenue', Review of Economic Studies, 75(4): 1011-38.

Bajpai, N. and J.D. Sachs (2000), 'Foreign Direct Investment in India: Issues and Problems', Development Discussion Paper No. 759, Harvard Institute for International Development (Cambridge, MA: Harvard University).

Baltagi, B.H., P. Egger and M. Pfaffermayr (2007), 'Estimating Models of Complex FDI: Are There Third-Country Effects?', Journal of Econometrics, 140(1): 260-81.

Becker, R. and V. Henderson (2000), 'Effects of Air Quality Regulations on Polluting Industries', Journal of Political Economy, 108(2): 379-421.

Besley, T. and R. Burgess (2004), 'Can Labor Regulation Hinder Economic Performance? Evidence from India', Quarterly Journal of Economics, 119(1): 91-134.

Bhaumik, S.K., S. Gangopadhyay and S. Krishnan (2009), 'Reforms and Entry: Some Evidence from the Indian Manufacturing Sector', Review of Development Economics, 13(4): 658-72.

Blonigen, B.A., R.B. Davis, G.R. Waddell and H.T. Naughton (2007), 'FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment', European Economic Review, 51(5): 1303-25.

Bobonis, G.J. and H.J. Shatz (2007), 'Agglomeration, Adjustment, and State Policies in the Location of Foreign Direct Investment in the United States', Review of Economics and Statistics, 89(1): 30–43.

Brulhart, M, M. Jametti and K. Schmidheiny (2007), 'Do Agglomeration Economies Reduce the Sensitivity of Firm Location to Tax Differentials?', CEPR Discussion Paper No. 6606 (London: Centre for Economic Policy Research).

Carlton, D. (1979), 'Why New Firms Locate Where They Do: An Econometric Model', in W. Wheaton (ed.), Interregional Movements and Regional Growth (Washington, DC: Urban Institute).

Carlton, D. (1983), 'The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables', Review of Economics and Statistics, 65(3): 440-49.

Chakrabarti, A. (2001), 'The Determinants of Foreign Direct Investment: Sensitivity Analyses', Kyklos, 54(1): 89-113.

Chakraborty, C. and P. Nunnenkamp (2008), 'Economic Reforms, FDI and Economic Growth in India: A Sector Level Analysis', World Development, 36(7): 1192-1212.

Chakravorty, S. (2003), 'Industrial Location in Post-Reform India: Patterns of Inter-regional Divergence and Intra-regional Convergence', Journal of Development Studies, 40(2): 120-52.

Cheng, L.K. and Y.K. Kwan (2000), 'What Are the Determinants of the Location of Foreign Direct Investment? The Chinese Experience', Journal of International Economics, 51(2): 379-400.

Chung, W. and J. Alcácer (2002), 'Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States', Management Science, 48, 12, 1534-54.

Coeurdacier, N., R.A. De Santis and A. Aviat (2009), 'Cross-border Mergers and Acquisitions and European Integration', Economic Policy, 57: 55-106.

Coughlin, C.C. and E. Segev (2000a), 'Foreign Direct Investment in China: A Spatial Econometric Study', World Economy, 23(1): 1-23.

Coughlin, C.C. and E. Segev (2000b), 'Location Determinants of New Foreign-owned Manufacturing Plants', Journal of Regional Science, 40(2): 323-51.

Coughlin, C.C., J.V. Terza and V. Arromdee (1991), 'State Characteristics and the Location of Foreign Direct Investment within the United States', Review of Economics and Statistics, 73(4): 675–83.

Crozet, M., T. Mayer and J.-L. Mucchielli (2004), 'How Do Firms Agglomerate? A Study of FDI in France', Regional Science and Urban Economics, 34(1): 27-54.

Davis, J.C. and J.V. Henderson (2008), ‘The Agglomeration of Headquarters’, Regional Science and Urban Economics, 38(5): 445-60.

Devereux, M.P., R. Griffith and H. Simpson (2007), ‘Firm Location Decisions, Regional Grants and Agglomeration Externalities’, Journal of Public Economics, 91(3-4): 413-35.

Duranton, G., L. Gobillon and H. Overman (2006), ‘Assessing the Effects of Local Taxation Using Microgeographic Data’, CEPR Discussion Paper No. 5856.

Feenstra, R.C. and G.H. Hanson (1997), ‘Foreign Direct Investment and Relative Wages’, Journal of International Economics, 42(3): 371-93.

Fujita, M. and D. Hu (2001), ‘Regional Disparity in China 1985-1994: The Effects of Globalization and Economic Liberalization’, Annals of Regional Science, 35(1): 3-37.

Government of India (2006), Annual Survey of Industries 2003-04 (Chandigarh: Labour Bureau).

Guimaraes, P., O. Figueiredo and D. Woodward (2000), ‘Agglomeration and the Location of Foreign Direct Investment in Portugal’, Journal of Urban Economics, 47(1): 115-35.

Guimaraes, P., O. Figueiredo and D. Woodward (2003), ‘A Tractable Approach to the Firm Location Decision Problem’, Review of Economics and Statistics, 85(1): 201-04.

Guimaraes, P., O. Figueiredo and D. Woodward (2004), ‘Industrial Location Modeling: Extending the Random Utility Framework’, Journal of Regional Science, 44(1): 1-20.

Hansen W.G. (1959), ‘How Accessibility Shapes Land Use’, Journal of American Institute of Planners, 25: 73-6.

Head, C.K. and T. Mayer (2004), ‘Market Potential and the Location of Japanese Investment in the European Union’, Review of Economics and Statistics, 86(4): 959-72.

Head, C.K. and J.C. Ries (1996), ‘Inter-City Competition for Foreign Investment: Static and Dynamic Effects of China’s Incentive Areas’, Journal of Urban Economics, 40(1): 38-60.

Head, C.K., J.C. Ries and D.L. Swenson (1995), ‘Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States’, Journal of International Economics, 38(3-4): 223–47.

Head, C.K., J.C. Ries and D.L. Swenson (1999), ‘Attracting Foreign Manufacturing: Investment Promotion and Agglomeration’, Regional Science and Urban Economics, 29(2): 197-218.

Holl, A. (2004), ‘Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain’, Regional Science and Urban Economics, 34(3): 341-63.

Kathuria, V. (2002), Liberalisation, FDI, and Productivity Spillovers: An Analysis of Indian Manufacturing Firms’, Oxford Economic Papers, 54(4): 688-718.

Klapper, L., L. Laeven and R. Rajan (2006), ‘Entry Regulation as a Barrier to Entrepreneurship’, Journal of Financial Economics, 82(3): 591-629.

Kochhar, K., U. Kumar, R. Rajan, A. Subramanian, and I. Tokatlidis (2006), ‘India’s Pattern of Development: What Happened, What Follows?’, Journal of Monetary Economics, 53(5): 981-1019.

Lall, S.V. and S. Chakravorty (2005), ‘Industrial Location and Spatial Inequality: Theory and Evidence from India’, Review of Development Economics, 9(1): 47-68.

Ledyaeva, S. (2009), ‘Spatial Econometric Analysis of Foreign Direct Investment Determinants in Russian Regions’, World Economy, 32(4): 643-66.

List, J.A. (2001), ‘US County-level Determinants of Inbound FDI: Evidence from a Two-step Modified Count Data Model’, International Journal of Industrial Organization, 19(6): 953-73.

Majumder, R. (2008), Infrastructure and Development in India: Interlinkages and Policy Issues (Jaipur: Rawat Publications).

Marshall, M.G. and D.R. Marshall (2008), ‘Crime in India: Annual Series, 1954-2006, Electronic Dataset and Codebook’, Center for Systemic Peace, http://www.systemicpeace.org/inscr/inscr.htm (accessed: March 2011).

McDowell, A. (2003), ‘From the Help Desk: Hurdle Models’, Stata Journal, 3(2): 178-84.

McFadden, D. (1974), ‘Conditional Logit Analysis of Qualitative Behaviour’, in P. Zarembka (ed.), Frontiers in Econometrics (New York: Academic Press).

Nunnenkamp, P. and R. Stracke (2008), ‘Foreign Direct Investment in Post-Reform India: Likely to Work Wonders for Regional Development?’, Journal of Economic Development, 33(2): 55-84.

Papke, L.E. (1991), 'Interstate Business Tax Differentials and New Firm Location: Evidence from Panel Data', Journal of Public Economics, 45(1): 47-68.

Purfield (2006), 'Mind the Gap - Is Economic Growth in India Leaving Some States Behind?', IMF Working Paper WP/06/13 (Washington, DC: International Monetary Fund).

Sachs, J.D., N. Bajpai, and A. Ramiah (2002), 'Understanding Regional Economic Growth in India', Asian Economic Papers, 1(3): 32-62.

Schmidheiny, K. and M. Brulhardt (2009), 'On the Equivalence of Location Choice Models: Conditional Logit, Nested Logit and Poisson', CESifo Working Paper No. 2726 (Munich: CESifo).

UNCTAD (1998), World Investment Report 1998 (New York and Geneva: United Nations).

UNCTAD (2004), World Investment Report 2004 (New York and Geneva: United Nations).

UNCTAD (2009), World Investment Report 2009 (New York and Geneva: United Nations).

World Bank (2004), India: Investment Climate Assessment 2004: Improving Manufacturing Competitiveness (Washington, DC).

Zhang, X. and K.H. Zhang (2003), 'How Does Globalisation Affect Regional Inequality Within a Developing Country? Evidence from China', Journal of Development Studies, 39(4): 47-67.

Tables

Table 3.A.1: Summary of Empirical Literature

Study	**Country**	**Sample**	**Methodology**	
			clogit	*poisson*
Carlton (1983)	USA	528 new firms; 1967-1971	√	
Papke (1991)	USA	8.3 million establishments; 1975-1982		√
Head, Ries and Swenson (1995)	USA	751 new firms; 1980-1987	√	
Becker and Henderson (2000)	USA	641 new births; 1963-1992		√
Guimaraes, Figueiredo and Woodward (2000)	Portugal	758 greenfield investments; 1982-1992	√	
List (2001)	California, USA	67 greenfield investments; 1983-1992		√
Head and Mayer (2004)	EU	452 firms; 1984-1995	√	
Crozet, Mayer and Mucchielli (2004)	France	3,902 firms; 1985-1995	√	
Guimaraes, Figueiredo and Woodward (2004)	USA	65,158 firms; 1989-1997		√
Holl (2004)	Spain	122,000 new plants; 1980-1994		√
Duranton, Gobillon and Overman (2006)	UK	21,813 new firms; 1984-1989		√
Brulhart, Jametti and Schmidheiny (2007)	Switzerland	13,768 new firms (1999-2002), and 12,465 new firms (2001-2002)		√
Devereux, Griffith and Simpson (2007)	UK	79,337 greenfield investments; 1986-1992	√	
Arzaghi and Henderson (2008)	New York county	502 new advertising firms; 1992-1997		√
Davis and Henderson (2008)	USA	11,990 new HQ firms; 1977-1997		√
Coeurdacier, De Santis and Aviat (2009)	EU	73 per cent of all M&As; 1985-2004		√

Table 3.A.2: Descriptive Statistics

Variable	Expected sign	#			Mean		
		1991	*1996*	*2001*	*1991*	*1996*	*2001*
Investment decisions* (new/cumulative)		1,234	1,669	1,905	-	4,634	12,927
HI	-			533			0.34
MA	+	434	423	530	224,984	240,099	231,256
Population	+	406	408	533	1,982,719	2,075,529	1,926,232
Wages	-			454			89.54
Electricity	+	334	350	533	0.84	0.88	0.56
Telephone	+	126	117	533	0.46	1.12	0.09
Education	+	231	305	533	0.17	0.22	0.06
Buses	+			533			0.50
Roads	+	207	184		61.25	69.30	
Banks	+	393	418	415	7.60	7.32	6.95
Health	+	167			2.59		
Labour Regulations	+	15	15	15	0.25	0.25	0.25
Riots	-	385	383	480	0.00012	0.00010	0.00008
FA	+		244	530		25.06	32.84

*Reference years: 1991, 1992, and 1993 (for 1991); 1996, 1997, and 1998 (for 1996); 2001, 2002, and 2003 (for 2001).
Note: # refers to the number of districts for which there are observations.

Table 3.A.3: Characteristics of the Data (2001 cross-section)

Variable	#	Mean	Std. Dev.	Min.	Max.
count	533	10.96	90.91	0	1,289
count>0	164	35.64	161.52	1	1,289
obs	533	0.6923	0.4619	0	1
fitp	423[1]	0.5835	0.3498	0	0.9996
fitnb	423	0.6385	0.3225	2.62E-06	0.9998
fitzip	423	0.6484	0.3241	0.0014	0.9999

[1] The number of observations is less than the number of cases in the dataset owing to missing values for some variables in the model.

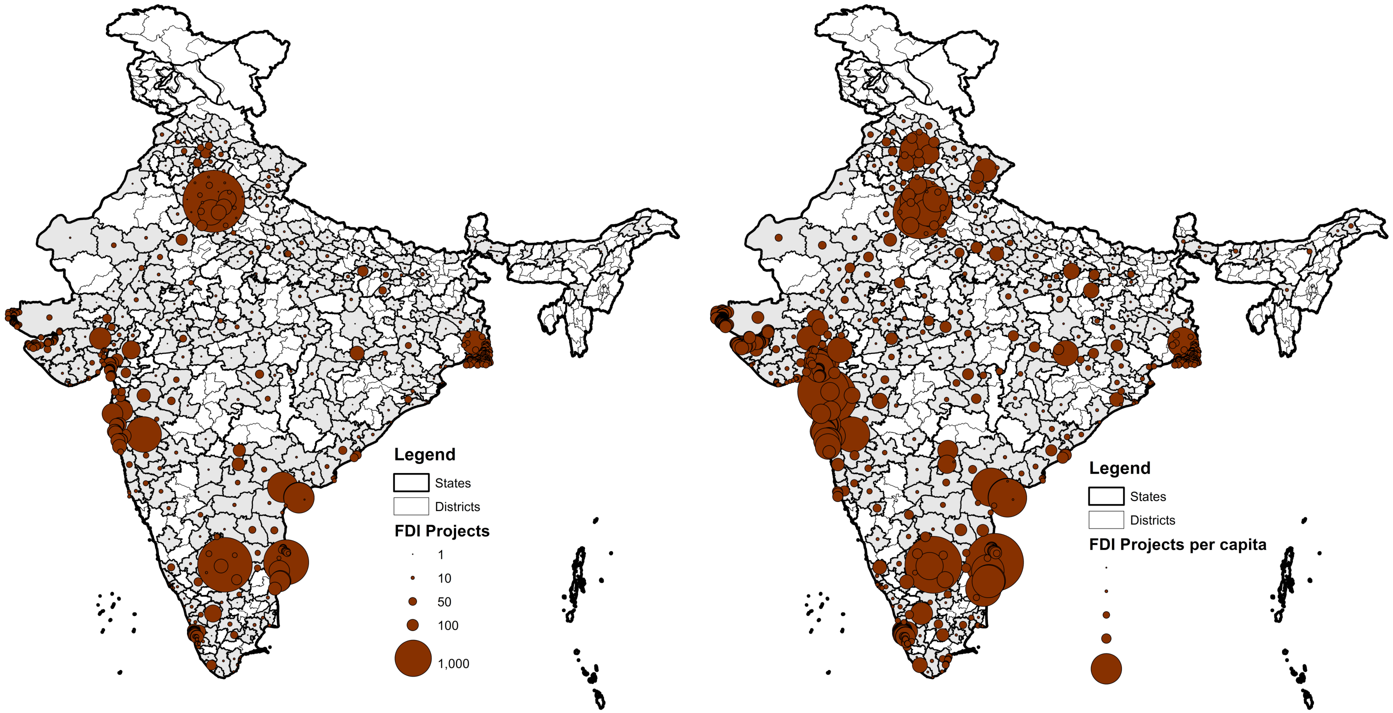
Table 3.A.4: Incidence Rate Ratios (2001 cross-section), negative binominal model for sub-samples

Variable	Secondary	Tertiary	Equity>50%	Equity<50%	Less than median	More than median
Population	0.4521***	0.6482***	0.1123	0.1311	0.1453*	0.1025
FA	0.0962***	0.0497***	0.7672***	0.6915***	0.7224***	0.7833***
HI	0.0237	-0.1035	-0.0249**	-0.0407	-0.0060	0.2303
MA	0.2296	-0.2517	-0.0091	-0.2470**	-0.2982**	-0.0499
Electricity	0.5665***	0.3277	-0.0735	-0.0234	-0.0234	-0.2234
Telephone	-0.1066	-0.1008	-0.0063	0.0202	-0.0256	-0.1315
Education	0.1621	0.3187**	0.5155***	0.2538*	0.1602	0.4772***
Buses	0.4429*	0.4847*	0.4397**	0.4634*	0.3737	0.8373***
Banks	0.3204**	0.5356**	0.0467	0.2578*	0.3579**	0.2481**
Wages	0.1422	-0.3765**	-0.1209	-0.1625**	-0.1986**	-0.1377
Labour Regulations	-0.0838	-0.1571	0.4321	0.2015	0.1049	0.1579
Riots	0.1099	0.1229**	-0.0363	0.0133	0.0748	-0.0166
#	331	300	72	83	77	81
AIC	778	580	367	434	408	402
BIC	831	632	399	468	441	436

Robust standard errors clustered at district level (not shown). Non-Exponentiated coefficients
* p<0.05, ** p<0.01, *** p<0.001

Figures

Figure 3.A.1: Spatial Distribution of FDI Projects



Source: Department of Industrial Promotion and Policy, Ministry of Commerce and Industry

CHAPTER 4

Does Agglomeration boost Innovation? An Econometric Evaluation*

* I am grateful to Eckhardt Bode, Wan-Hsin Liu, Ken Shadlen, Elizabeth Mack and participants at the Tinbergen Institute and the Martin Prosperity Institute for helpful comments and suggestions. (Revise and Resubmit, Spatial Economic Analysis).

Abstract

Innovation is crucial to regional economic competitiveness and to productivity growth. A salient feature of the Indian economy is the geographic clustering of both, economic activity and innovation. In this paper, I study to what extent the spatial distribution of economic activity drives innovation. I analyse patent applications between 1999 and 2007 across districts in India and my econometric findings suggest that R&D expenditures, industrial diversity and the distribution of human capital endowments can have an important effect on generating innovation. The estimates are robust to omitted variables bias, to different model specifications and to the type of applicant.

JEL Classification: R12, O3
Keywords: Innovation, agglomeration economies, human capital

4.1 Introduction

Economists and urban geographers have long noted the co-location of clusters of economic activity and of innovative activity. Theory dictates that people and organisations choose to live close to one another mainly because it is expensive to transport goods, people and ideas across space. Historically, transportation costs have been falling and the advent of new information and communication technologies led many to predict the 'death of distance'. On the contrary, economic activity continues to be concentrated and agglomerations continue to exist, and grow in size. If it is no longer as costly to transport goods, people and ideas from one place to the next, why do people choose to agglomerate?

It was Marshall (1919) who was the first to formally identify the benefits of agglomeration. Clustering of economic activity allows for the sharing of inputs, skilled labour and knowledge – producers can access both buyers and suppliers more easily and at lower costs, and can exploit the presence of a thick labour market with multiple specialisations. Agglomeration also facilitates face-to-face contact by shortening interaction distances, which results in knowledge spilling over, leading to innovative activity.

Romer (1986, 1990) developed a theory to explain economic growth by making endogenous the effect of the accumulation and spillover of technological knowledge. This seminal contribution provided an explanation of economic growth that went beyond a simple focus on the role of investment in physical capital and increases in the labour supply. Changes in technology constitute an important source of new knowledge.

And lastly, Feldman and Florida (1994) suggest that innovation is 'increasingly dependent on a geographically defined infrastructure that is capable of mobilising technical resources, knowledge, and other inputs essential to the innovation process'. Examples of such innovation-supporting advantages include the presence of networks of industry, concentration of Research and Development (R&D) activity, be it in the private or in the public sector, and the presence of a large pool of service activities with multiple specialisations.

In this paper, I employ a regional knowledge production framework to empirically identify the effect of agglomeration economies, knowledge spillovers and infrastructure on innovative output. I find that whilst R&D expenditures impact innovation, in fact clustering of innovative activity and factor inputs within a location also matter significantly. Section 4.2 outlines the literature and discusses the main focus of the most relevant papers. Section 4.3 provides a descriptive overview of patent activity in India. Section 4.4 describes the empirical model and the sources and description of the data. This is followed by a summary of the study results and a discussion of the key findings in Section 4.5. Section 4.6 addresses possible endogeneity concerns and carries out robustness checks, and the paper concludes with a discussion of key contributions and limitations.

4.2 Related Literature

The idea that clustering creates externalities that could lead to an increase in innovation and which in turn could drive economic growth, is not a new one – for instance, see Krugman 1991, Kelly and Hageman 1999, Paci and Usai 1999 and Hanson 2001. The literature has dealt jointly with the concepts of agglomeration and innovation under the general research umbrella of economic geography. Different studies have defined their clusters of observations differently – as industrial districts, technological clusters, learning regions, and innovation milieus (for an overview of these studies see Ibrahim 2009).

Griliches (1979) proposed a knowledge function approach at the level of the firm – a Cobb-Douglas function that measured a firm's innovative output tin terms of physical R&D inputs and some indicators of the knowledge stock. Jaffe (1986) was the first to use such a knowledge production function framework at the regional level[77], within which he measured the effect of knowledge spillovers on innovative output. The knowledge production approach has since been widely used in the empirical literature – see Table 4.A.1 for an overview of some of these studies.

[77] Pakes and Griliches (1980) used the knowledge production framework for patenting data, but at the level of the firm.

The empirical literature has many findings regarding the effect of regional differences on innovative output. For instance, some studies find evidence of a positive impact of localised knowledge spillovers resulting mainly from the presence of academic research (Anselin et al 1997, Fischer and Varga 2003, Fristch and Slavtchev 2007, Kantor and Whalley 2009, Ponds 2010). Others concern themselves with the effects of neighbours, i.e. the economic geography impact of R&D investments in surrounding regions (Bottazzi and Peri 2003, Bode 2004). Some authors, those who are able to access superior data, use patent citations to estimate the extent of knowledge spillovers (Jaffe et al 1993, Co 2005). A number of studies also draw out the difference between diversity and specialisation within a location and the associated impact on innovation (Feldman and Audretsch 1999, Carlino et al 2007).

Most studies have made use of data from the United States (Audretsch and Feldman 1996, Anselin et al 2000, Rodriguez-Pose and Crescenzi 2008, Crescenzi et al 2007[78]) or from the European region (Autant-Bernard 2001, Fischer and Varga 2003, Castro and Quevedo 2005, Fritsch and Slavtchev 2007). In short, there are very few studies that deal with developing countries, and none that analyse India. This paper attempts to fill this gap in the literature. One might expect that the dynamics of innovation and the relationship between R&D and patenting may differ in the context of developing countries. For instance, small and medium-sized firms in India can scarcely afford in-house R&D for technological innovations, and the spilling over of the industry-level stock of knowledge may matter more. However, one may expect that customer-driven innovations would depend on market demand for product diversification, and associated patenting activity may not differ from that in developed economies. Cumulative innovative processes could be internal to the firm or to the industry, but they could also depend crucially on institutional factors, which may be region or country-specific. The dynamics of innovative activity would depend on a country's labour markets, its legal systems and the quality of its infrastructure, factors which may differ between developing and developed countries, and even within countries. Ayyagari et al (2007) show that innovation varies greatly across countries, and that factors such as access to finance, governance, competition and educational attainment all remain important to explain differences across firms and regions within countries. Thus, one might expect

[78] Free availability of US patent data from the NBER website has helped to spur multiple empirical studies based on the data.

innovative activity to be more significantly affected by characteristics of the location and the workforce for firms and industries in developing countries.

4.3 Patents in India

Patenting activity is considered a yardstick for innovative activity within a country, and in keeping with the literature, this paper uses patents as a proxy for the outcome of the inventive process. Data on patents (applications) is taken from the Indian Patent Office (IPO). The IPO registers patent applications at its four branches located in Kolkata (West Bengal), Chennai (Tamil Nadu), Mumbai (Maharashtra) and Delhi (Delhi). Although it is easier to aggregate data on the basis of which office an application was made to, the analysis within this paper requires data on the address of the applicant, or the location of the organisation or firm, at the level of the district. Data disaggregated at the level of the state is made available in the annual reports of the IPO. However, information on the exact location of the applicant is only contained within the weekly journals published by the IPO[79]. Each patent file contains information on the following variables of interest: (1) date of application (2) the name of the inventor(s) (3) the address of the inventor(s) and, (4) the International Patent Classification (IPC) code.

Most patent jurisdictions publish a patent application 18 months after its original filing date, after which the application is considered to be in the public domain. Before this publication, the content and the very act of filing are considered proprietary information that is closely guarded by most filers. Although applications from residents within the country have increased steadily over the last 10 years or so, applications made by foreigners have accounted for the bulk of increase in total applications (see Figure 4.A.1). Between 1995 and 2008, a total of 189,577 patents were filed with the IPO, of which applications from foreigners accounted for 78.11%. According to the IPO's 2007-2008 Annual Report, the total patents in force, as of 31 March 2008, were 29,688, of which 7,966 comprised of patents granted to applicants residing in India. For a descriptive overview of pre-1995 patenting trends in India, see Rajeswari (1996). In this

[79] These weekly journals can be accessed from: http://www.ipcindia.nic.in/

paper, I consider only those applications that are made by those residing within the country.

The first reference line (year=1999) in the figure refers to the onset of the India-Pakistan war, and the second reference line (year=2005) refers to the year in which the Indian Patents Act (1970) was in full compliance with the Trade-Related Intellectual Property System (TRIPS) agreement of the WTO. Amendments to the Indian Patent Act were made in 1999, 2002, 2005 and finally in 2006 in anticipation of the country's obligations under the TRIPS agreement. Inventors, whether domestic or foreign, who were looking to file patent applications over this time period would have experienced a changing environment – for instance, pharmaceutical firms were not permitted to file product patents up until 2005. However, since this analysis does not deal with cross-country comparisons and is mainly concerned with the concentration of patenting activity within the country, changes in the national environment should not have differential geographical impacts. However, there is evidence to suggest that the impact may differ across industrial sectors. For instance, as expected, patent applications from pharmaceutical firms tripled from 765 in 2004-05 to 2,373 in 2008-09 (total applications doubled over the same period)[80]. As a check, I include industry fixed effects in the empirical model.

Patents are classified according to the International Patent Classification system, which is technology based rather than product based. Figure 4.A.2 provides an example of the different categories for patent applications for the latest year for which data is available. The product classification system limits the economic usefulness of the data, since it does not automatically allow an analysis of innovation by industry or sector. However, I match the IPC classifications with 2-digit industry level classifications, and am able to explore the impact of industry agglomeration on innovative output. The one-to-one matching is provided in Table 4.A.5. I carry out the matching by comparing the 2-digit National Industrial Classification (NIC) provided by the Indian government with the detailed IPC classification provided by the World Intellectual Property Organisation[81].

[80] See: http://www.nipo.in/nipoinnews.html
[81] See: http://www.wipo.int/classifications/ipc/ipc8/?lang=en

In addition, I also differentiate between different types of applicants - from public and private funded research organisations, universities, and from private firms, industry and individuals. I use these applicant classes as a further robustness check later in the paper. One of the shortcomings of the IPO data is that it does not differentiate patent applications made by individual laboratories of the Council for Scientific Research (CSIR) and lists the organisation's headquarters in Delhi as the default location for all such applications. I am thus forced to drop these observations, and then re-instate them using data from the CSIR which disaggregates applications made on the basis of the location of the laboratory [82].

Because patents contain geographic information about their inventors, I make use of this data to study the factors affecting the generation of knowledge in these locations. Patent activity in India shows signs of high levels of spatial concentration (see Figure 4.A.3). The maps provide an actual representation of the density of patent applications for the country – the size of the circle is proportional to the number of applications within the district. The map indicates that whilst some districts in the country host a lot of patent activity, others are virtually empty. Of the possible 604 districts, patent activity is present in only 190 districts. And of those, around 22 districts account for 90% of all patenting activity. It could be argued that the high degree of clustering of activity is explained mainly by the size of the district – after all districts like Mumbai, Delhi, Chennai, Hyderabad, Bangalore etc account for the lion's share of patents. However, the map on the right depicts patenting activity after controlling for the size of the district – in other words, the size of circle is proportional to the patents per capita within a district. Two results emerge: relative clustering increases, and clustering now spreads to neighbouring regions. Patenting per capita is higher in districts in the southern and eastern parts of the country. There is also evidence to show that the rate of patenting activity is high in satellite towns and cities such as Noida, Pune, Ghaziabad, Gurgaon, Vadodara, Coimbatore etc. In general, keeping in mind the simplest no-clustering (uniform distribution) benchmark, there is evidence of concentration of patent activity in the country.

[82] The full list of CSIR laboratories can be found on the website of the Department of Scientific and Industrial Research: http://www.dsir.gov.in/a_report/english/2005-06E/Annexures-8.pdf.

4.4 Empirical Model

I test the importance of geography and human capital for knowledge spillovers by implementing an extended version of the Griliches-Jaffe regional knowledge production function at the district level in India.

At the level of the firm, such a production function assumes that there exists a stable relationship between R&D investments by a firm and the production of economically useful knowledge. However, firm-level studies are unable to take into account the spillovers received by a firm from its geographical region and investments, which are made in one sector but can often spillover into other sectors, such that the total exceeds the sum of the individual components. The empirical link between knowledge inputs and innovative output becomes stronger with higher aggregation of the unit of observation, from the firm to the industry to the region.

What constitutes the right level of aggregation? A number of authors have pointed out that the effect of knowledge spillovers remains relatively localised and begins to diminish with increasing distance (see Jaffe et al 1993, Varga 1998, Acs et al 2002 for the United States, and Autant-Bernard 2001, Fischer and Varga 2003 for Europe). Whilst it is possible that the marginal cost of transmitting information across geographic space has been rendered invariant by the telecommunications revolution, the marginal cost of transmitting knowledge, and especially tacit knowledge, rises with distance (Audretsch 1998). This argues for as small a geographical unit of analysis as possible. Previous studies have used 'states' (Jaffe, 1989; Smith, 1999) or Metropolitan Statistical Areas (Anselin et al., 1997, 2000), or 'districts' (Keeble and Wilkinson 1999; Piergiovanni and Santarelli 2001) as their boundaries. Some studies approached comparison and analysis at the national level (UK Department of Trade and Industry 2001), where 'cluster' refers to every technological concentration within the country (Ibrahim 2009).

If I were to choose the state as the unit of observation, I would be able to expand my sources of data to patent applications made to the US and the European Patent Offices from applicants based in India. Krugman (1991) has emphasised that 'states aren't really the right geographical units', and although he was referring to states within the

United States, it does not make much sense to compare states like Uttar Pradesh (population: 190 million, area: 243 thousand km^2) with states such as Chandigarh (population: 900 thousand, area: 114 km^2) and city states such as Delhi (population: 12 million, area: 1,484 km^2). Thus, a better, albeit still far from perfect, unit of observation in this case would be the district. This assumes that knowledge flows are bounded within a relatively narrow geographical range.

In the typical regional production function approach, the innovative output of a region depends upon the level of R&D (measured either by employment or expenditures) within the region and in neighbouring regions. Agglomeration within a location is also theorised to be a catalyst to innovation, and I include economic geography variables to capture agglomeration at different levels. A vector of local economic characteristics that could contribute to the generation of new knowledge is also often included. These include the level of human capital, the presence of scientists and engineers, industrial diversity, the quality of technological infrastructure within the location etc. Following other empirical studies (Anselin et al 1997, Feldman and Audrestch 1999, Del Barrio-Castro and Garcia-Quevedo 2005, and Ponds 2010), this paper will use a count data model within the knowledge production framework.

Based on these specifications, the following model is estimated:

(I) $$P_{jkt} = \beta_0 + \beta_1 RD_{jkt} + \beta_2 Access_{jt} + \beta_3 \sigma_{jkt} + \beta_4 HI_{jt} + \beta_5 Edu_{jt} + \beta_6 X_{jk} + \delta_k + \gamma_t + \varepsilon_{jkt}$$

Where j indexes the districts, k indexes the IPC-industry classification and t refers to the years of observation. RD_{jkt} refers to private research and development expenditures within the district, $Access_{jt}$ refers to R&D expenditures in neighbouring districts, σ_{jkt} is a measure of agglomeration by industrial sector and district, HI_{jt} is a measure of industrial diversity in the district, Edu_{jt} is a measure of human capital, X_{jt} refers to other district-level characteristics, δ_k and γ_t refer to industry and year fixed effects respectively, and ε is a stochastic error term.

The dependent variable is the count of new patent applications within a district, which varies by industry and year. As mentioned before, these patent counts are taken from weekly journals of the Indian Patent Office. Only domestic applicants are included in the analysis, specifically when the address of the applicant is within the territory of India. The location of the patent refers to the location of the patent applicant, and not to the location of the firm – this is to avoid biases resulting from centralised patent applications from companies with multiple branches in different locations. Patents are assigned to the address of the first inventor named on the patent. The total patent applications gleaned from weekly journals total 15,782 applications made in 190 districts belonging to 35 states and union territories from 1999-2008. The choice of years of study is dictated by data limitations. Whilst annual patent application data exists for the period 1999-2008, I have data on district-level location variables for only a few years. Economic geography and infrastructure data is mostly taken from household-level surveys conducted by the NSSO. I use data from the Employment and Unemployment surveys: Round 55.10 (July 1999 – June 2000) and Round 61.10 (July 2004 – June 2005). This effectively restricts the sample to two years: 1999 and 2004. The sources of data used to construct the predictor variables are summarised in Table 4.1.

Industrial R&D refers to firm-level expenditures on research activities. The data is reported in Indian Rupee Crores[83]. This data is taken from the Prowess database and is then aggregated across different districts in India. Prowess is a corporate database that contains normalised data built on a sound understanding of disclosures of over 18,000 companies in India. The database provides financial statements, ratio analysis, fund flows, product profiles, returns and risks on the stock market etc. Unfortunately the data does not include information on R&D contracted to universities and to research institutions.

The accessibility indicator is defined as the potential for opportunities for R&D expenditures in neighbouring regions to spill over to the location:

$$Access_j = \sum_m \frac{RD_m}{d_{j-m}^b}$$

[83] 1 Crore = 10 Million

Where, $Access_j$ is the accessibility indicator estimated for district j, RD_m is the R&D expenditures in neighbouring districts m[84], d_{jm} is a measure of distance between the district j and its neighbour m, and b describes how increasing distance reduces the expected level of interaction[85]. The exponent value is an indicator of how distance is a restrictive factor, and in the simple model, accessibility is estimated with the exponent value set to 1. The accessibility measure is constructed by allowing transport to occur along the orthodormic distance connecting any two districts. Instead of calculating the distance between any pair of districts across the country, I restrict the links to districts within a 500-kilometre radius.

Localisation economies (σ_{jk}) can be measured by own industry employment in the region, own industry establishments in the region, or an index of concentration, which reflects disproportionately high concentration of the industry in the region in comparison to the nation. I measure localisation economies as the proportion of sector k's employment in district j as a share of all of sector k's total employment in the country. The higher this value, the higher the expectation of intra-industry concentration benefits in the district.

$$\sigma_{jk} = \frac{E_{k,j}}{E_k}$$

Since the data provided by the IPO lists the IPC classification, which is mainly a technology-based categorisation, it was necessary to relate this to the Indian National Industrial Classification (NIC) system to enable the computation of industrial agglomeration. A broad overview of the matching is presented in Table 4.A.5.

A higher level of industrial diversity may also translate into the presence of a wider selection of producer services essential to innovation, such as information technology, legal, marketing services etc. I use the Herfindal measure to examine the degree of economic diversity, as a measure of urbanisation in each district. The Herfindal index for district j (HI_j) is the sum of squares of employment shares of all industries in district j:

84 I am grateful to Eckhardt Bode for providing me with the syntax for computing the great-circle (orthodormic) distance calculations.

85 In the original model proposed by Hansen (1959), b is an exponent describing the effect of the travel time between the zones.

$$HI_j = \sum_k \left(\frac{E_{jk}}{E_j} \right)^2$$

Unlike measures of specialisation, which focus on one industry, the diversity index considers the industry mix of the entire regional economy. The largest value for HI_j is one when the entire regional economy is dominated by a single industry. Thus a higher value signifies lower level of economic diversity.

It is especially important to control for local inputs into the R&D process. Skilled workers endowed with a high level of human capital are a mechanism by which economic knowledge is created and transmitted. I include the share of the population with a higher education (defined as a high school degree or more) as a proxy for the general quality of human capital. However, more specifically, I also include the proportion of the population that possess a degree in a scientific subject – defined as agricultural sciences, engineering, medicine etc. Skilled workers endowed with a high level of human capital are a mechanism through which knowledge externalities materialise, and I would expect these to have a positive effect on the generation of patents.

Non-agricultural hourly wage rates are used as an indicator of labour input costs. The expected effect of this variable is hard to pin down theoretically. On the one hand, one would expect innovative activity to be inversely related to labour costs, since high costs within a location could drive down productivity. On the other hand, since I am unable, at this stage of the analysis, to differentiate wages on the basis of the skill set of workers, it is possible that there would be a positive effect since high wages are in effect accounting for the presence for highly skilled labour in the workforce.

Other district-level characteristics include variables which proxy for quality of infrastructure, in particular access to electricity and telephones. Descriptive statistics of the variables used is provided in Table 4.A.2. Please note that all covariates in the empirical model are lagged by one period in order to mitigate potential endogeneity concerns.

Table 4.1: Predictor Variables

	Variable	Indicator	Source(s)	Availability 1999-2000	2004-2005
R&D	R&D	Private R&D expenditures	Prowess	√	√
Economic Geography	Access	Neighbouring R&D	Orthodromic distance calculations	√	√
	HI	Economic Diversity	NSSO	√	√
	σ	Localisation Economies	NSSO	√	√
Infrastructure	Education	Proportion of population with a High-School education	NSSO	√	√
	Technical Education	Proportion of population with a technical diploma/degree	NSSO	√	√
	Wages	Non-agricultural hourly wages	NSSO	√	√

Notes: IPO: Indian Patent Office; NSSO: National Sample Survey Organisation

4.5 Results and Discussion

I illustrate the key characteristics of the data and the subsequent modelling choices, by using the data from 1999 as an example. One of the key characteristics of the data is that it is over-dispersed. In 1999, the mean number of patent applications per district is around 6.8, the standard deviation is over 75, i.e. over 10 times the mean (see Table 4.A.3). A Poisson model implies that the expected count, or mean value, is equal to the variance. This is a strong assumption, and does not hold for my data.

A frequent occurrence with count data is an excess of zeroes compared to what would be expected under a Poisson model. This is indeed a problem faced by this data – the mean number of non-zero patent counts is around 31 and the standard deviation is 161, i.e. around 5.2 times the mean. Also note that of 588[86] districts, a total of 462 districts have zero patent applications. This implies that one would need to take into account, both, over-dispersion and the excess of zeroes in the data, when selecting a model to fit the data[87].

Another way to reiterate the unsuitability of the Poisson model in this case is to show that such a model is unable to predict the excess zeroes found in the data. In Table 4.A.3, obs refers to actual observations in the data, and Fit_P, Fit_NB and Fit_ZIP refer to the predictions of the fitted Poisson, negative binomial and zero-inflated Poisson models respectively. It is clear that 78.57% of the locations in the sample have no patent applications, but the Poisson model predicts that only 47.97% would make zero patent applications. Clearly the Poisson model underestimates the probability of zero counts. The negative binomial model, which allows for greater variation in the count variable than that of a true Poisson, predicts that 75.9% of all districts will make no patent applications, much closer to the observed value.

One way to account for the excess zeroes would be to assume that the data comes from two separate populations, one where the number of investments is always zero, and another where the count has a Poisson distribution. The distribution of the outcome is

[86] Although there are a total of 604 districts in India, I exclude all districts for which I do not have data for the regressors.
[87] I am unable to use a log-linear specification since it would result in a loss of all zero observations that constitute an important part of the data.

then modelled in terms of two parameters – the probability of always zero and the mean number of patent applications for those locations not in the always zero group. The zero-inflated Poisson model (fit_ZIP) predicts that 76.36% of all locations will not apply for patents, marginally better than the predictions of the negative binomial model.

An alternative approach to deal with an excess of zeroes would be to use a two-stage process, with a logit model to distinguish between the zero and positive counts, and then a zero-truncated Poisson or negative binomial model for the positive counts. In the case of this paper this would imply using a logit model to differentiate between districts that make no patent applications and those that do, and then a truncated model for the number of districts that apply to at least one patent. These models are referred to as "hurdle models" – a binary probability model governs the binary outcome of whether a count variate has a zero or positive realisation; if the realisation is positive, the 'hurdle' is crossed and the conditional distribution of the positives is governed by a truncated-at-zero count model data model (McDowell 2003)[88].

The response variable is 'count', i.e. the number of patent applications per district. The Poisson regression models the log of the expected count as a function of the predictor variables. More formally, $\beta = \log(\mu_{x+1}) - \log(\mu_x)$, where β is the regression coefficient, μ is the expected count and the subscripts represent where the regressor, say x, is evaluated at x and $x+1$ (here implying a unit percentage change in the regressor[89]). Since the difference of two logs is equal to the log of their quotient, i.e. $\log(\mu_{x+1}) - \log(\mu_x) = \log(\frac{\mu_{x+1}}{\mu_x})$, thus one could also interpret the parameter estimate as the log of the ratio of expected counts. In this case, the count refers to the 'rate' of patent applications per district. Table 4.2 provides the results using different types of count models.

The coefficients can be interpreted as follows – if R&D expenditures were to increase by one unit (i.e. by INR 1 Crore), the expected number of patent applications would increase by 0.24 percent. On the other hand, if the Herfindahl Index were to increase by a unit (implying a reduction in industrial diversity), the expected count of patent

[88] I was unable to achieve convergence when using the zero-inflated and the zero-truncated negative binomial models, and these results are excluded from the paper.
[89] This is because the regressors are in logarithms of the original independent variables.

applications would decrease by 528 percent. A unit increase in the percent of the population with a technical degree and with a high-school degree is associated with an expected increase in patents of 25 and 10 percent respectively - see the coefficients for the Poisson model in Table 4.2).

As the model selection criteria I also examine and compare the Bayesian information criterion (BIC) and Akaike's information criterion (AIC). Since the models are used to fit the same data, the model with the smallest values of the information criteria is considered better. These regressions also control for industry and year, and so I rule out the possibility that these results are driven by changes specific to any one type of industry, or to any one of the two years in the pooled dataset.

The first result to note is that the effect of R&D expenditures remains positive and significant, irrespective of the model used – see Table 4.2. The main economic geography variables are localisation and industrial diversity, while market access is a measure of the effect of neighbouring regions. Notably, the effect of the agglomeration measure – localisation – is positive and significant across most models. This implies that an increase in clustering of firms within a given industry is associated with a positive rise in innovative output in that industry. Recall that the Herfindahl Index (HI) is a measure of the level of industrial diversity within the district. A higher HI implies higher employment concentration by one industry and lower industrial diversity, and vice versa. Thus, the negative coefficient for HI is evidence of a positive association between more industrial diversity and more innovation.

Table 4.2: Model Comparisons

Variable	Poisson	Negative Binomial	Zero-Inflated Poisson	Zero-Inflated Negative Binomial	Zero-Truncated Poisson	Zero-Truncated Negative Binomial
R&D	0.0024***	0.0022**	0.0022***	0.0033**	0.0022***	0.0042*
	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.002]
Access	0.3321*	-0.1077	0.2921*	0.0726	0.2976*	0.1191
	[0.191]	[0.197]	[0.176]	[0.167]	[0.177]	[0.174]
HI	-5.2841**	-3.6493***	-1.7552	-1.2203	-1.7516	-1.4180
	[2.085]	[1.002]	[1.925]	[1.938]	[1.935]	[2.004]
Localisation	3.2190***	3.6905***	2.2987***	3.4353	2.2927***	2.8107
	[0.884]	[5.107]	[0.816]	[2.673]	[0.814]	[5.126]
Education (technical)	0.2540***	0.2257**	0.1861***	0.0710	0.1861***	0.0619
	[0.045]	[0.093]	[0.042]	[0.109]	[0.042]	[0.165]
Education (high-school)	0.1058***	0.2377***	0.0550**	0.1426***	0.0552**	0.1274***
	[0.027]	[0.047]	[0.026]	[0.042]	[0.026]	[0.045]
Wages	-0.0030	-0.0047**	-0.0008	-0.0028**	-0.0008	-0.0038**
	[0.004]	[0.002]	[0.003]	[0.001]	[0.003]	[0.002]
Industry FEs	√	√	√	√	√	√
Year FEs	√	√	√	√	√	√
#	16,375	16,375	16,375	16,375	2,745	2,745
AIC	147296	27536	99722	25999	88974	15343
BIC	147420	27667	99961	26246	89069	15444

Robust standard errors clustered at district level.
*** p<0.01, ** p<0.05, * p<0.1

The effect of 'Access' is positive and significant, albeit only at the 10% level, in the case of the Poisson models. This positive effect could be seen as evidence of spilling over of expenditures in one region positively affecting innovation in others. As a measure of human capital within a location, I also include the percentage of the district population with a high-school degree or higher, and with a technical degree (i.e. a degree in agricultural sciences, engineering, medicine etc). Whilst the coefficient of the proportion with a high school education (or higher) is positive and significant across all models, that of technical education is positive and significant for most models with the exception of the zero-inflated and zero-truncated negative binomial. It is interesting that the magnitude of the effect of the percentage of the population with a technical degree, wherever statistically significant, outstrips the effect of that of a high-school degree.

Wages are included as an indicator of labour costs within a district, and as discussed earlier the expected coefficient is difficult to predict since the data does not allow me to control for the skill set of the workers. On average, however, the results seem to indicate that higher wages are associated with a lower count of patent applications.

Since research in this field has previously not been carried out for India, I limit myself to comparisons with studies for other countries. Autant-Bernard (2001), Bottazi and Peri (2003), Fischer and Varga (2003), Bode (2004) and Rodriguez-Pose and Creszensi (2008) find evidence of inter-regional spillovers, but find that these decay quickly over larger distances. It is possible that in this paper I am unable to measure spillovers at a scale lower than that of the district, and thus note only negative effects of R&D in neighbouring locations. However, Del Barrio-Castro and Garcia-Quevedo (2005) also fail to find any significant effect of R&D expenditures in neighbouring regions for Spain. Ponds et al (2010) find evidence of inter-regional spillovers, and of positive effects of human capital in the Netherlands. However, since their paper concentrates on university linkages, it is difficult to make one-to-one comparisons. Knudsen et al (2008) measure the effect of the presence of scientists and engineers for metropolitan areas in the US and are unable to find a statistically significant effect on patent activity. On the other hand, Carlino et al (2007) find a positive and significant effect of college-educated population on patents per capita in the US. For Italy, Paci and Usai (1999), find that localisation and industrial diversity have positive and significant effect on innovation. On the other hand, for the US, Feldman and Audretsch (1999) find that localisation has a negative effect on innovation, while industrial diversity has a positive effect. However, Aces et al (2002) find that the localisation (labelled location quotient in their paper) has no effect on patent activity. A number of these and other papers also measure the effects of university research on a location's patent activity and find evidence of positive effects – however, owing to lack of data I am unable to analyse the effects of university research in this paper.

At this stage the paper does not claim causality – potentially a number of variables have been omitted and the comparisons are being made across districts and not within districts. Nevertheless, the associations provide some interesting results. The next section addresses some of these concerns and provides some robustness checks.

4.6 Omitted Variables Bias and Robustness Checks

Although all the regressors have been lagged, there could remain endogeneity concerns that would bias the coefficients. One potential source of bias could be the possibility of variables omitted from the model that could lead to biased estimates of β_3. For instance, there could be factors intrinsic to the natural geography of the region that could be driving these results. In other words, some underlying features of the natural geography, could be jointly influencing agglomeration and innovation within a location. The introduction of district fixed effects would effectively control for the effect of unobservable variables at the level of the district.

Ideally, the introduction of district fixed-effects should be introduced at each point of the count models presented above, along with the industry and the year fixed effects. However, the introduction of district dummies does not allow the count models to converge in the specifications provided in Table 4.2. As explained by Silva and Tenreyro (2010), count models are less likely to converge in cases when the regressor is zero and the dependent variable is positive, otherwise being non-negative with at least one positive observation. Dummy variables often fit these characteristics, especially when the dummy equals zero for all observations with a positive dependent variable, and having positive value when the dependent variable equals zero. This is very much the case in this dataset when district-level dummies are introduced into the specification.

In a recent paper, Silva and Tenreyro (2011) define a new code within Stata to deal with this problem. The alternatives suggested, however, do not deal very effectively with the problem caused by the dummy variable. This is because their procedure would involve dropping some of the dummies, which would dimply an arbitrary redefinition of the reference category. This strategy, as pointed out by the authors themselves, is unlikely to be a sensible one. Another alternative might be to use generalised linear models that reweight the least squares algorithm to make it more stable[90]. This option does not solve the problems associated with the district-level dummy, as all the dummies are dropped from the procedure, effectively reducing the specification to one without any district fixed-effects.

In the absence of other options to reach convergence within the count models with the introduction of district fixed effects, I use ordinary least squares (OLS) estimations. The OLS method ignores observations where the count of patents is zero, but on the flip side it is able to account for the unobservables at the level of the district and the direction and magnitude of the coefficients would help provide some level of confidence in the estimates. The results are presented in Table 4.3.

The first three columns introduce the set of explanatory variables, column (4) includes industry fixed effects, column (5) includes industry and year fixed effects and column (6) includes industry, year and district fixed effects. The dependent variable is the log of the count of patent applications. The coefficients can be interpreted as follows: a unit increase (i.e. 1 Crore) in R&D expenditures leads to a 0.09 percent increase in the count of patent applications, a unit increase in the Herfindahl Index (i.e. a fall in industrial diversity) leads to an 85 percent reduction in patent applications and a unit increase in the percentage of the population with a technical degree leads to a 1.9 percent increase in the count of patent applications (see column 6). The effect of R&D expenditures is relatively stable and statistically significant across different specifications, although the magnitude of the effect is not very large and drops somewhat with the introduction of

[90] The authors provide a new command 'ppml' in Stata to carry out this estimation. But this continues to (correctly) drop the district dummy.

district fixed-effects. A fall in industrial diversity also has a relatively large and consistently negative effect on patent applications, and the percentage of the population with a technical degree only positive effects patent applications significantly when district dummies are introduced. Note that the number of observations falls considerably compared to the count models, since the dependent variable only includes non-zero values.

Table 4.3: Step-by-Step OLS

Variable	(1)	(2)	(3)	(4)	(5)	(6)
R&D	0.0037***	0.0031***	0.0029***	0.0029***	0.0029***	0.0009**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]
Access		0.1419	0.1105	0.0936	0.1013	-0.0544
		[0.140]	[0.135]	[0.148]	[0.150]	[0.045]
HI		-2.3228***	-1.4809***	-1.5192***	-1.0679**	-0.8555
		[0.504]	[0.530]	[0.555]	[0.526]	[1.343]
Localisation		3.5514*	3.2139*	4.0055**	3.7481**	0.8255
		[2.106]	[1.942]	[1.894]	[1.651]	[0.524]
Education (technical)			-0.0718	-0.0729	0.0280	-0.0195
			[0.045]	[0.049]	[0.058]	[0.032]
Education (high-school)			0.0475**	0.0529**	0.0599***	0.0098
			[0.020]	[0.022]	[0.020]	[0.017]
Wages			0.0002	0.0003	-0.0000	-0.0007
			[0.001]	[0.001]	[0.001]	[0.001]
Industry FEs	×	×	×	√	√	√
Year FEs	×	×	×	×	√	√
District FEs	×	×	×	×	×	√
#	2,766	2,745	2,745	2,745	2,745	2,745
R^2	0.339	0.398	0.416	0.460	0.493	0.778

Robust standard errors clustered at district level.
*** p<0.01, ** p<0.05, * p<0.1

As a robustness check, I also differentiate between different types of applicants - from government-funded research organisations, private-funded research organisations, industry and individuals. For each cross-section, I disaggregate the data according to the type of applicant and re-run the OLS regressions to observe if the effects of the predictor variables vary across each group - see Table 4.A.4. The specifications include industry, year and district fixed effects.

Recall that R&D represents only private R&D, and it is worth noting that it positively affects patents applications from firms and individuals, but has no statistically significant effect on private government-funded research[91]. Much more notably, the effect of the variable 'Access', i.e. R&D funding in neighbouring locations positively affects the innovative output of government-funded research agencies and firms. In other words, being situated close to a regions with higher R&D funding could lead to positive effects on the propensity to innovate for some types of patent applicants. The negative effect of the presence of more technical human capital on government-funded innovative output could be explained by the propensity of these agencies to locate in remote areas with low levels of technical human capital in the general population. As expected, however, technical human capital affects individual-level patenting positively, while the effect of high-school education has a positive effect on firms' innovative output.

4.7 Conclusion

The main findings of the paper are that (1) private R&D expenditures have a positive effect on fuelling innovation, (2) that the spatial spillovers of these expenditures seem to be positive for firms and individuals, (3) industrial diversity has a positive effect on the count of new patent applications within a location, and that (4) the level of human capital, both technical and more general, seems to matter as well.

---

[91] Government-funded research includes both, research carried out within government departments and within government-funded research institutions.

This paper contributes to the empirical literature on the geography of innovation in a number of ways. There is little research that has looked at emerging countries, and this is the first study of what drives innovative activity in India. Secondly, responding to the shortcomings of regions or states as the unit of observation, the spatial scale of the study goes down to the level of districts, which in a number of cases correspond to city-level boundaries in India. Indeed, data on patents has not been collected or used previously at the spatial scale of districts. And lastly, the study finds robust evidence to link the agglomeration of economic activity to innovative activity, after controlling for possible omitted variables bias.

The study has a few limitations. The analysis of patent applications made to the US and the European Patent Offices are not included. The analysis is also based on patent data, which certify new inventions, and thus innovation citations are excluded. As a number of authors have warned (Griliches 1990, Mansfield 1984, Scherer 1983 – quoted in Feldman and Florida 1994) that the number of patents is not directly equivalent to a measure of innovative output as many patented inventions never become commercially viable products while many successful products are never patented. Nonetheless patents continue to be used as a useful measure of the generation of ideas. In addition the empirical analysis is based on cross-sectional data, albeit at two points in time. Although the time lags between the dependent and independent variables lower the risk of endogeneity, future studies using panel data are necessary to come to more decisive conclusions.

The findings of this study could have important implications for policy. Firstly, the effect of private, firm-level R&D has a positive effect on not just industry-level innovation, but also spills over into patenting activity carried out by private individuals. While the effect of government R&D has not been studied in this paper, it would be interesting to see to what extent the public sector could further spur innovation. Secondly, education matters. This may be self-evident, but it is worth reiterating. A better-educated labour force increases the ability of a region to innovate, and investing in education would be an efficient means to spur on lagging regions. Lastly, the results indicate that agglomeration economies continue to have a significant effect on innovative activity – this result holds within the panel estimates and is robust to the type of innovator being examined. While government policy can have a direct effect on improving education and research infrastructure, its capacity to generate agglomeration economies remains unclear.

References

Acs, Z. J., Anselin, L., Varga, A. (2002) 'Patents and innovation counts as measures of regional production of new knowledge', Research Policy, 31: 1069–1085.

Anselin, L., Varga, A., Acs, Z. (1997) 'Local geographic spillovers between university research and high technology innovations', Journal of Urban Economics, 42: 422-448.

Anselin, L., Varga, A., Acs, Z. (2000) 'Geographic and sectoral characteristics of academic knowledge spillovers', Papers in Regional Science, 79: 435–445.

Audretsch, D.B. (1998) 'Agglomeration and the location of innovative activity', Oxford Review of Economic Policy, 14(2): 18-29.

Audretsch, D.B and Feldman, M.P. (1996) 'R&D spillovers and the geography of innovation and production', American Economic Review, 86(3): 630-640

Autant-Bernard, C. (2001) 'Science and knowledge flows: evidence from the French case', Research Policy, 30: 1069–1078.

Ayyagari, M, Demirguc-Kunt and Maksimovic, V. (2007) 'Firm innovation in emerging markets', World Bank Policy Research Working Paper 4157.

Barkey, D.L., Henry, M.S. and Lee, D. (2006) 'Innovative activity in rural areas: the importance of local and regional characteristics', Community Development Investment Review, Federal Bank of San Francisco.

Bode, E. (2004) 'The spatial pattern of localized R&D spillovers: an empirical investigation for Germany', Journal of Economic Geography, 4: 43-64

Carlino, G.A., Chatterjee, S. and Hunt, R.M. (2007) 'Urban density and the rate of invention', Journal of Urban Economics, 61: 389-419.

Co, C.Y. (2005) 'Co-evolution of invention activities among cities in New England', Papers in Regional Science, 85(1): 47-71.

Del Barrio-Castro, T. and Garcia-Quevedo, J. A. (2005) 'Effects of university research on the geography of innovation', Regional Studies, 39: 1217–1229.

Doring, T. and Schnellenbach, J. (2006) 'What do we know about geographical knowledge spillovers and regional growth? A survey of the literature', Regional Studies, 40: 375–395.

Feldman, M.P., and Audrestch, D.B. (1999) 'Innovation in cities: science-based diversity, specialization and localized competition', European Economic Review, 43: 409-429.

Feldman, M.P. and Florida, R. (1994) 'The geographic sources of innovation: technological infrastructure and product innovation in the United States', Annals of the Association of American Geographers, 84 (2): 210-229.

Fischer, M. and Varga, A. (2003) 'Spatial knowledge spillovers and university research: evidence from Austria', Annals of Regional Science, 23: 303–322.

Fritsch, M. and Slavtchev, V. (2007) 'Universities and innovation in space', Industry and Innovation, 14: 201–218.

Girma, S., Gong, Y. and Goerg, H. (2007) 'Foreign direct investment, access to finance, and innovation activity in Chiene enterprises', Research Paper Series, China and the World Economy, University of Nottingham.

Hanson, G. H. (2001) 'Scale economies and the geographic concentration of industry', Journal of Economic Geography, 1: 255–276.

Hardin, S., Schmiediche, H. and Carroll, R. (2003) 'Instrumental variables, bootstrapping, and Generalized Linear Models', Stata Journal, 3(4): 351-360.

Ibrahim, S.E., Fallah, M.H. and Reilly, R.R. (2009) 'Localised sources of knowledge and the effect of knowledge spillovers: an empirical study of inventors in the telecommunications industry', Journal of Economic Geography, 9: 405-431.

Jaffe, A.B. (1986) 'Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value', American Economic Review, 76 (5): 984-1001.

Jaffe, A.B. , Trajtenberg, M. and Henderson, R. (1993) 'Geographic localization of knowledge spillovers as evidenced by patent citations', The Quarterly Journal of Economics, 108 (3): 577-598.

Kantor, S. and Whalley, A. (2009) 'Do universities generate agglomeration spillovers? Evidence from endowment value shocks', Working Paper 15299, National Bureau of Economic Research, Cambridge, MA.

Keeble, D. and Wilkinson, F. (1999) ‘Collective learning and knowledge development in the evolution of regional clusters of high technology SMEs in Europe’, Regional Studies, 33: 295–303.

Kelly, M. and Hageman, A. (1999) ‘Marshallian externalities in innovation’, Journal of Economic Growth, 4 (1): 39-54.

Kerr, W.R. (2009) ‘Breakthrough inventions and migrating clusters of innovation’, Working Paper 15443, National Bureau of Economic Research, Cambridge, MA.

Knudsen, B., Florida, R., Stolarick, K. and Gates, G. (2008) ‘Density and creativity in US regions’, Annals of the Association of American Geographers, 98(2): 461-478.

Krugman, P. (1991) Geography and Trade. Cambridge: MIT Press.

Menon, C. (2009) ‘Stars and comets: an exploration of the patent universe’, Discussion Paper 37, Spatial Economics Research Centre, London.

Paci, R. and Usai, S. (1999) ‘Externalities, knowledge spillovers and the spatial distribution of innovation’, GeoJournal, 49: 381.

Pakes, A. and Griliches, Z. (1980) ‘Patents and R&D at the firm level: a first look’, NBER Working Paper 561.

Piergiovanni, R. and Santarelli, E. (2001) ‘Patents and the geographic localization of R&D spillovers in French manufacturing’, Regional Studies, 35: 697–702.

Ponds, R., van Oort, F. and Frenken, K. (2010) ‘Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach’, Journal of Economic Geography, 10: 231-255.

Rajeswari, A.R. (1996) ‘Indian patent statistics – an analysis’, Scientometrics, 36(1): 109-130.

Rodriguez-Pose, A. and Crescenzi, R. (2008) ‘R&D, spillovers, innovation systems and the genesis of regional growth in Europe’, Regional Studies, 42 (1): 51–67.

Santos Silva, J.M.C. and Tenreyro, S. (2010) ‘On the existence of the maximum likelihood estimates in Poisson regression’, Economic Letters, 107 (2): 310-312.

Santos Silva, J.M.C. and Tenreyro, S. (2011) ‘Poisson: Some convergence issues’, STATA Journal, 11 (2): 207-212.

Tables

Table 4.A.1: Summary of Empirical Literature

Study	Country	Sample	Unit of Observation
Jaffe (1986)	USA	432 firms; 1973, 1979	Firm
Jaffe et al (1993)	USA	950 and 1,450 patents; 1975, 1980	State
Feldman and Florida (1994)	USA	4,200 innovations; 1982	State
Audretsch and Feldman (1996)	USA	8,074 innovations, 1982	State
Anselin et al (1997)	USA	4,200 innovations; 1982	MSA
Feldman and Audretsch (1999)	USA	3,696 innovations; 1982	MSA
Paci and Usai (1999)	Italy	1978-1995	Local Labour System
Anselin et al (2000)	USA	1982	MSA
Piergiovanni and Santarelli (2001)	France	20,700 patents; 1991-1992	State
Autant-Bernard (2001)	France	1994-1996	Department
Acs et al (2002)	USA	4,476 patents; 1982	MSA
Bottazzi and Peri (2003)	Europe	3,010 patents; 1977-1995	NUTS 1/NUTS 2
Fischer and Varga (2003)	Austria	1993	District
Bode (2004)	Germany	27,361 patents; 1998	NUTS 3
Co (2005)	New England, USA	1975-1999	MSA
Del Barrio-Castro and Garcia-Quevedo (2005)	Spain	1996-2000	NUTS 2
Barkley et al (2006)	Southern counties, USA	125,589 patents; 1990-1999	MSA
Carlino et al (2007)	USA	581,000 patents; 1990-1999	MSA
Girma et al (2007)	China	239,085 firms; 1999-2005	Firm
Fritsch and Slavtchev (2007)	Germany	31,434 patents; 1995-2000	NUTS 3
Knudsen et al (2008)	USA	1999	MSA

Rodriguez-Pose and Crescenzi (2008)	Europe	1995-2003	NUTS 1/NUTS 2
Ibrahim et al (2009)	USA	5,353 patents; 2000-2003	Telecom clusters
Kantor and Whalley (2009)	USA	1981-1996	County
Kerr (2009)	USA	1975-1984	City
Menon (2009)	USA	1,161,650 patents; 1975-1999	MSA
Ponds et al (2010)	The Netherlands	3,332 patents; 1999-2001	NUTS 3

Notes: MSA: Metropolitan Statistical Area; NUTS: European Nomenclature for Territorial Units for Statistics

Table 4.A.2: Some Descriptive Statistics

Variable	Expected sign	#		Mean	
		1999-2000	*2004-2005*	*1999-2000*	*2004-2005*
Patents		587	579	6.77	10.80
R&D	+	589	579	3.7064	10.0654
Access	+	583	575	870246	868285
Localisation	+	485	575	0.0043	0.0036
Herfindahl Index	-	589	579	0.3377	0.4185
Education (High-School)	+	489	579	0.0566	0.0742
Technical Education	+	487	577	0.0024	0.0112
Wages	-/+	486	575	93.9052	101.0757

Notes: # refers to the number of districts for which data is available. There are a total of 604 districts in the country.

Table 4.A.3: Characteristics of the Data (1999)

Variable	#	Mean	Std. Dev.	Min.	Max.
count	587	6.77	75.22	0	1353
count>0	126	31.61	160.55	1	1353
obs	588	0.7857	0.41	0	1
Fit_P	485[92]	0.4797	0.29	0	0.94
Fit_NB	485	0.7590	0.15	0.06	0.97
Fit_ZIP	485	0.7636	0.20	0.00	0.98

[92] The number of observations are lesser than the number of cases in the dataset owing to missing values for some variables in the model.

Table 4.A.4: Robustness Checks

Variable	Government-funded	Private-funded	Firms	Individuals
R&D	0.0019	0.0002	0.0006**	0.0005*
	[0.001]	[0.000]	[0.001]	[0.001]
Access	1.4871***	0.2132	0.5283***	0.2461
	[0.409]	[0.198]	[0.183]	[0.198]
HI	-4.5803*	1.3518	0.3008	-0.7141
	[2.408]	[2.753]	[2.437]	[1.625]
Localisation	0.2765	0.8199	0.6344	1.3367
	[1.779]	[1.614]	[2.524]	[0.989]
Education (technical)	-0.1540	0.0615	0.0284	0.0938*
	[0.090]	[0.039]	[0.042]	[0.052]
Education (high-school)	0.0241	0.0166	0.0731**	0.0661
	[0.037]	[0.026]	[0.030]	[0.069]
Wages	-0.0107*	0.0016	0.0017	-0.0091
	[0.006]	[0.003]	[0.001]	[0.006]
Industry Fixed Effects	√	√	√	√
Year Fixed Effects	√	√	√	√
District Fixed Effects	√	√	√	√
#	547	430	1,018	423
R^2	0.840	0.587	0.805	0.727

Robust standard errors clustered at district level.
*** p<0.01, ** p<0.05, * p<0.1

Table 4.A.5: IPC and NIC Matching

IPC	IPC definition	NIC	NIC Definition
		1	Agriculture, hunting and forestry
		2	Forestry, logging and related service activities
		5	Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing
		15	Manufacture of food products and beverages
		16	Manufacture of tobacco products
		18	Manufacture of wearing apparel; dressing and dyeing of fur
		19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
		36	Manufacture of furniture; manufacturing N.E.C.
A	Human Necessities	85	Health and social work
B	Performing operations and transporting	20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
		22	Publishing, printing and reproduction of recorded media
		34	Manufacture of motor vehicles, trailers and semi-trailers
		35	Manufacture of other transport equipment
		37	Recycling
		50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel
		60	Land transport; transport via pipelines
		61	Water transport
		62	Air transport
		63	Supporting and auxilliary transport activities;

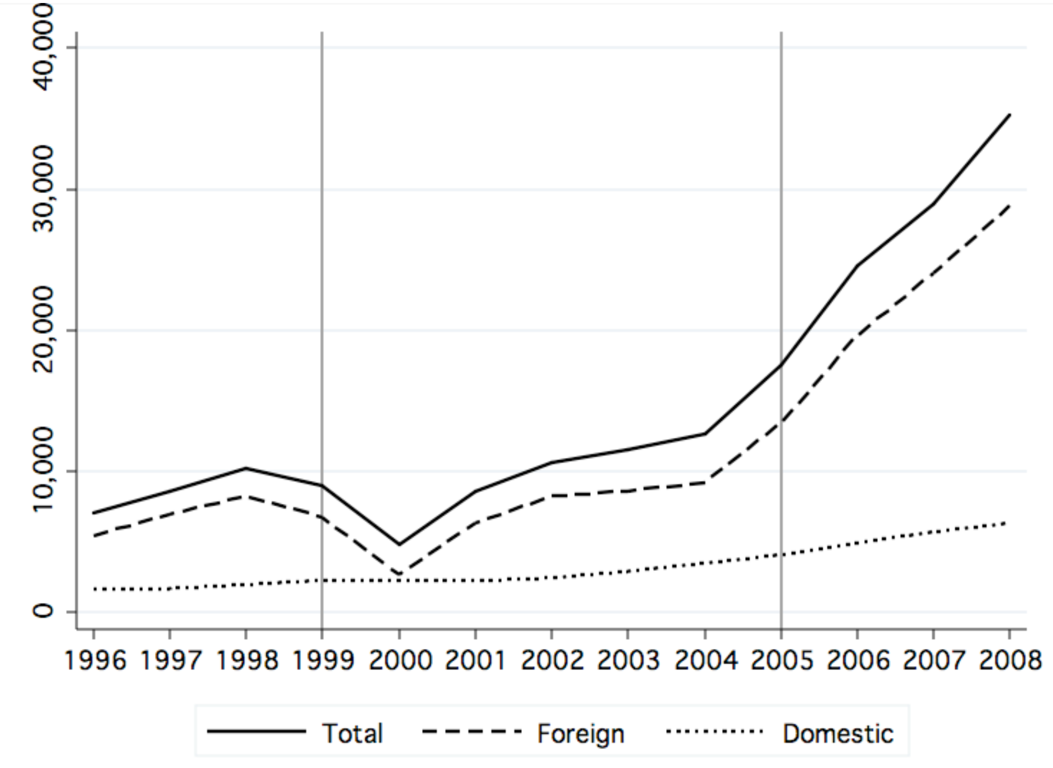
			activities of travel agencies
		64	Post and communications
		19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
		23	Manufacture of coke, refined petroleum products and nuclear fuel
		24	Manufacture of chemicals and chemical products
		25	Manufacture of rubber and plastic products
		26	Manufacture of other non-metallic mineral products
C	Chemistry and Metallurgy	27	Manufacture of basic metals
		17	Manufacture of textiles
D	Textiles and Paper	21	Manufacture of paper and paper products
		10	Mining of coal and lignite; extraction of peat
		11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction excluding surveying
		12	Mining of uranium and thorium ores
		13	Mining of metal ores
		14	Other mining and quarrying
		41	Collection, purification and distribution of water
		45	Construction
E	Fixed Constructions	90	Sewage and refuse disposal, sanitation and similar activities
		28	Manufacture of fabricated metal products, except machinery and equipments
		29	Manufacture of machinery and equipment N.E.C.
F	Mechanical Engineering	31	Manufacture of electrical machinery and apparatus N.E.C.
G	Physics		
		30	Manufacture of office, accounting and computing

			machinery
		32	Manufacture of radio, television and communication equipment and apparatus
		33	Manufacture of medical, precision and optical instruments, watches and clocks
		65	Financial intermediation, except insurance and pension funding
		72	Computer and related activities
H	Electricity	40	Electricity, gas, steam and hot water supply

Notes: IPC: International Patent Classification
NIC: National Industrial Classification

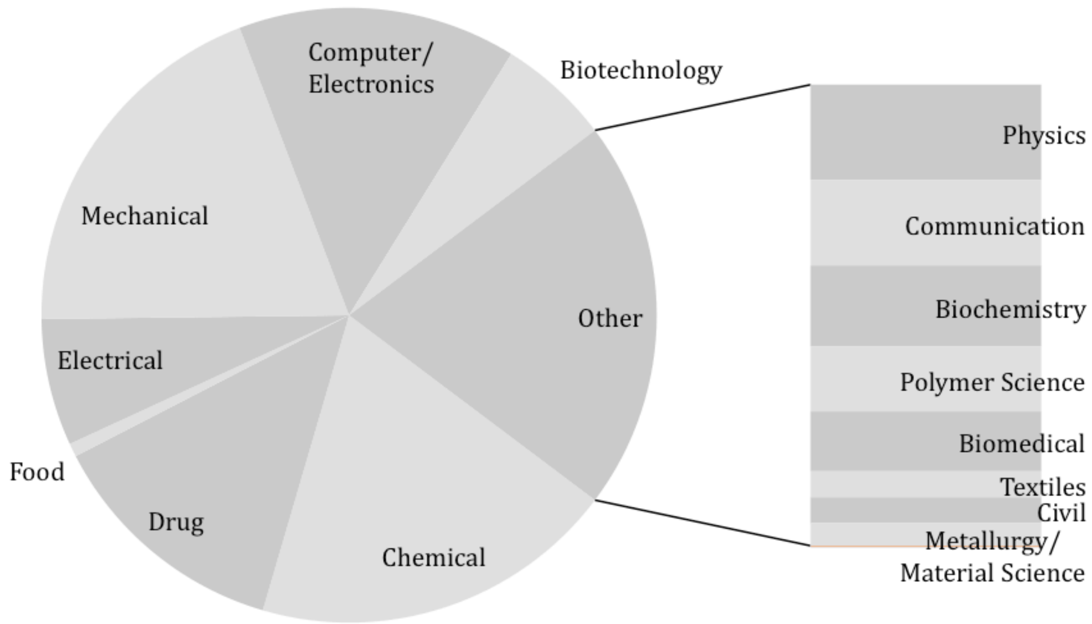
Figures

Figure 4.A.1: Patent Applications to the IPO (1995-2008)



Source: IPO Annual Reports

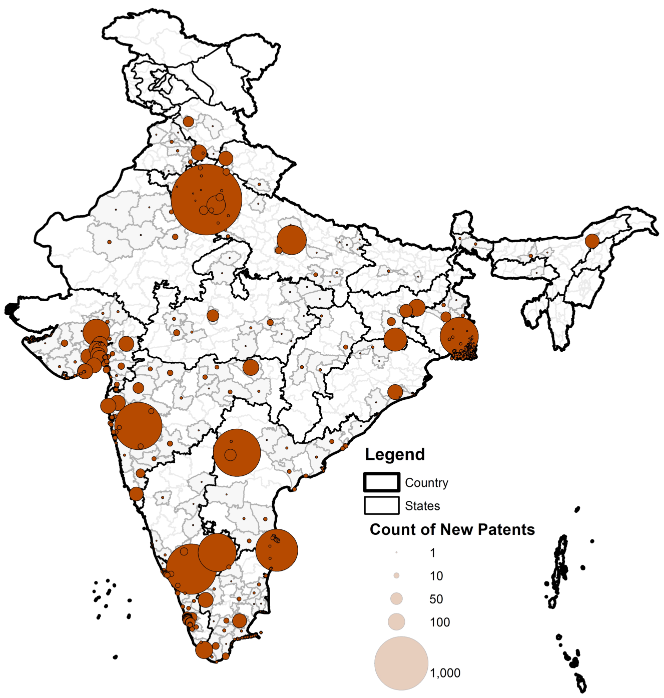
Figure 4.A.2: Patent applications by sector (2004)



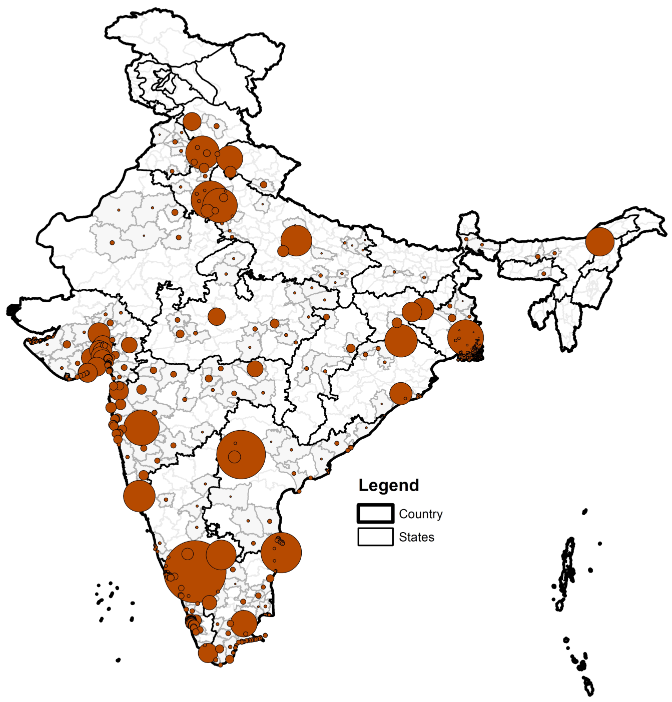
Source: IPO Annual Report (2004-2005)

Figure 4.A.3: Spatial distribution of patents

Total



Controlling for district size



Source: FAO and IPO