



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Spatial Networks, Clustering and Spillovers:

Lessons for Development

Emma Howard

A thesis submitted for the degree of:

Doctor of Philosophy

Supervised by:

Dr. Carol Newman

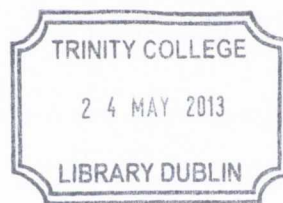
University of Dublin,

Trinity College,

Department of Economics,

Trinity College Dublin

2013



Thesis 9921

Declaration

I declare that this dissertation has not been submitted as an exercise for the degree of Doctor of Philosophy (Ph.D.) at this or any other university. All research contained herein that is not entirely my own but is based on research that has been carried out jointly with others is duly acknowledged in the text wherever included.

I agree that the library of Trinity College, University of Dublin may lend or copy this thesis upon request. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Summary

The geographic clustering of manufacturing activity has long been recognised as facilitating the growth and prosperity of firms and regions alike. Empirical evidence however is limited, particularly in a developing country context. This thesis uses an exceptionally rich data set on manufacturing firms in Vietnam to explore issues relating to the measurement of, and the driving forces behind, agglomeration.

Chapter 1 provides an introduction to the thesis.

Chapter 2 presents a new approach for the empirical investigation of agglomeration patterns. We first plot the location of each firm to generate a visual representation of the location pattern. We then examine the clustering of manufacturing firms by identifying patterns of spatial network formation that deviate from generated networks using a number of counterfactual spatial distributions. We calculate a new measure of the strength of clustering of manufacturing firms and test whether this is greater for the actual spatial network than for the generated networks. Our findings suggest that the extent of clustering is over and above that which can be attributed to the legal and regulatory framework, economic zoning, or population patterns. This is an important and often overlooked first step in understanding whether agglomeration effects are present in an industry.

In chapter 3 we use aggregated data and focus on between sector agglomeration in the manufacturing industry in Vietnam. We propose a new measure for the extent to which firms from two different industries locate in the same place, or coagglomerate. We examine what it reveals about the importance of transport costs, labour market pooling and technology transfers for agglomeration processes. We also disaggregate how agglomeration forces differ for high-tech and low-tech sector pairings. We find that firms in high-tech sectors pool skilled workers while firms in low-tech sectors compete for workers. We also find technology transfers are an important agglomerative force in Vietnam. We contrast our analysis with insights from existing measures in the literature and find very different underlying stories at work.

In chapter 4 we use micro data to conduct a fixed effects dyadic analysis which controls for regional characteristics and allows us to focus in on agglomeration economies in the location choice of the firm. In order to identify strategic clustering we focus on new

entrants and identify the types of firms they choose to locate beside. We find that new entrant firms are more likely to locate beside incumbent firms in the same 2 or 4-digit sectors, suggesting the presence of horizontal spillovers. We also find that new entrants choose to locate close to foreign or state-owned incumbents. If the new entrant is in a high-tech sector they are likely to locate beside a high-tech incumbent, suggesting technology transfers are an agglomerative force. Finally, new entrants are more likely to locate beside incumbents that have a different level of production technology. Foreign-owned new entrant firms are more likely to locate beside an incumbent firm with a higher level of technology, if both firms are in high-tech sectors. This again suggests that technology transfers are an agglomerative force as transfers are most likely to occur from a firm with a high level of technology to a firm with a lower level of technology.

Chapter 5 concludes the thesis and suggests areas for future research.

Acknowledgements

The completion of this thesis would not have been possible without the support and encouragement of a number of people.

First and foremost, I would like to thank my supervisor Carol Newman. Over the last three years she has been incredibly generous with her time, and her expertise and experience has been invaluable. She replies to emails and sends feedback so quickly, no matter where she is in the world, and thanks to her generosity and access to funding I have never had to worry about conference travel. I could not have wished for a better supervisor.

I would also like to thank Jacco Thijssen who supervised me in the first year of my PhD. Our weekly meetings helped to keep my research focused and provided insightful discussions.

Access to the data used in this thesis is thanks to Finn Tarp and I am very grateful to him and to the Central Institute of Economic Management (CIEM), Hanoi, Vietnam for the provision of the data.

This research was funded by Trinity College and I would like to thank Andrew Somerville and the Department of Economics for securing a College Award for me in my first year. I would also like to thank the administrative staff in the department, Colette Ding and Trish Hughes, for all their help over the years.

The staff and my PhD colleagues in the economics department have all been encouraging of my research and have provided helpful comments and feedback at various department presentations, and at the micro working group. I would particularly like to thank my colleagues Julia Anna Matz, Benjamin Elsner and Conor O'Toole for their advice and friendship throughout the PhD. John O'Hagan also warrants special mention; he has been extremely supportive ever since my time as an undergraduate in Trinity.

I am extremely grateful for the love and support, both emotional and financial, of my parents, during my PhD and indeed my many years as a student. I cannot thank them enough for all they have done for me.

Last, but by no means least, I want to thank my amazing husband Michael. He has been there for me since my first day as an undergraduate economics student and is a constant source of strength and inspiration. Everything I do in life is better because of him and this thesis is no exception.

Contents

1. Introduction	6
2. Are Spatial Networks of Firms Random?	
2.1 Introduction	10
2.2 Conceptual Framework	11
2.3 Data and Methodology	14
2.4 Results	17
2.5 Robustness Checks and Further Investigation	20
2.5.1 Firm Size and Statistical Significance	20
2.5.2 Ownership Structure	21
2.5.3 Location Restriction	23
2.5.4 Economic/Industrial Zones	24
2.5.5 Choice of Threshold Distance	25
2.5.6 Population Density	25
2.6 Conclusion	26
3. Measuring Industry Coagglomeration and Identifying the Driving Forces	
3.1 Introduction	28
3.2 Definition and Measurement of Agglomeration	31
3.3 Determinants of Agglomeration	35
3.3.1 Transport Costs	35
3.3.2 Labour Market Pooling	39
3.3.3 Technology Spillovers	40
3.3.4 Natural Advantages	41
3.4 Descriptive Statistics	41
3.5 Empirical Results	43
3.6 Discussion and Conclusion	53
4. Where do new entrant firms locate? A dyadic analysis	
4.1 Introduction	56
4.2 Motivation and Empirical Model	58

4.3 Data	62
4.4 Results	67
4.5 Conclusion	75
5. Conclusion	77
A Chapter 1 Appendix: Maps of Firm Location in Vietnam	87
B Chapter 2 Appendix: Description of Manufacturing Sector Codes	93

List of Tables

2.1 Number of Manufacturing Firms in Vietnam 2002-2007	14
2.2 Transitivity of the network of all manufacturing firms	18
2.3 Numbers of large manufacturing firms	20
2.4 Transitivity of the network of large manufacturing firms	20
2.5 Employment by size distribution of manufacturing firms	21
2.6 Transitivity of the networks of private firms only	23
2.7 Transitivity of the networks of firms within zones	24
2.8 Transitivity of the networks of firms outside zones	25
2.9 Transitivity of the networks of large firms: small threshold	25
2.10 Transitivity of the networks of large firms: large threshold	25
3.1 Highest colocation pairings: Commune	36
3.2 Highest colocation pairings: District	36
3.3 Highest colocation pairings: Province	37
3.4 Highest pairings using EG Index: Commune	37
3.5 Highest pairings using EG Index: District	38
3.6 Highest pairings using EG Index: Province	38
3.7 Descriptive statistics	42
3.8 OLS regression of determinants of coagglomeration using colocation index	44
3.9 OLS regression of the determinants of coagglomeration using absolute colocation measure	46
3.10 OLS regression of determinants of coagglomeration using EG measure	48

3.11 OLS regression of determinants of coagglomeration for tech pairings using colocation index	50
3.12 OLS regression of determinants of coagglomeration for tech pairings using EG measure	52
4.1 Dummy variables	65
4.2 Relational variables	66
4.3 New Entrants	66
4.4 Fixed effects dyadic regression	68
4.5 Fixed effects dyadic regression Foreign entrants only	71
4.6 Fixed effects dyadic regression Private domestic entrants only	72
4.7 Fixed effects dyadic regression including <i>Incumbent Higher-Tech</i> Dummy	74

List of Figures

A.1 Clustering of Manufacturing Firms 2002	88
A.2 Clustering of Manufacturing Firms 2003	88
A.3 Clustering of Manufacturing Firms 2004	88
A.4 Clustering of Manufacturing Firms 2005	88
A.5 Clustering of Manufacturing Firms 2006	89
A.6 Clustering of Manufacturing Firms 2007	89
A.7 Clustering of Large Firms 2002	89
A.8 Clustering of Large Firms 2007	89
A.9 Clustering of Small Firms 2002	90
A.10 Clustering of Small Firms 2007	90
A.11 Clustering of Medium Firms 2002	90
A.12 Clustering of Medium Firms 2007	90
A.13 Clustering of State-owned Firms 2002	91
A.14 Clustering of State-owned Firms 2007	91
A.15 Clustering of Privately-owned Firms 2002	91
A.16 Clustering of Privately-owned Firms 2007	91
A.17 Clustering of Foreign-owned Firms 2002	92
A.18 Clustering of Foreign-owned Firms 2007	92

Chapter 1

Introduction

The geographic clustering of economic activity has received considerable attention in the economics literature over the past century. The seminal theoretical work by Marshall (1920) postulated three reasons why firms locate in close proximity to other firms. Essentially, firms can reduce costs by locating in geographic clusters, as proximity to other firms reduces the cost of transporting goods, people and ideas. First, the cost of transporting goods is reduced when firms locate close to their customers, or close to their suppliers. Second, when firms locate in a cluster, a pool of workers emerges. This makes it easier for firms to hire new workers when their labour demand increases and also facilitates better matching of workers to jobs. Third, knowledge spillovers are more likely when firms are located close to other firms. The informal exchange of ideas in particular is more likely when firms are in close geographic proximity. These three theoretical reasons why firms cluster are integral to the economic geography literature and are known as Marshallian forces.¹

In the 1990s a new strand of agglomeration literature emerged, commonly termed ‘the new economic geography’. Beginning with Krugman’s (1991) model, this literature produced a number of general equilibrium models which attempt to explain the spatial structure of the economy. These models link Marshall’s sources of external economies, achieved by reducing costs associated with transporting mobile factors, with immobile factors such as land and resources. The emergence of these new theoretical models generated renewed interest in the geographic concentration of activity.

Enterprise concentration and the resulting agglomeration economies have long been recognised as an important mechanism for facilitating industrial growth, in both developed and developing countries (Krugman, 1991; Markusen and Venables, 1999).² The downside however is that economic concentrations can lead to spatial inequalities and there is evidence from many developing countries that this is not only occurring,

¹ See Section 3.3 for a more detailed discussion of the three Marshallian forces.

² Throughout this thesis when we refer to agglomeration we are considering industry agglomeration. An alternative interpretation is the agglomeration of people, or urbanisation, (see for example Uchida and Nelson, 2008) which we do not consider in this thesis.

but increasing over time (see for example Felkner and Townsend, 2011). Policy makers in developing countries therefore have a difficult task in formulating industrial policies that both foster clusters and promote growth, and reduce spatial inequality. Before such policies can be formulated it needs to be determined that agglomeration economies are in fact present in the country, and subsequently the driving forces behind agglomeration identified. Empirical evidence on the extent of industry agglomeration and the relative magnitude of the driving forces behind it is limited, particularly in a developing country context.

Following on from the important theoretical work of Krugman and others on agglomeration, an empirical literature has emerged. The majority of empirical work on agglomeration however has heretofore concentrated on developed countries. This thesis uses unique and rich data to investigate the extent of, and agglomerative forces behind, the clustering of manufacturing firms in Vietnam.

In the mid 1980s Vietnam began to move from a centrally planned to a market-oriented economy. This transformation in governance has resulted in a structural shift in the economy from agriculture to manufacturing; in 2007 the share of industry in GDP had risen to 42 percent. Industrialisation has been accompanied by rapid economic growth, however spatial disparity has become a big concern (Son, 2009). Substantial inflows of foreign direct investment (FDI) have been a driving force for growth in Vietnam but have also contributed to the spatial disparity; three quarters of FDI is concentrated in two regions (Huang and Bocchi, 2009).³ These factors make Vietnam a particularly interesting country context in which to investigate industry clustering and spillovers.

Two important issues have arisen in the analysis of the clustering of firms; the first is how agglomeration should be defined and measured, and the second is how the agglomerative forces at work in a country can be identified. This thesis contributes to both strands of the empirical agglomeration literature.

Chapter 2 of this thesis presents a new approach for the empirical investigation of agglomeration patterns using tools from network analysis. An important and often overlooked first step in understanding whether agglomeration effects are present in an

³ The Southeast (around Ho Chi Minh City) and the Red River Delta (around Hanoi).

industry is to first establish that the industry exhibits *significant* spatial clustering.⁴ We use panel data from 2002-2007 and plot the location of each firm in each year to generate a visual representation of the evolving location pattern. We then examine the clustering of manufacturing firms by calculating a new measure of the strength of clustering. We test whether this is greater than the measure calculated for a number of networks generated by counterfactual spatial distributions. Our findings show that the extent of clustering is over and above that which can be attributed to the legal and regulatory framework, economic zoning, or population patterns. This suggests that agglomeration economies are present in the manufacturing industry in Vietnam.

Given the presence of these agglomeration economies, chapter 3 uses a number of unique data sets to control for natural cost advantages between regions and test the relative importance of each of Marshall's three theoretical agglomerative forces (input-output links, labour market pooling and technology transfers) in the manufacturing industry in Vietnam. Although an extensive literature exists on agglomeration externalities, many empirical investigations focus on just one agglomerative force and very few provide evidence in a developing country context.⁵ Chapter 3 uses aggregated data and focuses on between sector agglomeration. We propose a new measure for the extent to which firms from two different sectors locate in the same place, or coagglomerate. We examine what this measure reveals about the relative importance of the agglomerative forces at work in Vietnam. We also disaggregate how the magnitude of these forces differs for firms in high and low-tech sectors. We find technology transfers are the most important agglomerative force in Vietnam. We also find evidence that firms in high-tech sectors pool skilled workers while firms in low-tech sectors compete for workers. We repeat our analysis using Ellison and Glaeser's (1997) measure, applied by Ellison et al's (2010) to measure coagglomeration, and find very different underlying stories at work. The contrasting results highlight how finding an

⁴ Notable exceptions include; Henderson (1997) who uses data from the US and Brazil, Ellison and Glaeser (1997) who use US data, and Duranton and Overman (2005) who use UK data.

⁵ For example; Kim (1999) and Ellison and Glaeser (1999) use US data to determine if natural advantages explain the geographic distribution of manufacturing firms; Greenstone et al. (2010) focus on identifying spillovers (measured by productivity) between firms in the US; Overman and Puga (2010) investigate if labour pooling is a source of agglomeration in the UK.

appropriate measure for agglomeration is crucial to the process of identifying agglomerative forces.

Identifying the reasons why firms cluster can be empirically difficult due to reverse causality. In chapter 3 we identify technology transfers as the most important agglomerative force in Vietnam, however it is hard to determine whether firms locate in clusters in order to benefit from technology transfers, or if firms first choose a location and then technology transfers occur with other firms in close proximity. In chapter 4 we focus on new entrants in order to identify strategic clustering and the types of firms new entrants choose to locate beside. We use micro data to conduct a fixed effects dyadic analysis which controls for regional characteristics. This allows us to focus on agglomeration economies in the location choice of the firm. We find that new entrant firms are more likely to locate beside incumbent firms in the same 2 or 4-digit sectors, suggesting the presence of horizontal spillovers. We also find that new entrants choose to locate close to foreign or state-owned incumbents. If the new entrant is in a high-tech sector they are likely to locate beside a high-tech incumbent, supporting the results of chapter 3 that technology transfers are an agglomerative force. Finally, new entrants are more likely to locate beside incumbents that have a different level of production technology. When the new entrant firms are disaggregated into domestic and foreign-owned, the results suggest that foreign-owned firms are more likely to locate beside incumbents with higher levels of technology. No significant relationship is found between domestic entrants and incumbents with higher levels of technology. This suggests that while foreign new entrants locate beside incumbents with higher levels of technology in order to maximise potential spillovers, domestic firms do not appear to engage in this type of strategic clustering.

Chapter 5 concludes the thesis. It summarises the key contributions and results of the thesis. The chapter also discusses the direction that future research will take.

Chapter 2

Are Spatial Networks of Firms Random?⁶

2.1 Introduction

The clustering of economic activities and the resulting agglomeration effects on firm outcomes has received considerable attention in the economics literature. It is well established that the clustering of firms not only facilitates specialisation leading to efficiency improvements, but the concentration of knowledge and the existence of a tight social network can also lead to productivity enhancing technology spillovers (Feldman, 2001; Fujita et al., 1999; Kenney and von Burg, 1999; Baptista, 2000; Krugman, 1991; Markusen and Venables, 1999). Most of the literature focuses on why firms cluster, and the existence and prevalence of clustering is usually taken as given. Two notable exceptions include Ellison and Glaeser (1997) and Duranton and Overman (2005) who provide measures of the extent of clustering in the US and the UK⁷.

In this chapter, we do not examine the reasons why firms cluster but take a step back and examine the overall pattern of clustering of manufacturing firms in Vietnam. Like Ellison and Glaeser (1997) and Duranton and Overman (2005) we believe that it is helpful to first determine that *significant* clustering exists before investigating the exact combination of natural endowments and external economies that lead to the particular pattern of clustering observed. Understanding to what extent the observed location patterns of firms in a country are actual departures from randomness is an important, often overlooked step in understanding whether agglomeration effects are present in an industry. While these questions are relevant in all contexts, they are of particular importance for developing countries at earlier stages of the industrialisation process where most evidence on the spatial distribution of firms is anecdotal.

In this chapter, we attempt to address the gap in the literature in this area in the following ways. First, we analyse the clustering patterns of manufacturing firms in

⁶ This chapter is based closely on a paper co-authored with Carol Newman and Jacco Thijssen.

⁷ Felkner and Townsend (2011) use a spatial statistic test to detect significant spatial clusters in Thailand.

Vietnam, a country that has undergone rapid industrialisation in recent years that has been characterised, anecdotally, by dynamic cluster formation. Second, we propose a new framework for understanding and measuring clustering that builds on the methodology proposed by Duranton and Overman (2005). Third, we make use of a unique data set of the population of registered firms in Vietnam that includes firm location.

We find a high degree of clustering in the manufacturing industry in Vietnam relative to that of a randomly generated network. As using the benchmark of a randomly generated network may overstate the extent of strategic agglomeration (Picone et al., 2009) we also construct a number of other counterfactual spatial distributions based on zoning, population density and the regulatory framework.⁸ In each case we find a high degree of clustering in the spatial network of manufacturing firms relative to the generated network. We find that clustering is driven to a large extent by large enterprises but also by the location of state-owned enterprises. The results of our analysis could be helpful in the design of industrial policy, in particular in designing industrial spatial strategies. To our knowledge no similar analysis has been conducted on firms in this country or in a developing country context.

The remainder of the chapter is organised as follows. Section 2.2 presents a conceptual framework for the formation of spatial clusters. The data and methodological approach are presented in Section 2.3. Section 2.4 presents the results while Section 2.5 describes a range of robustness checks. Section 2.6 concludes the chapter.

2.2 Conceptual Framework

The behaviour of economic agents, such as firms, countries, or individuals, is influenced by their relationships with other economic agents. In turn aggregate outcomes are influenced by relationships between agents. Relationships between and among agents can have a number of different dimensions. In order to meaningfully analyse the impact of an agent's relationships on their decisions, and on aggregate

⁸ We construct separate counterfactual distributions either based purely on zoning, population density or the regulatory framework. To attempt to construct a counterfactual distribution based on all three factors would require assumptions on the relative importance of each factor, and these assumptions would be difficult to justify. The analysis of each counterfactual in turn, although not ideal, is an informative extension on the work of Duranton and Overman, who examine only one counterfactual.

outcomes, we need a framework to systematically examine these relationships. Network analysis provides such a framework. A network describes a collection of nodes and the links between them. Nodes can represent individuals, firms, countries or even collections of such entities and the links between them can be defined in many different ways. For example, a network could consist of individuals as the nodes and define links as the friendships between them, or nodes could be defined as countries and links defined as trade between those countries.

A particular class of complex networks is composed of those embedded in real space, in other words networks whose nodes occupy a precise position in two/three dimensional Euclidean space and whose edges are real physical connections. When nodes and links are embedded in a physical space this induces a distance between nodes and such networks are designated 'spatial'. For example, manufacturing firms in Vietnam form a spatial network where the nodes are firms and the links are the physical distances between them. As the links in the spatial network are defined purely in terms of physical distances there is not necessarily any interaction between the firms. This is suited to our purpose given that we are specifically interested in identifying geographical clustering of firms, regardless of whether or not interactions occur. The position of a firm within the spatial network may impact on the costs and production decisions of the firm itself and on other firms within the network.

The 'symmetric connections model' (JW model) by Jackson and Wolinsky (1996) motivates our approach to analysing cluster formation. The JW model is a social network model where the nodes are agents/players and the links are the friendships between them. The model assumes that players benefit not only from direct links but also from indirect links or 'friends of friends'. This benefit is assumed to deteriorate in the distance between nodes. It is assumed that there is some cost associated with the formation of a direct link and therefore the network is not fully connected. Jackson and Rogers (2005) extend the JW model to include a simple geographic cost structure of link formation whereby agents are based on a number of 'islands' and the cost of forming a link with another agent depends on the geographic location of that agent. This cost structure, together with the deteriorating benefit structure of indirect links, produces a network that exhibits small world characteristics.

Jackson and Rogers (2005) (JR) apply the geographic cost structure extension of the JW model to the case of a social network. We argue that with one simple extension the model is a good fit for a spatial network of firms. Firms in clusters can experience agglomeration effects. For example, the costs of production may fall as firms have multiple competing suppliers or greater specialisation and division of labour can lead to efficiency gains. There can also be diseconomies of agglomeration. For example, a greater number of firms can increase competition and drive down prices. A manufacturing firm is affected by the production decisions of all firms in the network, but the magnitude of the impact decreases as the distance between the two firms increases. Agglomeration effects also decrease with increasing distance. In order to apply the JR model to a spatial network of firms we introduce the concept of a threshold distance and define firms to be linked if they are within a threshold distance, d^* . This implies that firms receive the 'benefit' if a firm is within the threshold distance. For the network of manufacturing firms this can be positive or negative. Firms cannot sever spatial links and they do not have any control over which firms, if any, locate near them, once they have entered the market and made their own location choice. In our extension of the JR model, the 'benefit' is simply the impact of the direct link between two firms i and j , or the effect that i locating within d^* of j has on the firms. Other firms in the spatial network are also effected by i 's location choice but the 'benefit' deteriorates with increasing distance.

On entering the market each firm decides on a location. Their choice of location determines their position in the spatial network and also determines the spatial links they form with other firms. The cost associated with link formation is the cost of locating in a particular area. This cost clearly has a geographic structure as the cost of entering the market will differ depending on the area in which the firm decides to locate. Costs such as rent and land prices will vary from region to region and firms receive different incentives and levels of state investment depending on the area in which they choose to locate. What we aim to establish empirically in this chapter is that this extension of the JR model provides a good explanation of the observed pattern of clustering of firms in Vietnam. Hence, we expect to observe 'small world' characteristics, in other words a high degree of clustering, in the spatial network of manufacturing firms.

2.3 Data and Methodology

The data are taken from the Vietnamese Enterprise Survey for 2002-2007 inclusive, provided by the General Statistics Office of Vietnam (GSO, 2010). The dataset includes all registered manufacturing enterprises at the end of each year with more than 30 employees, plus a random sample of 15 per cent of small registered enterprises with less than 30 employees. The data includes information on the main economic activity of the firm, the type of ownership (state-owned, privately-owned or foreign-owned), the number of employees, and the commune in which the firm is located. There are three levels of administrative areas in Vietnam; communes, districts and provinces. Communes are very small areas so commune level location information is sufficient for the analysis. The diameter of the largest commune is 66.37km. Table 2.1 presents the number of manufacturing firms and the percentage increase in the number of firms in each year.

Table 2.1: Number of manufacturing firms in Vietnam 2002-2007

Year	Number of firms	% Increase
2002	61,997	
2003	71,135	12.8%
2004	91,499	28.6%
2005	109,947	20.2%
2006	128,637	17.0%
2007	155,520	20.9%

Our methodological approach to measuring clustering is similar to that proposed by Duranton and Overman (2005); however we use an alternative measure of clustering which is based on measures used in network analysis. We measure the distance between every set of firms, and define two firms to be linked if the distance between them is less than the threshold distance. In other words, $L_{ij} = 1$ if and only if $|l_i - l_j| \leq d^*$, where l_i is the location of firm i , l_j is the location of firm j , and d^* is the threshold distance. We define a *triad* as a subgraph of three nodes (i, j, k). A triad is called *transitive* if all three nodes are linked ($L_{ij} = 1, L_{ik} = 1$ and $L_{jk} = 1$) and *potentially transitive* if two of these three links exist. The *transitivity* of a network (τ) is defined as the ratio of the number of transitive triads to the number of potentially transitive triads, i.e.

$$\tau = \frac{3 \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N L_{ij} L_{ik} L_{jk}}{\sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N L_{ij} L_{ik} + L_{ij} L_{jk} + L_{ik} L_{jk}} \quad (2.1)$$

The spatial network is assumed to be undirected in that $L_{ij} = L_{ji}$, $L_{ik} = L_{ki}$ and $L_{jk} = L_{kj}$. The denominator is the sum of the transitive triads and the potentially transitive triads. For any group of three firms i , j and k if only two of the three links exist then two of the terms in the numerator will be zero. For example if firm i is linked to firm j and k but firms j and k are not linked the first term in the denominator will be one and the second two terms will zero. If the triad is transitive, in other words if all three potential links exist, all three terms in the denominator will equal 1. To counter this triple counting of the transitive triads in the denominator the numerator is multiplied by three. The transitivity has an upper bound of 1 and a lower bound of zero. If firm i is linked to firm j and firm i is also linked to firm k , the transitivity can be interpreted as the probability that firms j and k are also linked.

The threshold distance is chosen to be the diameter of the largest commune and firms are therefore defined to be linked if the distance between them is less than or equal to 66.34km. The choice of the diameter of the largest commune as the threshold distance is to ensure that if firms are located beside one another but in adjacent communes, they will still be linked. The use of a threshold distance rather than a geographic region, such as a commune, avoids problems associated with firms located on the border of regions (see for example Gorman and Kulkarni (2004)). A matrix of links is then created and the transitivity of the network is calculated. This procedure is repeated for all years.

The measure proposed here is similar to that of Duranton and Overman (2005). They plot the location of each firm and measure the distance between all pairs of firms. They then construct a measure of localisation called a K-density. For each 4-digit sector they compute the density of bilateral distances between all pairs of establishments and compare this density to counterfactuals generated by randomly re-shuffling plants across sites. Their measure assesses departures from randomness in the location of firms in a four-digit sector, conditional on overall manufacturing agglomeration. Our approach to measuring agglomeration departs from Duranton and Overman (2005) in

several ways that represent a number of unique methodological contributions to the literature on clustering and spatial networks.

First, we use transitivity as an alternative measure of agglomeration.⁹ Unlike the K-density measure, the use of a threshold distance allows us to calculate transitivity when data on the location of firms is not exact, as is the case in our study.¹⁰ In addition, as we are interested in the clustering of firms and agglomeration effects, it makes sense in the context of our conceptual framework to choose a threshold distance. The transitivity measure also has the benefit that it is not heavily influenced by outliers.¹¹ The transitivity of a network is a measure commonly used in social network analysis to assess the degree to which nodes tend to cluster together. It is a global measure which gives an overall indication of clustering in the network.¹² Second, rather than analysing 4-digit sub-sectors of the manufacturing industry, we look at the manufacturing industry as a whole. Duranton and Overman (2005) acknowledge that analysing four digit sectors can create problems, as two firms within the same sector do not necessarily produce the same thing¹³, and suggest that one way to overcome this problem is to consider the whole population of manufacturing firms, which is what we do in this chapter. Finally, the transitivity measure of clustering also has the advantage of being comparable not only across industries but also across countries and other networks (provided the same threshold distance is used).

⁹ We do not suggest that transitivity is a superior measure, but rather where it is possible to calculate both, our measure can be thought of as complimentary to the K-density measure.

¹⁰ Our data tell us which commune the firms are in, but not the exact addresses of the firms. Consequently we could not do a full spatial statistical analysis as the exact distance between firms is not known. Given these data limitations the use of a threshold distance is appealing.

¹¹ This makes transitivity preferable to simply measuring the average distance between firms as the arithmetic mean is heavily influenced by outliers.

¹² As transitivity is a global measure it is again preferable to simply measuring the average distance between firms. Consider a spatial network where all firms are located in two clusters. The further apart the two clusters are the higher the average distance will be, even though the overall clustering of the network does not change. The computed transitivity remains the same over increasing distance once the two clusters are more than the threshold distance apart, reflecting the overall pattern of clustering.

¹³ Duranton and Overman (2005) use the example of naval constructions; large firms will mostly be located along the coast, but there may be smaller firms inland, in the same sector, but producing entirely different goods.

2.4 Results

Before investigating the extent of clustering of manufacturing firms in Vietnam, we first plot the location of each firm to generate a visual representation of the location pattern. In order to plot the location of the firms, the midpoint of each commune is calculated. As the exact location of a firm within a commune is unknown, all firms are plotted in the midpoint of the commune in which they are located. As communes are very small in area this has a negligible impact on the plot of the firms.

Figures A.1 to A.6 in the Appendix to this chapter show the plots of the locations of all manufacturing firms for the years 2002 to 2007. The firms are plotted semi-transparently so that the clusters appear darker the more firms are located in a particular commune. Of particular interest is the pattern of clustering of firms. It is easy to see that in 2002 there are two main large clusters, located in the north near Hanoi and in the south near Ho Chi Minh City (HCMC). We can also see that over the time period of the data set new smaller clusters appear along the coast in the east most region in south central Vietnam.

Table 2.2 presents the transitivity of the spatial network of all manufacturing firms for each year of the data set. Consistent with the pattern illustrated on the maps, the transitivity of the network is increasing over the time period of the data. Transitivity can be interpreted in the following way; in 2002, the reported value of 0.9482 indicates that if firm i is linked with firm j and firm i is also linked with firm k , then there is a 94.82 per cent probability that firms j and k are also linked. This is true for all firms i , j and k in the network. Note that because of the way we have defined the links, this means that if firm i is located within 66.34km of both firms j and k , there is a 94.9 per cent probability that firms j and k are located within 66.34km of each other.

We can say that there appears to be a high degree of clustering as the transivities are all large numbers, however this may be purely because there are a large number of firms in a small area. In other words, the more firms in Vietnam the higher we would expect the transitivity to be, and so we need to control for the density of the network. We generate random networks for each year and compare the transitivity of the network of manufacturing firms to the transitivity of the randomly generated networks (RGNs) (Erdos and Renyi, 1960). To generate the random networks for each year we take the

same number of firms from that year and randomly allocate each of them to a commune. This process produces a random distribution of firms in the same way as the ‘dartboard approach’ proposed by Ellison and Glaeser (1997). Using the same threshold distance as for the original networks we then calculate the transitivity for the random distribution of firms. The results for each year are presented in column 3 of Table 2.2. For each year the transitivity of the network of firms is considerably higher than the transitivity of the randomly generated network.

Table 2.2: Transitivity of the network of all manufacturing firms

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9486	0.6869	0.7974	1.381	1.190	0.0032
2003	0.9541	0.6862	0.8004	1.390	1.192	0.0025
2004	0.9667	0.6865	0.7956	1.408	1.215	0.0028
2005	0.9693	0.6857	0.7882	1.414	1.230	0.0028
2006	0.9716	0.6870	0.7797	1.414	1.246	0.0023
2007	0.9675	0.6883	0.7705	1.406	1.256	0.0023

If the data are consistent with our extension of the JR model for the location choice of firms, then we expect to observe a network with ‘small world’ characteristics, meaning a network that exhibits a high degree of clustering relative to that of a randomly generated network. To test if this is in fact the case, we use the benchmark suggested by Watts and Strogatz (1998). They propose that if the ratio of the clustering coefficient (or transitivity) of the actual network to the randomly generated network is greater than 1, the network is a ‘small world’. Column 5 of Table 2.2 presents this ratio for each year of our analysis. We find that the ratio is greater than 1 for all years and hence we can conclude that the network of manufacturing firms in Vietnam exhibits small world characteristics.

In the case of firms in Vietnam it could be argued that generating purely random networks to control for density is not an appropriate comparison. Randomly allocating firms to communes will of course result in firms being allocated to communes that do not in reality have any manufacturing firms, and it may not be possible for firms to locate in these communes due to legal or geographical restrictions. Therefore, to control for the regulatory framework, we follow Duranton and Overman (2005) and consider that the set of all existing sites for firms constitutes the set of all possible locations for any firm. We generate a second random network, denoted RGN*, for each year, where

the firms are randomly allocated only to communes where at least one firm is located in that particular year. However comparing the observed transitivity in each year to a single RGN is somewhat problematic, as a single RGN is just one draw from the distribution of RGNs. To offer a more complete analysis we generate 30 RGN*s and consider the average transitivity. The average transivities of these 30 random networks, for each year, are presented in column 4 of Table 2.2, and the ratio of the transivities of the actual networks to the average transivities of RGN*s is presented in the second last column of Table 2.2. Controlling for the regulatory framework we can still conclude that the networks of manufacturing firms exhibit a high degree of clustering, as evidenced by the fact that the ratio of transivities is greater than one in all time periods.

We wish to test the null hypothesis that firms in Vietnam randomly choose where to locate, from the set of all feasible locations.¹⁴ In order to test this hypothesis we compare the observed spatial network to the mean of the distribution of random networks. We can see from table 2.2 that in all cases the observed network exhibits a high degree of clustering relative to the mean of the random networks. However, in order to conduct a hypothesis test we also need to consider the spread of the distribution of random networks. The last column of Table 2.2 shows the standard deviation (σ) of the distribution of RGN*s.¹⁵ If we assume that the RGN*s are normally distributed, then we know that 99.7 per cent of RGN*s lie within 3σ of the mean.¹⁶ Take 2007 for example, the observed transitivity of 0.9675 is 85.6σ higher than the mean of the RGN* distribution. The associated p-value for the hypothesis test is zero. Therefore we can reject the null hypothesis, that the observed spatial network in 1997 is random, at the 1% significance level. We conclude that the observed spatial network of firms is not random, as it is not drawn from the distribution of random networks. Similar analyses of the other years leads to the same conclusion.

¹⁴ An alternative way of thinking about this is that individuals are randomly scattered across the country, and they choose to establish a firm where they live, so the location of firms is simply random.

¹⁵ Standard deviations are needed to conduct a hypothesis test but they are also useful in quantifying how much greater the observed transitivity is than that the mean transitivity of the randomly generated networks.

¹⁶ The p-value for the Shapiro-Wilk test for normality is 0.31 so we fail to reject the null hypothesis of normality. Therefore our assumption that the RGN*s are normally distributed is justified.

2.5 Robustness Checks and Further Investigation

2.5.1 Firm Size and Statistical Significance

As detailed in Section 2.3, our data includes all large registered enterprises, but only a sample of small enterprises. The dataset includes all registered manufacturing enterprises at the end of each year with more than 30 employees, plus a random sample of 15 per cent of small registered enterprises with less than 30 employees. Therefore, as a robustness check, we remove all firms with fewer than 30 employees and repeat the analysis. These data represent the population of manufacturing firms with more than 30 employees and so the calculated transitivities are parameters rather than statistics. Table 2.3 shows the number of manufacturing firms remaining in our dataset each year after the small firms have been removed. These represent the population of ‘large’ manufacturing firms in Vietnam. The results of the analysis using only those firms with 30 employees or more are presented in Table 2.4. We find very similar measures of clustering to those presented in Table 2.1 when all firms are included in the analysis, revealing that the clustering of manufacturing firms is robust to the exclusion of small firms.

Table 2.3: Numbers of large manufacturing firms

Year	Large Manufacturing Firms
2002	14,794
2003	16,915
2004	20,531
2005	24,044
2006	26,208
2007	31,057

Table 2.4: Transitivity of the network of large manufacturing firms

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9606	0.6896	0.8194	1.393	1.172	0.0049
2003	0.9696	0.6920	0.8488	1.401	1.142	0.0037
2004	0.9750	0.6914	0.8381	1.410	1.163	0.0042
2005	0.9720	0.6915	0.8311	1.010	1.170	0.0040
2006	0.9709	0.6904	0.8238	1.406	1.179	0.0043
2007	0.9644	0.6886	0.8101	1.401	1.190	0.0036

As Duranton and Overman (2005) note, firms with a small number of workers may have different location patterns than firms with a large number of employees. In many industries small firms may be involved in very different activities to much larger firms. It is also possible that the location behaviour of small and large plants simply differ. To further investigate the possibility that firms with different numbers of employees have

different clustering patterns, the firms with 30 employees or more are divided into small, medium and large firms. The large firms are defined as those in the 95th percentile in terms of numbers of employees. These firms represent a large percentage of employment in the manufacturing industry in Vietnam, as revealed in Table 2.5. Small firms are defined as those below the 50th percentile in terms of employment while medium firms are defined as those within the 50th percentile number of employees and above, but below the 95th percentile threshold.

Table 2.5: Employment by size distribution of manufacturing firms

Year	Total Employment	95 th Percentile Lower Bound	Employment in 95 th Percentile	%Total Employment in 95 th Percentile	50 th Percentile Lower Bound
2002	4,221,624	1,152	868,066	20.6%	107.5
2003	2,448,586	1,126	1,022,720	41.8%	106
2004	2,757,753	1,095	1,182,508	42.9%	100
2005	2,938,227	1,050	1,282,116	43.6%	96
2006	3,136,891	1,069	1,402,208	44.7%	97
2007	3,569,523	1,062	1,593,183	44.6%	96

These three groups of manufacturing firms (small, medium and large) are plotted on three different maps for 2002 and 2007. Figures A.7 to A.12 in the Appendix show these plots based on firm size.¹⁷ It is clear that the firms in the 95th percentile are at the very centre of the two largest clusters, near Hanoi in the north and HCMC in the south. In addition, the smaller clusters on the coast that have formed over the period of the dataset appear to be caused by large firms locating there and smaller firms clustering around them. Medium firms are located in the clusters but some are also in seemingly random locations around the country. In contrast, very few small firms are located away from clusters.

2.5.2 Ownership Structure

We also explore the extent to which ownership structure impacts on the pattern of clustering in the manufacturing industry in Vietnam. The ownership structure can have an important impact on the location choice of the firm. State-owned firms may not base their choice of location purely on profit maximisation. As governments are more likely to be directly involved in the location decision of state-owned firms they are more likely to take other factors into consideration, such as spatial strategies and employment generation. Foreign ownership may also impact on the location decision of the firm. A

¹⁷ Colour maps provide a clearer picture of the clustering pattern and are available on request.

number of empirical studies have found that the nationality of the firm ownership is important in location patterns, and that foreign firms are more likely to cluster (see for example Crozet et al, 2004 and Devereux et al, 2007).

We divide the manufacturing firms into three groups, this time defined by their ownership structure. We define state-owned firms as those either wholly or partly owned by the state, private owned firms as those owned wholly by domestic non-state owners, and foreign-owned firms as those either partly or wholly owned by foreigners. We plot the locations of the firms in 2002 and 2007 once again on three separate maps for the three ownership categories.¹⁸ This allows us to see not only the pattern of clustering but also its progression over time. The resulting maps are illustrated in Figures A.13 to A.18 in the Appendix. Once again a clear pattern of clustering is revealed. Private enterprises are very clearly clustering around state-owned enterprises in the two largest clusters while there are very few privately-owned enterprises located outside of the clusters.

If we compare Figures A.15 and A.18 it is clear that the number of foreign-owned enterprises increased over the period of the data set. The raw data reveal a particularly pronounced increase in foreign-owned firms between the end of 2006 and the end of 2007. This sharp increase is most likely due to changes in investment laws that came into effect in July 2006. Prior to July 2006 foreign and domestic investments were governed by two separate laws; the new law was intended to apply equally to domestic and foreign-owned enterprises. The Investment Law introduced in 2006 was considered a significant watershed for improvement of the legal environment on investment activities and corporate governance (Foreign Investment Agency, 2009). The introduction of this new law in 2006 most likely assured and encouraged foreign investors in Vietnam, leading to an increase in the number of firms with some level of foreign ownership. In addition, although the purpose of the new investment law was to ensure both domestic and foreign-invested companies would be regulated in the same way, it has been argued that in some instances the regulatory burden to establish a business under the new law was actually greater for domestic investors (Freshfields Bruckhaus Deringer, 2006). This effect may also have lead to further increases in the

¹⁸ Colour coded maps available on request.

number of foreign-owned firms relative to domestically owned firms between 2006 and 2007.

2.5.3 Location Restriction

It could be argued that privately-owned domestic firms have the greatest freedom in terms of choosing where to locate. The locations of state-owned firms are determined by government rather than management of the firms, and foreign-owned firms are more restricted than domestically owned enterprises in terms of where they are allowed to locate. Therefore the high degree of clustering we observe in the network could be driven by the state-owned and foreign-owned firms and consequently purely a result of government or planning policies. If this is the case then we may observe a less clustered network if we run our analysis for the network of private domestic owned firms only. The results of this analysis using only large domestic firms are presented in Table 2.6.

Table 2.6: Transitivity of the networks of private firms only

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev RGN*
2002	0.9883	0.7698	0.9464	1.284	1.044	0.0068
2003	0.9888	0.7762	0.9386	1.274	1.053	0.0065
2004	0.9908	0.7657	0.9445	1.294	1.049	0.0064
2005	0.9901	0.7718	0.9432	1.283	1.050	0.0065
2006	0.9874	0.7746	0.9423	1.275	1.048	0.0061
2007	0.9844	0.7827	0.9260	1.258	1.063	0.0093

As we can see from the results presented in Table 2.6 when only private firms are used the resulting networks have higher transivities than when all firms are used. So location restrictions on foreign and state-owned firms are not driving the results. We can also say that the observed transivities are significantly higher than the average transitivity of the distribution of transivities from RGNs. If we take 2006 as an example, the observed transitivity is 7.4σ higher than the mean of the RGN* distribution. We can conclude then that the spatial network of private firms is not drawn from the RGN* distribution and exhibits a high degree of clustering.

2.5.4 Economic/Industrial Zones

In 1991 the Vietnamese government introduced a policy to develop special economic zones (Foreign Investment Agency, 2009).¹⁹ There are 96 different zones located across the country. The zones cover large areas and can span a number of communes. The zones have a high quality of infrastructure and zoned land is made available exclusively for industrial use. The government offers fiscal incentives to firms to locate within zones (Foreign Investment Agency, 2009). It is therefore possible that economic zoning is driving the clustering observed in the spatial network of firms.²⁰ To determine if this is in fact the case we split the data set into firms located in districts with economic zones and firms located in districts with no economic zones. We then run the analysis to determine if significant clustering occurs within economic zones and if significant clustering occurs outside economic zones. The results of the analysis are presented in Tables 2.7 and 2.8. Our results reveal that there is significant clustering outside economic zones as well as within economic zones. For example from table 2.7 we can see that in 2004 the within zone transitivity is 40 σ greater than the mean of the RGN* distribution. Table 2.8 shows that in the same year outside the zones the observed transitivity is 91.3 σ greater than the mean of the RGN* distribution. This is evidence that it is not economic zoning that drives the clustering observed in the spatial network of firms in Vietnam. Significant clustering occurs both inside and outside of economic zones.

Table 2.7: Transitivity of the networks of firms within zones

Year	Transitivity <i>Within Zones</i>	Transitivity RGN	Ratio: $c(g)/c(g^{rd})$	Std. Dev. RGN
2002	0.9742	0.8220	1.185	0.004
2003	0.9572	0.7978	1.200	0.003
2004	0.9684	0.8080	1.199	0.004
2005	0.9680	0.8015	1.207	0.004
2006	0.9277	0.8024	1.156	0.004
2007	0.9646	0.7963	1.211	0.003

¹⁹ www.business-in-asia.com/countries/vietnam_industrial_zones.html has a map of Vietnam's industrial zones.

²⁰ It is also possible that other physical characteristics of provinces drive the observed pattern of clustering. We explore this using a linear regression of the density of firms within provinces as the dependent variable and a number of province characteristics such as the number of ports, train stations, airports and industrial zones as explanatory variables. The number of industrial zones is the only significant variable, and so we exclusively focus on this in our analysis. (Results available on request).

Table 2.8: Transitivity of the networks of firms outside zones

Year	Transitivity <i>Outside Zones</i>	Transitivity RGN	Ratio: $c(g)/c(g^{rd})$	Std. Dev. RGN
2002	0.9555	0.8266	1.156	0.007
2003	0.9653	0.8097	1.192	0.005
2004	0.9728	0.7903	1.231	0.002
2005	0.9767	0.7847	1.245	0.004
2006	0.9780	0.7747	1.262	0.003
2007	0.9753	0.7589	1.285	0.003

2.5.5 Choice of Threshold Distance

The results of our analysis may also be driven by the choice of threshold distance in the construction of the spatial network. As an additional robustness check we repeat the analysis for all large firms using approximately half the original threshold distance (33km), and using approximately double the original threshold distance (132km). The results are presented in Tables 2.9 and 2.10, respectively. The choice of threshold distance does not have a significant impact on the results of the analysis. The networks still exhibit a high degree of clustering relative to the RGN.

Table 2.9: Transitivity of the networks of large firms: small threshold

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9385	0.6645	0.8651	1.412	1.085	0.0045
2003	0.9303	0.6635	0.8842	1.402	1.052	0.0042
2004	0.9338	0.6625	0.8742	1.410	1.068	0.0047
2005	0.9273	0.6705	0.8606	1.383	1.078	0.0041
2006	0.9168	0.6610	0.8728	1.387	1.050	0.0040
2007	0.9674	0.6615	0.8376	1.462	1.155	0.0045

Table 2.10: Transitivity of the networks of large firms: large threshold

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9577	0.7722	0.8765	1.240	1.093	0.0025
2003	0.9731	0.7771	0.9017	1.252	1.078	0.0051
2004	0.9736	0.7779	0.9015	1.252	1.073	0.0055
2005	0.9730	0.7760	0.9056	1.254	1.074	0.0018
2006	0.9722	0.7773	0.8984	1.251	1.082	0.0027
2007	0.9708	0.7783	0.8949	1.247	1.085	0.0024

2.5.6 Population Density

It is possible that the pattern of clustering we observe is driven by population density; firms could simply choose to locate where there are workers available to them, and so clusters form in areas with high population density. To investigate if population density drives the observed high degree of clustering in the spatial network of manufacturing

firms we use provincial level population density data for 2008 to construct a set of province location probabilities.²¹ We assume that the higher the population density is, the more likely firms are to locate in a particular province. This approach is similar to that used by Devereux et al. (2004) who consider that a ‘random allocation’ of firms could be determined by a random allocation of individuals who set up firms in the areas in which they reside. They therefore construct counterfactuals based on total population.

Using the same number of firms as in our dataset in 2007 we randomly allocate firms to communes based on the province location probabilities. As before, we define any commune in which a firm is actually located in 2007 to be a commune in which a firm may choose to locate and restrict the random allocation of firms to these communes. We generate 30 random networks based on the province location probabilities and calculate the transitivity of the random network for each. The resulting distribution of random networks has a mean transitivity of 0.7894 and a standard deviation (σ) of 0.0015. If we compare this to the observed transitivity for the spatial network of firms in 2007 of 0.9644 it is 116.7σ greater than the mean of the distribution of RGN*s based on population density. We can therefore conclude that the population density alone is not driving the high degree of clustering observed.

2.6 Conclusion

In this chapter, we take a step forward in the literature by examining the extent to which significant clustering of manufacturing actually occurs in Vietnam. We do this by comparing, statistically, whether the clustering of firms is better explained by a random graph network or by a network generated by our extension of the JR model. Although a random graph network might not be the spatial distribution of firms we would expect in the absence of agglomeration economies, it is an important initial counterfactual in order to establish the statistical significance of clustering. We extend the analysis to also consider counterfactual spatial distributions based on population patterns, economic zoning and the regulatory framework. With the exceptions of Ellison and Glaeser (1997) and Duranton and Overman (2005) most papers take clustering as given and make no attempt to measure the extent and existence of clustering within an

²¹ Data are provided by the General Statistics Office of Vietnam (2008). Data are not available for other years and so it is not possible to complete the analysis for all years of our data.

industry. Any analysis of why firms cluster should begin by determining that the observed pattern of clustering is a significant departure from randomness. This chapter uses tools of network analysis to measure the clustering of manufacturing firms in Vietnam. This research is the first to apply the transitivity measure (a network analysis measure of clustering) to firms, and the first to attempt to measure clustering defined in this way in a developing country context.

We find that there is significant clustering in the spatial network of manufacturing firms in Vietnam. The difference between the observed transitivity and the mean transitivity of the distribution of randomly generated networks is large enough (in terms of standard deviations) in all years to reject the null hypothesis that the manufacturing firms in Vietnam form a random network.

Marshall (1920) proposed three theoretical reasons why firms cluster; the clustering of economic activity reduces the transport costs for firms; where industry is concentrated a large pool of labour will also emerge facilitating better matching of workers to employers; and information and technology spillovers are more likely when firms are geographically located close together. Before we can establish which of these agglomerative forces are at work in a country, we must first establish that there is significant clustering, and that this clustering is not purely driven by the regulatory framework, population patterns or government industrial policy. We have established that this is the case for Vietnam. Manufacturing firms are highly clustered, and this clustering is not driven by economic zoning, population density or location restrictions. This chapter provides valuable insight into the extent and pattern of clustering within Vietnam, and paves the way for similar investigations for other developing countries.

Chapter 3

Measuring industry coagglomeration and identifying the driving forces²²

3.1 Introduction

The geographic clustering of manufacturing activity has long been recognised as an important mechanism for facilitating industrial growth in both developed and developing countries (Krugman, 1991; Markusen and Venables, 1999). Also more recent contributions speak to the issue. Deichmann et al. (2008) use microdata for India and Indonesia and find that agglomeration benefits outweigh the costs of congestion and higher wages in clusters. Collier and Page (2009) examine case studies of firms in Chile, China and Malaysia and find anecdotal support for positive agglomeration externalities in the form of knowledge transfers, productivity gains and the development of a thick labour market. Bigsten et al. (2011) investigate the effects of clustering on firm performance and find that all else being equal, Ethiopian manufacturing firms located in clusters have higher productivity.

Although an extensive literature exists on the benefits to firms in clusters from agglomeration externalities, there is little empirical evidence, particularly in developing country contexts, as to which agglomerative forces are at work within a country and their relative importance. Identifying the driving forces of agglomeration is critical for governments in the formulation of industrial policy.²³ Three well established theoretical reasons for firm clustering exist over and above that which can be explained by natural advantages. First, the clustering of economic activity reduces transport costs and so firms along the supply chain have more incentive to locate near each other. Second, where industry is concentrated a large pool of labour will emerge facilitating better

²² This chapter is based closely on a paper of a similar name co-authored with Carol Newman and Finn Tarp.

²³ See Pack and Saggi (2006) for a full discussion of industrial policy in developing countries.

matching of workers to employers.²⁴ Third, information and technology spillovers are more likely when firms are clustered (Marshall, 1920; Krugman, 1991; Fujita et al, 1999). Ellison et al. (2010) test these theories in the case of manufacturing firms in the US and set a new and welcome standard for empirical testing of agglomerative forces. They propose a measure of coagglomeration that is closely related to the covariance in employment shares between two industries within defined geographical regions.

In chapter 2 we examined the overall pattern of clustering for the manufacturing industry in Vietnam. We used transitivity, a global measure of clustering from network analysis, to test the hypothesis that the spatial network of firms was more clustered than a number of counterfactual distributions. We found that manufacturing firms in Vietnam exhibited a far higher degree of clustering than could be attributed to the regulatory framework, economic zoning or population patterns. Having completed this crucial initial first step in the analysis of the agglomeration of firms in Vietnam, we now turn to developing a new measure for the agglomeration of firms from different sectors, and relating this measure to measures of Marshall's agglomerative forces.

In this chapter, we aim to advance the literature by first proposing a different measure of agglomeration based on the physical location of firms. Like Ellison et al (2010) we focus on agglomeration across sub-sectors or coagglomeration. Agglomerative forces between firms in the same sub-sector of course may also exist but we do not consider these in this chapter. We exploit the fact that we have an exceptional set of data sources for manufacturing firms in Vietnam. Second, we also consider an absolute measure of coagglomeration where clusters are measured in terms of absolute size. Third, our data allow us to test the impact of transport costs, labour pooling and technology spillovers on the clustering of firm activity along the lines of Ellison et al. (2010). Fourth, we are able to capture informal channels of technology diffusion between firms which adds an important dimension to existing studies that have so far focussed on formal or contracted technology transfers (Jaffe, 1986; Ellison et al., 2010). Fifth, our analysis addresses whether agglomerative forces are different for high-tech and low-tech sectors. Sixth and finally, we perform the analysis at three levels of spatial disaggregation (commune, district and province).

²⁴ See Helsley and Strange (1990). Another interpretation is that there is a risk sharing aspect to a large pool of labour and therefore labour market pooling makes workers and firms better off when firms face idiosyncratic demand shocks (Krugman, 1991, Overman and Puga, 2010).

Our key result is that the two most important forces behind agglomeration are technology transfers and skills correlations. The magnitude of the effect of technology transfers would appear to be twice as large in Vietnam as that found by Ellison et al. (2010) for the US. Moreover, we find that technology transfers occur primarily between high-tech firms but also between high-tech and low-tech firms within clusters. Importantly, when the analysis is repeated using the EG measure of coagglomeration, technology spillovers have an almost negligible role to play. This highlights that empirical analysis using different agglomeration measures may lead to different results. It is therefore crucial to consider both the country context, and the construction of agglomeration measures, in order to choose an appropriate measure for use in cluster analysis.

Turning to skills correlations, our results capture both competition for labour and labour pooling which impact on agglomeration in opposite directions.²⁵ It emerges that competition for unskilled labour acts as a negative agglomerative force while the pooling of skilled labour contributes to agglomeration through the clustering of high-tech firms. In contrast, when we use the EG measure of agglomeration, skills correlations are a positive agglomerative force for all firms. This reflects the different way in which the agglomeration measures are defined in this chapter as compared with the EG measure used in Ellison et al. (2010).

We also amend our coagglomeration measure to consider *absolute* agglomeration between sectors. When we repeat the analysis using this alternative measure, in contrast to our results using the relative measure, we find natural cost advantages are the most important agglomerative force at all three levels of measurement. This is as expected when clusters are measured in terms of absolute size. The significance and relative importance of the other three agglomerative forces are consistent with the results using our relative measure.

The remainder of the chapter is organised as follows. In section 3.2 we present our measure of coagglomeration and provide evidence on the extent of coagglomeration of industry pairings in Vietnam. Section 3.3 describes each of the agglomerative forces

²⁵ Combes and Duranton (2003) argue that when firms employ workers from the same local labour market they face a trade-off between the benefits of labour pooling and the costs of labour poaching.

considered and presents the measures used in our analysis. Section 3.4 presents and discusses the results and section 3.5 concludes the chapter.

3.2 Definition and measurement of agglomeration

In spite of the importance attached to agglomeration as a force in economic transformation and development, few attempts have been made in the empirical literature to explicitly define and measure the extent of clustering. A notable exception is the Ellison and Glaeser (1997) (EG) index adapted by Ellison et al. (2010) to measure coagglomeration of two sectors. They use this as the dependent variable in their study of the impact of transport costs, labour correlations and technology spillovers on coagglomeration. Specifically, the EG coagglomeration index for two industries A and B is given by equation (3.1).

$$\gamma_{AB}^C = \frac{\sum_{m=1}^M (s_{mA} - x_m)(s_{mB} - x_m)}{1 - \sum_{m=1}^M x_m^2} \quad (3.1)$$

where m indexes administrative areas; s_{mA} is the share of industry A 's employment in area m ; s_{mB} is the share of industry B 's employment in area m ; and x_m is the mean employment share in the area m across all industries.

The EG measure is a 'model based' index of agglomeration. The index is derived on the basis of a simple model which assumes that agglomeration is a result of a sequence of profit maximising location decisions by individual firms.²⁶ We note that this index is closely related to the covariance of the area-industry employment shares in the two industries. The EG index for two sectors A and B depends not only on the distribution of employment in industries A and B but also on the distribution of employment in all other sectors. This means that even if all firms in sector A and all firms in sector B are located in the same area, the EG index will not necessarily equal 1, even though the sectors are completely coagglomerated.²⁷ The index therefore captures correlations in

²⁶ See Ellison and Glaeser (1997) for a full derivation of the model and the EG index.

²⁷ In the case where all sector pairs are fully clustered in different areas the EG measure will take a value of 1.

the relative size of the two sectors, in terms of employment shares in each area, compared with the relative size of all other sectors in all other areas.

Measuring coagglomeration in this way potentially leads to missing out on an important channel for technology transfer in the empirical analysis. Where high-tech firms are small in terms of number of employees the EG index may fail to identify the relative importance of high-tech clusters. To illustrate, we consider two high-tech clusters of sectors A and B of the same size located in different regions. Both consist of many small firms but in one region they account for a small proportion of overall employment while in another they account for a large proportion of overall employment. In the former case, the relative importance of the high-tech cluster will be less than in the latter, on the basis of the EG measure.

Moreover, as the EG measure for two sectors is closely related to the covariance of the area-industry employment shares, it will also be closely related to the correlation in employment patterns for the two sectors. Therefore high skills correlations will be associated with large values of the EG index and skills correlations are likely to emerge in the econometric analysis as a positive force. Consequently empirical analysis using the EG index as a measure of coagglomeration may fail to capture competition for workers between sectors.

We believe both of these aspects are critical in developing country contexts. When agglomeration is thought of as the clustering of firms regardless of their size there is room for further development of the agglomeration measure. We therefore propose a measure of coagglomeration for use as the dependent variable that is based on the physical location of firms. Accordingly, for every possible set of two sectors A and B we calculate a colocation index which measures the extent to which they are located in the same area.²⁸ We calculate this measure at the three different levels (or areas); commune, district and province. More precisely, for m firms in sector A and n firms in sector B we take each firm i in sector A and sum the number of firms in sector B that

²⁸ Ellison et al. (2010) also consider the exact location of firms in an alternative measure of coagglomeration based on Duranton and Overman's (DO) (2005) index. They find similar results to the EG measure. The DO index requires the Euclidean distance between sets of firms and our data are not detailed enough to compute this. However, we would expect that in a developing country context the results using the two different measures would not necessarily be similar for reasons we explore in this paper.

are located in the same area. We then take the number of colocated pairings as a proportion of all possible pairings across the two sectors (i.e., $m \times n$). This measure will be bound by 0 and 1. The colocation formula is given by equation (3.2).

$$colocation_{AB} = \frac{\sum_{i=1}^m \sum_{j=1}^n C_{ij}}{m \times n} \quad (3.2)$$

where $C_{ij} = 1$ if firms i and j are located in the same area, and 0 otherwise.

We also consider an absolute colocation measure. The absolute formula is given by equation (3.3).

$$AbsoluteColocation_{AB} = \sum_{i=1}^m \sum_{j=1}^n C_{ij} \quad (3.3)$$

where $C_{ij} = 1$ if firms i and j are location in the same area and 0 otherwise. This formula simply counts the number of firms in sector A that are located in the same area as firms in sector B. As the this measure does not control for the total number of firms in sectors A and B it is a measure of absolute coagglomeration. It will therefore take on larger values for larger sectors.

Accordingly, we compute pairwise colocation measures for 43 manufacturing industries in Vietnam using the Enterprise Survey for 2007 provided by the General Statistics Office (GSO, 2010).²⁹ The dataset includes all registered manufacturing enterprises at the end of the year with more than 30 employees, plus a random sample of 15 percent of small registered enterprises with less than 30 employees. Along with the standard financial information the data also include the name of the commune that each firm is located in. There are three levels of administrative areas in Vietnam: communes, districts and provinces. In 2007 there were 10,995 communes, 749 districts and 67 provinces.

Tables 3.1, 3.2 and 3.3 present the top ten colocation pairings for the manufacturing sector in Vietnam for each of the different regions. While there are differences in the

²⁹ The full list of sectors considered are listed in the Appendix. It should be noted that while it is possible using our data to construct coagglomeration indices for 4-digit ISIC sector pairings we are constrained by the level of sector disaggregation available for the other variables used in our analysis. For this reason we must aggregate the 4-digit industry codes into a common set of sector codes that are available for all measures.

important pairwise colocation patterns depending on whether the indices are constructed at the commune, district or provincial level, some distinct patterns emerge. The main sectors that are likely to be coagglomerated with other sectors are the manufacture of various types of machinery. For example, a high level of coagglomeration is found between the manufacture of electrical machinery and the manufacture of precision and optical equipment, sectors where two-way technology spillovers are likely. Similarly, we find a high degree of coagglomeration between chemical products and processed rubber and by-products which may also use common technologies and so potentially may benefit from spillovers. We also find coagglomeration along the value chain. For example, firms manufacturing leather goods, plastics and ready-made apparel are all likely to be colocated with firms manufacturing domestic appliances, suggesting that transport costs of inputs from the former upstream sectors to the latter downstream sector may be a motivating factor. Similarly, the printing and publishing sector is likely to be colocated with sectors that are likely to require information booklets including regulations or instructions for the products that they produce, for example the manufacture of medical and surgical equipment or the manufacture of precision and optical equipment.

The tables also reveal some pairings that are less obvious, such as the high level of coagglomeration between the manufacture of milk and dairy products and the manufacture of machinery used for broadcasting, telecommunications and information. This suggests that in at least some cases natural advantages also play a role in firms' location decisions.

Tables 3.4, 3.5 and 3.6 present the top 10 coagglomeration pairs on the basis of the EG measure at the commune, district and province levels, respectively. Although there are some distinct differences between the pairings depending on the geographical area chosen, some clear patterns emerge. In particular, three sector pairings are highly coagglomerated at all three levels of measurement; electrical machinery and precision and optical equipment; production, processing and preserving of meat and meat products and the manufacture of milk and dairy products; and the manufacture of milk and dairy products and publishing. The most notable difference between the two Tables is that the top ten pairings based on the EG index contain far fewer high-tech sectors than the top pairings based on the colocation index. This supports our contention that the EG index may fail to identify the relative importance of high-tech clusters,

particularly where high-tech sectors employ fewer workers. Although there are clear differences between the most coagglomerated pairings identified by each measure there is also some overlap; for example both measures reveal a high level of coagglomeration between the manufacture of electrical machinery and the manufacture of precision and optical equipment. Manufacture of electrical machinery and of precision and optical equipment are both high-tech sectors and they are large in terms of employment shares. In one district they account for the third and second highest employment shares respectively, over all sectors and over all districts. For this reason both the EG index and the colocation index identify these sectors as being highly coagglomerated.

3.3 Determinants of agglomeration

In this section we describe how we construct measures of each of the three drivers of agglomeration and data sources for these measures. In addition, we describe how we control for natural advantages in our empirical analysis.

3.3.1 Transport costs

Firms wish to minimise their costs in order to maximise profits. Transportation of goods, both to customers and from suppliers, is more expensive the further the goods need to be transported. Therefore firms will have an incentive to locate near to their suppliers and/or their customers in order to reduce costs. To capture the extent to which transport costs matter in the location decisions of firms we use the input-output linkages between different industrial sectors. Following the approach of Ellison et al (2010) we use the Vietnam Supply and Use Tables (SUT) for 2007 (CIEM, 2007).

Table 3.1: Highest colocation pairings: Commune

Sector 1	Sector 2	Commune
05: Milk and dairy products	39: Machinery used for broadcasting, television and information activities	0.005
26: Chemical products	39: Machinery used for broadcasting, television and information activities	0.005
28: Processed rubber and by-products	37: Electrical machinery	0.005
37: Electrical machinery	41: Precision and optical equipment	0.004
26: Chemical products	28: Processed rubber and by-products	0.004
36: Domestic appliances	39: Machinery used for broadcasting, television and information activities	0.004
26: Chemical products	37: Electrical machinery	0.004
27: Pharmaceuticals, medicinal chemicals and botanical products	39: Machinery used for broadcasting, television and information activities	0.004
21: Publishing	40: Medical and surgical equipment	0.004
36: Domestic appliances	41: Precision and optical equipment	0.004

Table 3.2: Highest colocation pairings: District

Sector 1	Sector 2	District
29: Plastic and by-products	36: Domestic appliances	0.030
22: Printing	40: Medical and surgical equipment	0.027
36: Domestic appliances	39: Machinery used for broadcasting, television and information activities	0.027
21: Publishing	40: Medical and surgical equipment	0.027
21: Publishing	41: Precision and optical equipment	0.027
18: Leather and leather products	36: Domestic appliances	0.026
36: Domestic appliances	41: Precision and optical equipment	0.026
21: Publishing	22: Printing	0.025
05: Milk and dairy products	36: Domestic appliances	0.025
17: Ready-made apparel	36: Domestic appliances	0.025

Table 3.3: Highest colocation pairings: Province

Sector 1	Sector 2	Province
36: Domestic appliances	41: Precision and optical equipment	0.410
17: Ready-made apparel	36: Domestic appliances	0.367
17: Ready-made apparel	41: Precision and optical equipment	0.346
05: Milk and dairy products	36: Domestic appliances	0.342
29: Plastic and by-products	36: Domestic appliances	0.340
18: Leather and leather products	36: Domestic appliances	0.336
21: Publishing	41: Precision and optical equipment	0.328
22: Printing	41: Precision and optical equipment	0.327
29: Plastic and by-products	41: Precision and optical equipment	0.325
05: Milk and dairy products	41: Precision and optical equipment	0.318

Table 3.4: Highest pairings using EG Index: Commune

Sector 1	Sector 2	Commune
37: Electrical machinery	41: Precision and optical equipment	0.084
08: Prepared feeds for animals	41: Precision and optical equipment	0.055
04: Vegetable and animal oils and fats	39: Machinery used for broadcasting, television and information activities	0.040
38: Electrical equipment	41: Precision and optical equipment	0.038
01: Production, processing and preserving of meat and meat products	05: Milk and dairy products	0.037
36: Domestic appliances	41: Precision and optical equipment	0.029
05: Milk and dairy products	21: Publishing	0.027
08: Prepared feeds for animals	37: Electrical machinery	0.023
34: General purpose machinery	40: Medical and surgical equipment	0.022
05: Milk and dairy products	28: Processed rubber and by-products	0.021

Table 3.5: Highest pairings using EG Index: District

Sector 1	Sector 2	District
37: Electrical machinery	41: Precision and optical equipment	0.143
05: Milk and dairy products	21: Publishing	0.073
08: Prepared feeds for animals	37: Electrical machinery	0.061
08: Prepared feeds for animals	41: Precision and optical equipment	0.059
01: Production, processing and preserving of meat and meat products	05: Milk and dairy products	0.039
01: Production, processing and preserving of meat and meat products	21: Publishing	0.035
04: Vegetable and animal oils and fats	39: Machinery used for broadcasting, television and information activities	0.035
38: Electrical equipment	41: Precision and optical equipment	0.030
34: General purpose machinery	40: Medical and surgical equipment	0.030
01: Production, processing and preserving of meat and meat products	25: Fertilizers	0.028

Table 3.6: Highest pairings using EG Index: Province

Sector 1	Sector 2	Province
01: Production, processing and preserving of meat and meat products	05: Milk and dairy products	0.194
37: Electrical machinery	41: Precision and optical equipment	0.124
01: Production, processing and preserving of meat and meat products	21: Publishing	0.123
11: Alcohol and liquors	35: Special purpose machinery	0.121
02: Processing and preserving of fish and fish products	06: Processing of rice and flour	0.100
21: Publishing	22: Printing	0.099
11: Alcohol and liquors	21: Publishing	0.095
04: Vegetable and animal oils and fats	05: Milk and dairy products	0.085
01: Production, processing and preserving of meat and meat products	14: Cigarettes and other tobacco products	0.082
05: Milk and dairy products	21: Publishing	0.082

The SUT provides information on the use of 138 commodities in 112 production activities. Commodities and production activities are mapped to the 4-digit ISIC codes used in the Vietnamese Enterprise Survey and 43 manufacturing industry codes that are comparable across the two datasets are generated. Details of these codes are provided in the Appendix. The information in the SUT is aggregated to the new industry codes. We consider three different measures of input-output linkages as in Ellison et al. (2010). First, we consider the flow of inputs between each industry pairing. We compute the proportion of total inputs that sector A buys from sector B and vice versa and take the maximum as a measure of the degree of input linkages between the two sectors. Second, we consider the proportion of total output that sector A provides to sector B and vice versa and take the maximum as a measure of the degree of output linkages between the two sectors. Third, we take the maximum of the input and output measures to produce the variable $InputOutput_{AB}$. Finally we also take the maximum of the input and output measures in absolute terms, measured as the maximum of the total value of inputs/outputs that A sells/buys from B and vice versa. We expect that firms that are highly linked on the basis of this measure will be more likely to cluster.

3.3.2 Labour market pooling

To assess the importance of labour market pooling we examine the correlation in the types of occupations of workers employed in each of the different sectors. We use the Vietnam Household Living Standards Survey (VHLSS) for 2006 which is a representative sample of all households in Vietnam (GSO, 2006). The database contains information on the occupation of household members together with the 2-digit sector in which they are employed. These different occupation groups represent different skills sets that employers require from their workers. There are 32 different occupation groups specified. In order to measure the extent to which two different sectors employ workers with similar skills sets, and therefore have incentives to locate near a common ‘pooled’ labour market, we calculate the correlations between the employment patterns of all sets of two sectors. First, we calculate the share of each sector’s employment in each occupation (o) $Share_{AO}$. We then calculate the correlation between $Share_{AO}$ and $Share_{BO}$ across all sectors A and B. This results in the $SkillsCorrelation_{AB}$ variable.

3.3.3 Technology spillovers

Of the three main agglomerative forces technology spillovers are the hardest to quantify or measure. Measuring the flow of ideas between industries is difficult although a number of different proxies have been used in the literature. Audretsch and Feldman (1996) use industry R&D, university R&D and skilled labour as measures of knowledge spillovers. Greenstone et al (2010) quantify agglomeration spillovers as the change in total factor productivity (TFP) experienced by incumbent manufacturing firms when a large manufacturing firm locates in the same area. Ellison et al. (2010) use two measures. First, they use a technology matrix similar to the input-output matrix for the US which captures how R&D activity in one industry flows out to benefit another industry. Second, they use a patent database to construct measures of patents in and out of pairs of sectors. These are however imperfect measures as they only capture official exchanges of technology. It is likely that technology exchanges are more informal or accidental, particularly in a developing country context.

We go further in our technology spillover measure using a specially designed module on technology usage in the Vietnamese manufacturing sector included in the 2010 Vietnamese Enterprise Survey (GSO, 2010). All surveyed manufacturing firms are asked the question ‘Do most contracts include technology transfer from the supplier to the enterprise?’ If they answer yes to this question they are asked whether the technology transfer is mainly ‘intentional and part of the legal contract’, ‘intentional but not part of the legal contract’ or ‘unintentional’. The question therefore captures both formal and accidental or informal technology transfers from the supplier to the enterprise. We construct our technology transfer variable by calculating the proportion of firms in each sector that answer yes to this question and weight this by firm revenue. To construct industry pairings we map each sector’s weighted technology transfer variable to each of the other sectors by interacting it with a measure of the proportion of inputs it buys from the other sectors. In other words, for sector A we multiply the weighted measure of technology transfer for the sector by the proportion of sector A’s inputs that come from sector B to produce the measure $TechTransfer_{AB}$. If technology transfer from suppliers to firms is common in a sector, then we would expect firms in that sector to locate near their suppliers. Therefore, a positive relationship between our technology variable and the coagglomeration index would provide evidence that technology transfer is an important agglomerative force.

3.3.4 Natural advantages

In this chapter, we hypothesise that clustering may be driven by transport costs, labour market pooling and technology transfers. In addition, some areas may have natural advantages over others that result in cluster formation that are important to control for in our analysis. For example areas that are rich in minerals will attract clusters of mining companies. Following Ellison and Glaeser (1999) we develop a predicted spatial distribution of firms based upon cost differences between regions (commune/district/province). We construct a set of probabilities for each region which captures the probability that a firm will locate in that region if cost is the only factor in its location decision. From the Enterprise survey we take data on the average wage paid by firms and the percentage tax that they pay (calculated by tax paid divided by total revenue). We then express the cost per region as a percentage of the total costs faced by firms in Vietnam. As firms are more likely to locate in a region with lower costs we take the reciprocal of this percentage and compute location probabilities. We then randomly allocate firms to regions in Vietnam using these probabilities. We calculate the colocation measures for this predicted spatial distribution of firms and use this variable as a control for differences in costs across regions in the analysis.

3.4 Descriptive Statistics

Table 3.7 presents summary statistics for the dependent and explanatory variables used in our analysis. The colocation measure is calculated at the three levels of administrative area. The mean and the maximum values of the colocation measure increase as the size of the administrative area increases as is expected. Descriptive statistics for the EG index calculated at the commune, district and province levels is also presented. The mean of the EG index is approximately zero. This is largely by definition as x_m , the measure of an area's 'size', is the share of manufacturing employment, so the deviation of each sector from the benchmark will be approximately uncorrelated with the average of the deviations of all other industries (Ellison et al, 2010).

Table 3.7: Descriptive statistics

	Mean	Std. Dev	Min	Max
<i>Colocation Measures</i>				
Colocation (Commune)	0.001	0.001	0.000	0.005
Colocation (District)	0.009	0.005	0.000	0.030
Colocation (Province)	0.110	0.071	0.001	0.410
<i>EG Measures</i>				
EG (Commune)	0	0.013	-0.026	0.224
EG (District)	0	0.149	-0.374	0.208
EG (Province)	0	0.338	-0.108	0.247
<i>Natural Advantage Measures</i>				
Cost advantage (Commune)	0.001	0.0002	0.000	0.002
Cost advantage (District)	0.001	0.0002	0.000	0.002
Cost advantage (Province)	0.016	0.001	0.009	0.032
<i>Marshallian Factors</i>				
Input-Output maximum	0.046	0.113	0.000	0.893
Technology Transfer	0.002	0.009	0.000	0.124
Skills correlation	0.460	0.358	-0.166	1.000

The natural advantage proxy (i.e. colocation calculated for the hypothetical spatial distribution of firms, generated based purely on costs) is of a much lower magnitude at all regional definitions than for the actual distribution of firms. This suggests that agglomerative forces go beyond pure cost advantages associated with regions. The input-output maximum is expressed as a fraction and so is bound by 0 and 1. It has a mean of 0.046 which means that, on average, approximately 5 percent of inputs/outputs are supplied/purchased between sector pairs. This is much higher than Ellison et al's (2010) measure for the US of 0.007 suggesting that, on average, less than 1 percent of inputs/output are traded between 3-digit sector pairs. Our data also reveal that there are a number of sectors that do not buy or sell any goods to one another. The maximum, however, is high at 0.893 suggesting that there is a lot of variation between sector pairs in the extent of input-output linkages and some sectors are particularly well integrated.³⁰

The skills correlation measure is bound by -1 and 1, where 1 is perfect positive correlation. The mean value is 0.46 suggesting a relatively high degree of correlation in the types of workers that different sectors employ. This measure is also comparable to Ellison et al's (2010) who find a mean of 0.47 in the correlation between occupation

³⁰ Ellison et al (2010) find a maximum of 0.823 for their measure of input-output linkages between 3-digit sector pairs in the US.

types among sectors. Finally, the technology transfer variable, which is weighted by revenue, has a mean of 0.002, a minimum of zero (implying that no technology transfers occur between some sectors) and a maximum of 0.124. The higher this value is the greater are the technology linkages between sectors.

3.5 Empirical Results

Our core empirical model is given in equation (3.4).³¹

$$\begin{aligned} Colocation_{AB} = & \beta_0 + \beta_1 CostColocation_{AB} + \beta_2 InputOutput_{AB} \\ & + \beta_3 TechTransfer_{AB} + \beta_4 SkillsCorrelation_{AB} + e \end{aligned} \quad (3.4)$$

Each of the variables are transformed to have unit standard deviation for ease of comparison of the estimated coefficients on each of the different variables, and to assess the relative importance of each factor in explaining overall coagglomeration patterns. As our unit of analysis is sector pairings residuals may be correlated; to correct for the cross-observation correlation in the error terms involving the same sectors, we report bootstrapped standard errors. We use a nonparametric approach which involves drawing 50 random samples, with replacement, from the data. For each of these random samples we estimate the regression coefficients. We then calculate the sample standard deviation of the sampling distribution to obtain the bootstrapped standard errors. Bootstrapping also deals with the econometric issues arising from the use of the generated regressor, *CostColocation_{AB}* (Pagan, 1984; Ellison et al, 2010).

Table 3.8 presents the results of the regressions where *Colocation_{AB}* and *CostColocation_{AB}* are measured at different levels of geographical aggregation (commune, district and province). We find that coagglomeration when measured at the commune level is not determined by cost advantages. The positive and significant coefficient on the input-output index in Column (2) suggests that the more linked along the value chain the sectors are the more likely they are to be coagglomerated. Column (3) reveals that the relationship between input-output linkages and coagglomeration may be motivated by technology transfers between firms. This is confirmed in columns (5) and (6) by the fact that when these variables are included together the positive effect on coagglomeration is solely due to the technology transfer variable. This result also

³¹ This model is estimated using OLS. We also estimate a generalized linear model which imposes a logistic distribution on the dependent variable and achieve the same results.

holds when the output-index, the input-index and the absolute input/output values are considered in isolation (results available on request). At the commune level our results suggest that a one unit standard deviation increase in the technology transfer variable is associated with a 0.215 standard deviations increase in the colocation value. Technology transfers are the most important agglomerative force at the commune level.

Table 3.8: OLS regression of the determinants of coagglomeration using colocation index

	(1)	(2)	(3)	(4)	(5)	(6)
Commune Level						
Natural Advantage	0.025 (0.032)				0.029 (0.035)	0.020 (0.047)
Input-Output maximum		0.069** (0.031)			-0.034 (0.034)	-0.040 (0.035)
Technology Transfer			0.178*** (0.055)		0.195*** (0.049)	0.215*** (0.068)
Skills Correlation				-0.154*** (0.038)		-0.162*** (0.035)
R-squared	0.001	0.005	0.031	0.025	0.033	0.071
Observations	946	946	903	703	903	703
District Level						
Natural Advantage	0.080** (0.034)				0.092*** (0.035)	0.089* (0.052)
Input-Output maximum		0.043 (0.027)			-0.024 (0.029)	-0.013 (0.037)
Technology Transfer			0.111*** (0.045)		0.122*** (0.048)	0.135*** (0.053)
Skills Correlation				-0.223*** (0.034)		-0.230*** (0.034)
R-squared	0.006	0.002	0.012	0.048	0.021	0.070
Observations	946	946	903	703	903	703
Province Level						
Natural Advantage	0.021 (0.036)				0.032 (0.046)	-0.034 (0.067)
Input-Output maximum		0.020 (0.028)			-0.035 (0.029)	-0.020 (0.027)
Technology Transfer			0.076** (0.034)		0.093** (0.047)	0.105** (0.047)
Skills Correlation				-0.313*** (0.033)		-0.319*** (0.041)
R-squared	0.0005	0.0004	0.006	0.093	0.008	0.103
Observations	946	946	903	703	903	703

Bootstrapped standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Variables are transformed to have unit standard deviation for ease of interpretation

In contrast to expectations we find that the correlation in occupations across sectors has a negative effect on the extent of coagglomeration. This means that in Vietnam sectors requiring similar skill sets are not likely to locate close to each other. There are two possible explanations for this. First, employers may believe that by locating close to other firms that demand similar skills of their workers, they run the risk of their

employees being poached by other firms. Second, if specific skills are in high demand within a commune it will place upward pressure on wages which firms will try to avoid. In other words, the negative relationship between the labour correlation variable and coagglomeration suggests that firms are competing for labour rather than pooling labour. Locating in the same area as firms who employ similar types of workers can lead to the loss of some key workers to competing firms and a higher wage bill to retain others. Our results suggest that correlations in skills demand across sectors is a deterrent to colocation within communes.

Focussing on the district level we find in contrast to the commune level definition, that natural advantages play a small but significant role in the colocation decision. This may be explained by the fact that cost advantages are similar across communes but not across districts. As such, when the region of analysis is the district, natural advantages will play a greater role. The role of technology transfer remains robust to the district level definition of coagglomeration suggesting that even for larger geographical units technology transfer plays an important role in the location choice of firms. A negative relationship is also found between the labour correlation measure and the coagglomeration index giving further support to our suggestion that firms compete for labour and factor this into their location decisions. Labour correlation effects are larger when measured at the district level while technology transfer effects are smaller; a one standard deviation increase of technology transfer is associated with a 0.135 standard deviation increase in colocation while a unit standard deviation increase in the labour correlation is associated with a 0.230 standard deviation decrease in colocation.

The results for the province level definition reveal that cost advantages no longer explain the level of coagglomeration. Taken with the results for the commune and district levels, this suggest that natural advantages are not a consideration made by firms in deciding either which province to locate in or which specific commune within provinces to locate in, but are a factor in deciding between districts. The important role of technology transfers is once again evident in the province level regressions as is the negative relationship between the labour correlation variable and coagglomeration.

We also consider the absolute measure of colocation given in equation (3.3). We substitute this absolute measure for the relative measure in our regression equation

(3.4). Table 3.9 presents the results of this analysis at each of the three levels of geographic area.

Table 3.9: OLS regression of the determinants of coagglomeration using absolute colocation measure

	(1)	(2)	(3)	(4)	(5)	(6)
Commune Level						
Natural Advantage	0.874*** (0.080)				0.872*** (0.072)	0.870*** (0.070)
Input-Output maximum		0.099** (0.042)			-0.088 (0.021)	-0.011 (0.029)
Technology Transfer			0.093** (0.049)		0.078*** (0.029)	0.082** (0.035)
Skills Correlation				-0.057 (0.043)		-0.073*** (0.021)
R-squared	0.763	0.013	0.008	0.003	0.767	0.761
Observations	946	946	903	703	903	703
District Level						
Natural Advantage	0.887*** (0.067)				0.886*** (0.061)	0.892*** (0.073)
Input-Output maximum		0.083** (0.041)			-0.010 (0.016)	-0.011 (0.025)
Technology Transfer			0.059** (0.030)		0.044** (0.019)	0.053** (0.023)
Skills Correlation				-0.072* (0.042)		-0.087*** (0.019)
R-squared	0.787	0.007	0.003	0.003	0.787	0.785
Observations	946	946	903	703	903	703
Province Level						
Natural Advantage	6.42*** (0.567)				6.33*** (0.571)	6.46*** (0.694)
Input-Output maximum		0.064* (0.035)			-0.025 (0.022)	-0.027 (0.027)
Technology Transfer			0.047 (0.037)		0.040*** (0.015)	0.046*** (0.018)
Skills Correlation				-0.088** (0.042)		-0.100*** (0.021)
R-squared	0.707	0.004	0.002	0.006	0.707	0.705
Observations	946	946	903	703	903	703

Bootstrapped standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Variables are transformed to have unit standard deviation for ease of interpretation

In contrast to our results when using the relative measure of coagglomeration, we find that natural cost advantages are the largest agglomerative force at all three areas of measurement. This result is unsurprising when we think of clustering in absolute terms. Firms in large industries face more competition and so care more about costs than firms in less competitive industries with fewer firms. The absolute measure captures this effect at all three levels of aggregation and the magnitude of the effect is greatest when measured at the province level. This makes sense as cost differences are greatest across provinces. The results for input-output links, technology transfer and skills correlation

are in agreement with those found when using the relative measure of coagglomeration; input-output links are not significant at any level of measurement; skills correlations are significant and negative at all three levels increasing in magnitude with increasing geographic area; technology transfer is positive, significant and decreasing with increasing geographic area. These results suggest that even if we think of clustering in absolute terms, technology transfers and skills correlations are important coagglomeration forces in Vietnam.

The above summarizes our econometric analysis of agglomeration in Vietnam as we have defined it. We now turn to bringing into focus how our results differ from the results that would emerge based on the EG approach. We therefore substitute the EG measure of agglomeration (γ_{AB}) given in equation (3.1) for our measure in the empirical model given in equation (3.4). Table 3.10 presents the results of this analysis for all three levels of geographical aggregation.

Table 3.10: OLS regression of determinants of coagglomeration using EG measure

	(1)	(2)	(3)	(4)	(5)	(6)
Commune Level						
Natural Advantage	0.184** (0.100)				0.175 (0.108)	0.165 (0.110)
Input-Output maximum		-0.004 (0.026)			-0.032 (0.022)	-0.036 (0.023)
Technology Transfer			0.038** (0.016)		0.044*** (0.014)	0.046** (0.025)
Skills Correlation				0.147*** (0.028)		0.127*** (0.033)
R-squared	0.034	0.000	0.0014	0.031	0.033	0.074
Observations	946	946	903	703	903	703
District Level						
Natural Advantage	0.124** (0.068)				0.117** (0.061)	0.119** (0.072)
Input-Output maximum		0.001 (0.022)			-0.017 (0.025)	-0.013 (0.024)
Technology Transfer			0.031** (0.017)		0.033 (0.021)	0.029 (0.027)
Skills Correlation				0.102*** (0.027)		0.087*** (0.031)
R-squared	0.0155	0.000	0.0009	0.018	0.015	0.045
Observations	946	946	903	703	903	703
Province Level						
Natural Advantage	-0.067 (0.046)				-0.041 (0.044)	-0.031 (0.036)
Input-Output maximum		0.015 (0.028)			0.014 (0.033)	-0.039 (0.034)
Technology Transfer			0.015 (0.019)		0.008 (0.022)	0.036 (0.025)
Skills Correlation				0.062* (0.036)		0.065* (0.039)
R-squared	0.0044	0.0002	0.0002	0.076	0.002	0.007
Observations	946	946	903	703	903	703

Bootstrapped standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1
 Variables are transformed to have unit standard deviation for ease of interpretation

As before, all variables are normalised to have a standard deviation of one for ease of comparison and bootstrapped standard errors are reported in parenthesis. At the commune level, only the measures of technology transfer and skills correlations are found to be statistically significant agglomerative forces; skills correlations is the stronger of the two. Contrary to the analysis using our measure, the labour correlation variable is found to have a positive effect on the EG measure of coagglomeration. This effect is also evident, decreasing in magnitude, at the district and province levels. In contrast with our results above, technology transfers, although significant at the commune level, are small in magnitude. The effect of the technology transfers at the commune level is almost five times larger when the colocation index is used to measure

agglomeration. At the district and province levels, the analysis using the EG measure does not find any evidence that technology transfers are a significant agglomerative force, contrary to our results above. We find no evidence that input-output linkages are a significant agglomerative force using either measure of coagglomeration³². This does not necessarily mean that transport costs and proximity to suppliers or customers are not an important factor in firms' location decisions. Our results only suggest that firms do not locate near *domestic* suppliers or customers in other sectors. Input-Output linkages may however be a force for *within* sector coagglomeration. Additionally, firms that import their inputs or export their outputs may consider transport costs in their location decisions and so locate near to ports or airports; this is not captured by our input-output measure.

To investigate further and in particular to explore the role of technology transfers, we split the sample into high-tech and low-tech sectors.³³ We consider three subsets of sector pairings: (i) both sectors are low-tech; (ii) both sectors are high-tech; and (iii) the two sectors are respectively high and low-tech. Table 3.11 shows the results for each pairing when our colocation measure is used as the dependent variable.

³² We run the analysis using each of the three relative input-output measures and the absolute input-output measure described in section 3.1 and we find no significant effect on coagglomeration.

³³ Full list of sectors is available in the appendix. Sectors 1-22 inclusive, plus sector 43, are defined as low-tech.

Table 3.11: OLS regression of determinants of coagglomeration for tech pairings using colocation index

	<i>Commune</i>	<i>District</i>	<i>Province</i>
Low-tech			
Natural Advantage	0.206*** (0.076)	0.252*** (0.068)	-0.105 (0.100)
Input-Output maximum	0.060 (0.070)	0.093 (0.078)	0.087 (0.069)
Technology Transfer	-0.083 (0.067)	-0.118 (0.078)	-0.093 (0.083)
Skills Correlation	-0.208*** (0.062)	-0.290*** (0.065)	-0.356*** (0.064)
R-squared	0.09	0.16	0.16
Observations	171	171	171
High-tech			
Natural Advantage	-0.030 (0.077)	0.026 (0.105)	-0.003 (0.091)
Input-Output maximum	-0.108 (0.071)	-0.131 (0.081)	-0.263*** (0.091)
Technology Transfer	0.393*** (0.082)	0.282*** (0.065)	0.239*** (0.075)
Skills Correlation	0.153** (0.080)	0.144** (0.075)	-0.006 (0.072)
R-squared	0.15	0.08	0.06
Observations	171	171	171
Mixed pairings – high-tech/low-tech			
Natural Advantage	0.022 (0.082)	0.098 (0.079)	-0.073 (0.070)
Input-Output maximum	-0.028 (0.051)	-0.037 (0.054)	-0.001 (0.057)
Technology Transfer	0.157*** (0.047)	0.145*** (0.051)	0.102* (0.060)
Skills Correlation	-0.176*** (0.044)	-0.245*** (0.054)	-0.324*** (0.048)
R-squared	0.06	0.09	0.11
Observations	361	361	361

Bootstrapped standard errors are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Variables are transformed to have unit standard deviation for ease of interpretation

For low-tech sector pairings the technology transfer variable is not significant at the commune, district or province levels, suggesting that technology transfers are not an agglomerative force for between sector agglomeration for low-tech firms. The labour correlation measure is negative and significant and increases as the area increases from commune to province. This implies importantly that low-tech firms choose to locate away from other low-tech firms with similar skills requirements, so these firms compete for, rather than pool, unskilled workers. Natural cost advantages are a significant positive force at the commune and district levels but not at the province levels. This means that low-tech firms consider cost differences between regions when deciding in which commune or district to locate but cost differences do not play a role in province locations.

For high-tech sector pairings the technology transfer coefficient is positive and statistically significant at all three geographical levels, decreasing in magnitude as the area increases. The magnitude of the impact of technology transfers is greater for the high-tech sector pairings than for the whole data set and is the most important agglomerative force for high-tech firms. The labour correlation coefficient is positive and significant at the commune and district levels. This suggests that high-tech firms locate close to other high-tech firms with similar skills requirements and consequently pools of skilled workers emerge in agreement with Ellison et al.'s (2010) findings for the US.

For mixed sector pairs, the technology transfer coefficient is also positive and significant at all three levels and decreases with the size of area. This implies that technology transfers are an agglomerative force for the colocation of low-tech firms with high-tech firms from other sectors. This suggests that technology transfers occur from high-tech to low-tech firms. As we would expect the magnitude of the technology transfer force is smaller between high and low-tech sectors than for the high-tech sector pairings. The labour correlation coefficient is negative which implies that if a high-tech firm has similar labour requirements to a low-tech firm they are more likely to locate in different geographic areas. This is further evidence that firms compete for unskilled labour; if a high-tech firm has similar labour requirements to a low-tech firm this is likely because the high-tech firm requires a large number of unskilled workers. For example computer manufacturers may require low skilled workers for assembly line or packaging tasks. If the high-tech firm has low skilled worker requirements then they locate in areas where there are fewer low-tech firms and hence low skilled workers are cheaper and more abundant.

We finally conduct the analysis for the three subsets of sector pairings using the EG measure as the dependent variable. Table 3.12 shows the results.

Table 3.12: OLS regression of determinants of coagglomeration for tech pairings using EG measure

	<i>Commune</i>	<i>District</i>	<i>Province</i>
	Low-tech		
Natural Advantage	0.528*** (0.072)	0.442*** (0.083)	0.007 (0.128)
Input-Output maximum	0.005 (0.034)	0.017 (0.066)	-0.088 (0.127)
Technology Transfer	0.007 (0.037)	-0.063 (0.068)	-0.004 (0.124)
Skills Correlation	0.033 (0.030)	0.041 (0.048)	-0.004 (0.049)
R-squared	0.42	0.18	0.01
Observations	171	171	171
	High-tech		
Natural Advantage	0.128 (0.098)	0.118 (0.078)	0.062 (0.081)
Input-Output maximum	-0.085 (0.053)	-0.031 (0.036)	0.003 (0.060)
Technology Transfer	0.084* (0.047)	0.071 (0.051)	0.117 (0.044)
Skills Correlation	0.102 (0.081)	0.039 (0.064)	0.052 (0.091)
R-squared	0.07	0.06	0.02
Observations	171	171	171
	Mixed pairings – high-tech/low-tech		
Natural Advantage	0.152 (0.198)	0.093 (0.089)	-0.084 (0.053)
Input-Output maximum	-0.044 (0.031)	0.009 (0.040)	-0.020 (0.042)
Technology Transfer	-0.001 (0.043)	0.019 (0.033)	-0.004 (0.046)
Skills Correlation	0.128*** (0.045)	0.101*** (0.033)	0.065 (0.050)
R-squared	0.04	0.03	0.01
Observations	361	361	361

Bootstrapped standard errors are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Variables are transformed to have unit standard deviation for ease of interpretation

First, for low-tech pairings natural cost advantages are significant at the commune and district levels but not at the province level, and are the only significant agglomerative force. Contrary to the results when using our index there is no evidence of low-tech firms competing for unskilled workers. Second, for high-tech pairings, technology transfers are significant only at the commune level, and small in magnitude; once again the effect of the technology transfers for high-tech firms is almost five times larger when the colocation index is used to measure agglomeration. This reinforces our understanding that the EG measure does not capture the important role of technology transfers between high-tech firms as an agglomerative force, nor does it adequately capture the pooling of skilled labour. Third, for mixed sector pairings, skills correlations are a positive force at the commune and district levels. This is in contrast to

the results using our measure of agglomeration, labour pooling rather than labour competition appears as a driving force, while technology transfers from high-tech to low-tech firms are not captured.

3.6 Discussion and Conclusion

Understanding firm agglomeration is analytically challenging and the existing empirical evidence is fragmented and scarce. Ellison et al. (2010) advanced the literature by putting forward a coherent overall framework linking agglomeration to three drivers in the case of the US: transport costs, labour pooling and technology transfers. In this chapter, we have explored the implications of an alternative coagglomeration measure for our understanding of the impact of agglomerative forces. We also developed a measure of technology transfers that encompasses both formal and informal channels which we believe are important in developing country contexts. Furthermore, we disaggregated how agglomeration forces differ for high-tech and low-tech sector pairings. Our contribution was made possible by an unusually rich set of data for Vietnam. Interesting differences in results emerge.

The two major differences between the analysis using the EG index and the analysis using our measure are the ways in which technology transfer and labour pooling impact on agglomeration. First, using our measure we find that technology transfers are an important agglomerative force. The magnitude of the force decreases the larger the geographical area considered, consistent with expectations. When the EG measure is used the technology transfer variable plays a negligible role. This notable difference can be explained by the fact that technology spillovers need not necessarily depend on the number of employees in a firm but could take place between many small firms that form a cluster. In fact, in developing countries, firms that are more technologically advanced may have fewer employees³⁴. Large firms tend to be labour intensive simply because they use low levels of technology and employ large amounts of unskilled labour. Therefore the EG measure, by construction, will not capture the effect of technology spillovers on coagglomeration choices of these firms. The results of the high-tech/low-tech split suggest that technology transfers are an important agglomerative force between high-tech sectors and also between high-tech and low-

³⁴ In our data the mean number of employees for a high-tech firm is 92 while the mean number of employees for a low-tech firm is 143.

tech sectors. As might be expected, technology transfers are not found to be an agglomerative force for low-tech sectors. This is consistent with the results for the EG measure which gives greater weight to larger low-tech firms in its construction.

Second, using our agglomeration measure we find a negative and significant effect of the skills correlation variable on coagglomeration. When the data are split into high-tech/low-tech sector pairings, it emerges that the pooling of skilled labour is an important agglomerative force for high-tech firms. In contrast, we have shown that low-tech firms compete for unskilled labour and that the competition intensifies as the region size increases from commune to district to province. When we use the EG index as the measure of coagglomeration, the opposite result emerges. This is contrary to our understanding of the mobility of labour in Vietnam where restrictions on the movement of workers are in place, particularly at the province and district levels. Moreover, this result may also reflect the priorities of the Vietnamese Government as reflected in the *Vietnam Strategy for Socioeconomic Development 2001-2010* (Government of Vietnam, 2000) in which the government explicitly states that they will encourage labour intensive firms to locate in rural areas in order to boost employment in rural regions. If labour intensive firms are incentivised to locate in rural areas where there are smaller populations and therefore fewer workers, they will necessarily choose to locate in rural areas where there are no firms or few firms with similar labour needs.

We also consider an absolute measure of coagglomeration. We find that natural cost advantages are the most important agglomerative force when we think of clusters in absolute terms. We also find that technology transfers are an important positive agglomerative force and skills correlation is a significant negative agglomerative force, in agreement with our results using the relative measure.

Understanding the driving forces behind agglomeration is analytically challenging. It is also critical for governments in the formulation of industrial policy. We have shown that the definition and measurement of agglomeration are crucial to analytical outcomes when agglomeration is linked to underlying driving forces. Ellison et al. (2010) use a measure of coagglomeration that captures correlations in the relative size of two sectors, in terms of employment, across areas. While this may explain an important part of the agglomeration story, in a developing country context high-tech sectors tend to be smaller in terms of employment, and within high-tech sectors the more technologically

advanced firms tend to be even smaller again. In this setting, the EG index fails to capture the relative importance of high-tech clusters. This is certainly the case in Vietnam. It is left for future research to establish the extent to which the differences we observe reflect underlying characteristics of developed vs developing country contexts (i.e. US vs Vietnam) or are embedded in the way in which agglomeration is defined and measured.

Chapter 4

Where do new entrant firms locate?

A dyadic analysis

4.1 Introduction

Where firms choose to locate has been a question of interest to both theoretical and empirical economists for decades. In this chapter we use unique micro data on manufacturing firms in Vietnam to conduct an empirical analysis of the location choice of new entrants. We control for regional characteristics in order to determine other observable factors that drive the location decisions of firms.

Hotelling's (1929) seminal theoretical paper analysed the behaviour of two firms when choosing where to locate on a finite line. He asserted that spatial competition would lead to minimum differentiation, whereby in equilibrium both firms would choose to locate side by side at the centre of the market. Extensions of the model, however, have found that this result is very sensitive to the assumptions made, and that changing the number of firms, the type of space, or the demand function used in the analysis, can lead to equilibria where firms locate away from each other (see for example Eaton and Lipsey, 1975; Smithies, 1941; Prescott and Visscher, 1977). In the real world, however, a geographic concentration of firms is often observed, and so factors other than demand, market access and price need to be considered when analysing the location choice of firms. The presence of agglomeration economies is one explanation as to why, despite the predictions of many theoretical models, firms often locate in geographic proximity to other firms.

Empirical studies on the location choice of firms generally consider both the factors included in theoretical location choice models, and measures of agglomeration economies. A typical location choice model is based on the firm's profit maximisation problem. A firm chooses a region in order to maximise its profits and the profit from choosing region j can be expressed as a function of the observable characteristics of region j (see for example Bartik, 1985). The choice of location is then modelled in a discrete choice framework as a function of the characteristics of each region.

In this chapter we focus on which types of firms new entrants into the manufacturing sector in Vietnam choose to locate beside. Chapter 2 focused on the overall pattern of clustering in Vietnam while chapter 3 analysed why firms from different sectors locate in the same area, or colocate. This chapter uses micro data and focuses on new entrant firms. This approach will give us insight into not only between sector agglomeration but also within sector agglomeration.

We exploit the fact that we have a rich and unique data set of manufacturing firms in Vietnam. We use a fixed effects model that controls for the unobservable characteristics of the region. We can therefore isolate and focus on the observable factors that drive the firm's location decision. We generate dyads which consist of all possible unique pairings of new entrants and incumbants.³⁵ For each pairing or dyad we determine if they are located in the same cluster (same commune) or not.³⁶ This binary variable is then the unit of analysis. Using a dyadic fixed effects model for the analysis has the advantage that we do not need to explicitly measure agglomeration and so we do not need to choose a specific index. Therefore the analysis is not sensitive to the measure of agglomeration chosen.³⁷ To our knowledge this method of empirical analysis has not been previously used to investigate the clustering of firms, and this approach, together with the unique dataset, are the major innovations of this chapter.

Vietnam is a particularly interesting country to consider as over the past two decades it has opened up to foreign investors and moved from a centrally planned to a market economy. This transformation in governance has resulted in a structural shift in the economy from agriculture to manufacturing. For example, in 1991 the Vietnamese government introduced a policy to develop special economic zones. The zones have a high quality of infrastructure; zoned land is made available exclusively for industrial use and the government offers fiscal incentives to firms to locate within zones (Foreign

³⁵ See Fafchamps and Söderbom (2011) and De Weert (2004) for other empirical applications of this approach.

³⁶ A commune is the smallest administrative area in Vietnam.

³⁷ Ellison et al. (2010) find that input-output links between firms are the most important agglomerative force in the US. Howard et al. (2012) using a different measure of co-agglomeration find that technology transfers are the strongest agglomerative force in Vietnam, but also find this is dependent on the measure that is used. Finding an appropriate measure is therefore crucial to the process of identifying agglomerative forces.

Investment Agency, 2009). These incentives have led to huge growth in the number of manufacturing firms in Vietnam, but as shown in chapter 2, zoning alone cannot explain the clustering patterns. Moreover, in July 2006 a new investment law was introduced to apply equal governance to domestic and foreign-owned enterprises. Prior to this separate laws applied to domestic and foreign firms. The Investment Law introduced in 2006 was considered a significant watershed for improvement of the legal environment on investment activities and corporate governance (Foreign Investment Agency, 2009). In this chapter we focus on the location choice of new entrants in 2007; a particularly interesting year given these changes in the law.

Our results show that new high-tech firms are more likely to locate beside other high-tech firms and that they are likely to locate beside a high-tech firm that has a different level of technology. We do not find a similar relationship between low-tech firms. This suggests that technology transfers are an agglomerative force for high-tech firms. We find that firms are also more likely to cluster if they are in the same 2 or 4-digit sector, suggesting horizontal spillovers. We also find that new firms locate beside foreign-owned and state-owned firms. We disaggregate to determine if foreign and privately-owned domestic firms differ in their choice of location. We find that privately-owned firms locate beside firms in the same 4-digit sector while foreign-owned high-tech firms locate beside other high-tech firms. This suggests that different agglomerative forces are at work for the different types of firms; horizontal spillovers are important for domestic privately-owned firms while technology transfers are important for foreign firms.

The remainder of the chapter is organized as follows. Section 4.2 details the set up of the econometric model while Section 4.3 describes the data and methodology. Section 4.4 presents and discusses the results and section 4.5 concludes the chapter.

4.2 Motivation and Empirical Model

A number of empirical papers have analysed the location decision of firms. Guimaraes et al (2000) investigate the location choice of foreign firms in Portugal and find that while the distance to principal cities is a significant factor in their choice of location there is no evidence that labour costs are important. Coughlin and Segev (2000) examine the location decision of foreign-owned firms in the U.S. and find that their decision is positively correlated with the economic size, educational attainment,

existing manufacturing base, and transportation infrastructure of the county. They find that firms are less likely to locate in counties with higher average labour intensiveness and higher taxes. Head and Mayer (2004) analyse the location of Japanese firms in Europe and find that both demand (or market potential) and agglomeration economies play important roles in the decision process. Crozet et al (2004) study the determinants of location choice by foreign firms in France. They find that the nationality of the investor is important in location patterns and that firms cluster with their competitors. They also find that foreign firms are not sensitive to regional investment incentives.

Most empirical studies assume that firms are profit maximising and choose between a discrete number of regions when deciding where to locate.³⁸ Firms choose to locate in the region which will maximise their profits. This location decision is based on a number of region characteristics, agglomeration economies and policy variables. More formally if firm i locates in region j it will derive a profit π_{ij} . This profit is composed of a deterministic term and a stochastic term, as given by equation (4.1);

$$\pi_{ij} = U_{ij} + \varepsilon_{ij} \quad (4.1)$$

The deterministic part of profit is affected by a set of m regional specific variables and so, as a profit maximiser, these variables effect the location decision of the firm. The set of m variables generally include measures for demand or market size, labour costs, policy variables and agglomeration economics. Typically, it is assumed that U_{ij} is a linear combination of these regional variables;

$$U_{ij} = \beta_1 X_{ij}^1 + \beta_2 X_{ij}^2 + \dots + \beta_m X_{ij}^m \quad (4.2)$$

Many empirical papers analysing the location choice of firms model the location decision using the discrete choice conditional logit model (CLM) developed by McFadden (1974). In estimating the CLM the dependent variable is the region in which each firm chooses to locate. The analysis then determines, based on the m explanatory variables, how likely a firm is to locate in a particular region.

³⁸ An alternative approach is an occupation choice model where people are spatially distributed and choose whether to set up a firm or not (See for example Felkner and Townsend 2011).

We take an alternative approach and use a dyadic analysis to investigate the reasons why firms locate in clusters in Vietnam. We consider that incumbent firms in Vietnam form a spatial network. In choosing where to locate in the existing spatial network new entrant firms form dyads with the incumbent firms. In social network analysis a dyad is a subgraph containing two nodes and the possible links between them. In the context of this chapter the 'links' between the firms are not necessarily interactions between the two firms but rather the similarities/differences between them. It is possible to focus on the local structure of a network by studying the network at the level of subgraphs. If the level of analysis is local then these local structures give a local view of the entire network. (Wasserman and Faust, 1995).

One issue with investigations of firm location choice is that many regional characteristics can not be directly measured, therefore empirical studies have to devise proxies. For example; Crozet et al. (2004) use the real GDP of a region to proxy for market size; Siedschlag et al. (2009) use a region's GDP per capita to proxy labour costs and Bartik (1985) uses the land area of a region to proxy the number of available sites. The problem with this approach, as highlighted by Head et al. (1994), is that selecting appropriate and correct proxies for regional characteristics is extremely difficult and some important regional characteristics are likely to be omitted. Another issue is that empirical studies of location choice tend to over estimate the importance and/or the extent of strategic agglomeration. Agglomeration economies are likely to be a result of a dynamic process; firms may choose to locate close to their suppliers in order to reduce transport costs but they may also choose a location and then buy inputs from a supplier that is located close by. Both scenarios result in agglomeration economies but only in the former case is the firm engaged in strategic agglomeration. Ex-post empirical analysis of location patterns cannot distinguish between strategic and non-strategic agglomeration.

To overcome the issues associated with developing measures to proxy regional characteristics and omitted variable bias we use a district fixed effects analysis in this chapter. There are three administrative areas in Vietnam; provinces, districts and communes. Provinces are the largest in area and are controlled by provincial councils, elected by the inhabitants of each province. This structure is repeated at the district level. Zoning and planning is determined at each level of government down to the district level. The provinces and districts can therefore be very different in terms of

their laws and regulations. Firms wishing to locate in Vietnam are likely to first choose which province to locate in and then choose the district within that province before considering any other factors. By using a district fixed effects model we can control for regional policy variables, as well as the regional specific factors previously discussed such as market size and labour costs. Controlling for policy variables and regional characteristics in this manner allows us to focus on the extent to which agglomeration economies factor into the firm's location decision.

As noted above, identifying the reasons why firms cluster can be empirically difficult due to reverse causality; it is difficult to determine for example whether firms choose to locate close to their suppliers or firms choose a location and then find suppliers close by. Consequently few papers are successful in identifying if and why firms engage in strategic clustering. In this chapter we focus on newly formed dyadic relationships between new entrants and the incumbent firms they choose to locate beside. This provides some insight into the direction of causality of the agglomerative forces at work in Vietnam. Howard et al (2011) find a far higher level of agglomeration of manufacturing firms in Vietnam than would occur purely because of economic zoning, population patterns or regulatory restrictions alone. Howard et al. (2012) conduct an extensive empirical investigation into the driving forces of between sector clustering of firms and find that technology transfers are the strongest agglomerative force. Determining the types of firms new entrants choose to locate beside provides important insights into the mechanisms of these technology transfers. Knowing the mechanisms through which these transfers are occurring is critical for governments to formulate policies that foster clusters and facilitate technology spillovers.

As we conduct a fixed effects analysis and since communes are geographically small in area, once two firms are located in the same district, we can consider them locating in the same commune analogous to them locating in the same cluster. Comparing the dyads where both firms are located in the same commune to dyads where the two firms are located in different communes in the same district allows us to identify the types of firms that cluster, controlling for the district characteristics and the planning and regulatory framework. An additional benefit of using this type of analysis as noted in the introduction is that we do not need to measure the extent of agglomeration and so we do not need to choose an agglomeration index.

In line with other empirical investigations into the location choice of firms, we consider location choice in the profit maximisation framework. However, we use a new empirical approach and conduct a dyadic fixed effects analysis. In our framework, the new entrant firm chooses which incumbent firms to locate beside in order to maximise profits, given that the entrant has already chosen in which district to locate in order to maximise profits. Agglomeration economies reduce costs and therefore increase profits. Therefore new entrants locate beside incumbents or in clusters where the potential for agglomeration economies is highest. This location decision creates the dyads which are our unit of analysis. The new entrant chooses to locate in the same commune as an incumbent firm, or to locate in a different commune in the same district as the incumbent.

We therefore estimate the district fixed effects regression given by equation (4.3);

$$y = X_{ij}\beta + \alpha + \varepsilon \quad (4.3)$$

where the dependent variable y takes the value 1 if the two firms in the dyad are located in the same commune, and 0 if they are located in the same district but different communes. X_{ij} contains relational and dummy variables pertaining to the two firms (i and j) in the dyad, α is the district fixed effects, the unknown intercept for each district, and ε is the error term.

4.3 Data

The data are taken from the Vietnamese Enterprise Survey for 2006 and 2007 provided by the General Statistics Office of Vietnam (GSO, 2010). The dataset includes all registered manufacturing enterprises at the end of each year with more than 30 employees. The data includes information on the main economic activity of the firm, the type of ownership (state-owned, privately-owned or foreign-owned), the number of employees, the income of employees, 4-digit sector to which the firm belongs, and the commune (the smallest administrative area) in which the firm is located.

We generate dyads which consist of all possible unique combinations of two firms in the data set. These dyads are the units of analysis. Dyadic analysis has been used in a number of papers with economic applications. Most recently Fafchamps and Söderbom (2012) use dyad analysis to explore network proximity and business practices in

Ethiopia. They use survey data and investigate if firms who do business with each other, or have a common supplier, or a common client, tend to be more similar than firms without these links. De Weert (2004) uses dyadic analysis to look at risk sharing in a small village in Tanzania. He uses survey data and calculates the strength of the link between all sets of two houses (dyads) in the village to present stylized facts on the formation of the risk sharing network.

For each dyad we construct a number of relational and dummy variables. Relational variables are the difference between the two firms in the dyad for a numerical variable. The dummy variables are used for categorical variables, for example to indicate if one of the firms in the dyad is foreign-owned. For each of these dyads we determine the degree of spatial connection between the two firms. We assign a spatial connection value of 1 if the firms are in the same commune and 0 if they are in the same district but not the same commune. We drop the dyads that contain firms that are located in different districts and so focus on within-district clustering of firms. Comparing the dyads where both firms are located in the same commune to dyads where the two firms are located in different communes in the same district allows us to identify the types of firms that cluster, controlling for district specific characteristics.

We use the enterprise survey data from 2006 to identify new entrants in 2007. We drop any dyads that contain two new firms or two old firms. Once a firm has entered the market and chosen a commune in which to locate, it has no control over what firms subsequently chose to locate in that commune. Therefore in order to say something about the strategic clustering of firms we focus on the dyads containing one new firm/entrant.

The final dataset contains 204,333 dyads. Each contains two firms, one of which entered the market in 2007. Of these new firm dyads, in 35,191 cases (17 percent) the two firms are in the same commune. The remaining 169,142 dyads consist of two firms that are located in the same district, but different communes. Zoning and planning in Vietnam is determined at various governmental levels down to the district level. Therefore differences in district regulations which are unobserved in the data could be correlated with the explanatory variables included in the analysis. In order to allow for this and to ensure that district differences are not driving our results we use a district fixed effects model. Additionally we report bootstrapped and clustered (at district level)

standard errors. When using dyadic data observations are typically not independent. In the case of our data firm i may appear in many dyads and so the residual from the dyad including firms i and j is likely to be correlated with the residual from the dyad containing firms i and k . Bootstrapping obtains consistent standard errors and corrects for the cross-observation correlation in the error terms involving the same enterprises (Fafchamps and Söderbom, 2012).³⁹

As previously discussed, using a district fixed effects model controls for many of the factors that may influence a firm's decision where to locate. Given that a firm has decided to locate in a particular district, a firm need only choose in which commune in that district to locate. The main, and arguably the only factor in this location choice as communes in the same district are so similar, is the type of firms already located in that commune. We construct a number of variables to measure how similar (or different) the new entrants and incumbent firms in the dyads are in terms of the types of production technologies they use, their ownership structure, their size, and what they produce. In chapter 3 we find technology transfers between firms to be the strongest agglomerative force at work in Vietnam, and so we pay particular attention to the technological similarities between firms in our analysis.

Table 4.1 summarises the dummy variables used in the regression analysis. The variable *high-tech* takes the value 1 if both firms in the dyad are in sectors with typically high levels of production technologies and zero otherwise.⁴⁰ The variable *low-tech* takes the value 1 if both firms in the dyad are in sectors with typically low levels of production technology and zero otherwise. We found in chapter 3 that technology transfers were particularly important for the agglomeration of firms in high-tech sectors, so we would expect to find a positive and significant result for the high-tech dummy variable. Table 4.1 illustrates that 29 percent of dyads consist of high-tech

³⁹ Note that it is not possible to cluster standard errors at the firm level as the unit of analysis is the dyad, which contains two firms.

⁴⁰ 4-digit sectors 1500 to 2233 and sectors 3600 to 3699 (inclusive) we define as low-technology manufacturing sectors. All other sectors are defined as high-technology. See appendix for a full list of manufacturing sector codes. This closely follows the OECD's classification of industrial sectors by level of technology however they classify industries as low-tech, middle low-tech, middle high-tech and high-tech (Hatzichronoglou, 1997). As only about 20 percent of firms in our data are middle high-tech and high-tech we consider only the low-tech firms in the OECD classification as low-tech and define all other sectors as (relatively) high-tech.

firm pairings while 24 percent of dyads consist of low-tech firm pairings. This means that 47 percent of dyads consist of mixed pairings, where one firm is high-tech and the other is low-tech.

Table 4.1: Dummy Variables

Dummy Variable	1	0
High-tech	59,219 (29%)	145,114 (71%)
Low-tech	49,632 (24%)	154,701 (76%)
Same 4-digit	16,300 (8%)	188,033 (92%)
Same 2-digit	32,729 (16%)	171,604 (84%)
State Dummy	87,982 (43%)	116,351 (57%)
Foreign Dummy	74,447 (36%)	129,886 (64%)
Large Firm 90th	31,397 (15%)	172,936 (85%)

A dummy variable for two firms in the same 4-digit sector is also included along with an indicator for dyads where both firms are in the same 2-digit sector. In 8 percent of dyads the firms are in the same 4-digit sector and in 16 percent of dyads the firms are in the same 2-digit sector.

The state and foreign dummy variables indicate, respectively, whether the incumbent firm in the dyad is state-owned or foreign-owned; 43 percent of dyads contain a state-owned incumbent firm while 36 percent of dyads contain a foreign-owned firm. This variable is included to investigate if ownership structure plays a role in the location decision of the firm.⁴¹

An indicator for large firms captures if the incumbent firm in the dyad is very large; it takes the value one if the firm is defined as large, and zero otherwise. The visual analysis of the clustering pattern by firm size in chapter 2 indicated that large firms appear to be located at the centre of clusters. We include this large firm dummy variable to investigate if this is in fact the case. We define a firm to be large if they are in the 90th percentile of all manufacturing firms in the dataset in terms of number of employees. As illustrated in Table 4.1, 15 percent of dyads contain a large incumbent firm.

⁴¹ A priori we expect ownership structure will play a role, for reasons discussed in Section 2.5.2

Table 4.2 summarises the relational variables used in the regression analysis. The relational variables are the difference in values between the two firms in each dyad. We calculate the capital-labour ratio for each firm to measure the firm's capital intensity. For each dyad we take the absolute value of the difference between the capital labour ratios for each firm in the dyad. We then take the log of this for inclusion in the analysis. This variable measures the difference in capital intensity for the two firms in the dyad, which we consider a proxy for the difference in the types of technology used by the two firms.

Table 4.2: Relational Variables

Relational Variable	Mean	Standard Dev	Min	Max
Log KL Ratio	5.04	1.51	0.0005	10.59
Log No. Comp per Employee	0.09	0.10	0	0.86

Second, we calculate the absolute difference in the number of computers per employee between the two firms and take the log. This is included as an additional proxy for the difference in the types of technology used by the two firms.

In 2007, 9,436 new manufacturing firms are observed in our data. Of these new entrants 5,577 or 59 percent are entirely or partly owned by the state. A further 632 firms or 7 percent, are foreign-owned firms, and 3,227, or 34 percent, are privately-owned domestic firms. Table 4.3 illustrates the differences between state, foreign and private firms in terms of the mean and spread in capital intensity and number of computers per employee. The mean number of computers per employee is highest for foreign-owned entrant firms and lowest for domestic privately-owned enterprises. Foreign-owned entrant firms are also the most capital intensive while privately-owned domestic entrants are the least capital intensive.

Table 4.3: New Entrants

	State-owned	Foreign-owned	Private Domestic
Variable	<i>Mean (Std Dev)</i>	<i>Mean (Std Dev)</i>	<i>Mean (Std Dev)</i>
KL Ratio	268 (1322)	1011 (4110)	240 (740)
Comp per Employee	0.14 (0.17)	0.22 (0.25)	0.13 (0.16)

This difference in production technologies between the three different types of firms is unsurprising when we consider the sectors these firms are entering into; 52 percent of foreign entrant firms are entering into sectors with typically high levels of production technology compared with 41 percent of state-owned and only 40 percent of privately-owned domestic firms.

4.4 Results

Table 4.4 presents the results of the regression model given in equation (4.3). Bootstrapped and clustered (at district level) standard errors are presented in parenthesis. The first column shows the results of the baseline regression when the dummy variables high-tech, low-tech, same 4-digit and same 2-digit sector are included. The results show that the two firms in the dyad are most likely to be located in the same commune if they are in the same 4-digit sector. They are also more likely to be located in the same commune if they are in the same 2-digit sector. This suggests that new entrant firms are locating beside incumbent firms in the same 2 or 4-digit sectors. This points to the presence of horizontal spillovers between firms. Howard et al. (2012) explore and identify the driving forces of between sector agglomeration in Vietnam. These results indicate that there are also important within-sector agglomerative forces at work.

The results also show that the two firms in the dyad are more likely to be located in the same commune if both firms are high-tech, compared to the base group of a mixed pairing (where one firm is low-tech and the other is high-tech). Therefore a high-tech firm entering the market is more likely to locate in the same commune as another high-tech firm. This suggests that technology transfers may be an agglomerative force for high-tech firms; one reason why they are more likely to locate beside other high-tech firms is because they are then more likely to experience technology spillovers. We find a negative and significant effect for low-tech firms, indicating that low-tech firms entering the market, who are not in the same 2 or 4-digit sector, choose to locate away from other low-tech firms. Controlling for sector overlap of the two firms in the dyad, they are less likely to be in the same commune if they are both low-tech firms. This suggests that firms in low-tech sectors are less likely to experience horizontal technology transfers and so choose to locate away from other firms in low-tech sectors in order to reduce competition. Firms in the same sector are subject to two competing

forces; if they locate close to other firms in their sector they are more likely to experience technology transfers, however locating close to another firm in the same sector increases direct competition and could consequently reduce their market share. The results in this chapter suggest that for low-tech firms within sector competition dominates, while for high-tech firms, technology transfers dominate.

Table 4.4: Fixed effects dyadic regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High-tech	0.017** (0.008)	0.016* (0.009)	0.015 (0.010)	0.015* (0.008)	0.018** (0.008)	0.020*** (0.007)	0.018** (0.008)
Low-tech	-0.010*** (0.004)	-0.006** (0.003)	-0.008** (0.004)	-0.005** (0.003)	-0.011*** (0.003)	-0.013*** (0.003)	-0.008** (0.004)
Same 4-digit	0.047*** (0.012)	0.043*** (0.011)	0.052*** (0.016)	0.046*** (0.013)	0.046*** (0.014)	0.045*** (0.012)	0.043*** (0.014)
Same 2-digit	0.021*** (0.007)	0.021** (0.008)	0.020*** (0.007)	0.020** (0.009)	0.020** (0.008)	0.021*** (0.008)	0.020** (0.008)
Ln Comp per emp		0.120* (0.072)		0.093 (0.066)			0.078 (0.053)
Ln KL Ratio			0.007** (0.003)	0.005*** (0.001)			0.005*** (0.002)
State dummy					0.010** (0.005)	0.010** (0.004)	0.010** (0.005)
Foreign dummy					0.075*** (0.018)	0.074*** (0.019)	0.074*** (0.016)
Large Firm						0.018** (0.008)	0.020*** (0.007)
R-sqd	0.0033	0.012	0.0049	0.0045	0.019	0.019	0.021
Obs	204,333	194,639	204,172	194,483	204,333	204,333	194,483

Bootstrapped and clustered (at district level) standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The second column of Table 4.4 shows the regression results when the first measure of the difference in the production technologies between the two firms, the Log of Computers per Employee, is included in the model. This is the log of the absolute difference between the number of computers per employee for the two firms in the dyad. The results show that the larger the difference in the number of computers per employee, the more likely the two firms are to locate in the same commune. Therefore the new firm entering the market is more likely to locate close to a firm that has a different production technology. As technology transfers are more likely to take place from a firm with high levels of technology to a firm with a lower levels of technology this result also suggests that technology transfers are an agglomerative force.

The results of the third specification of the model are shown in column 3 of Table 4.4. This model includes the second proxy for the difference in levels of production

technology between the two firms, the Capital Labour (KL). The KL ratio is the log of the absolute difference between the KL ratios of the two firms in the dyad and is a measure of the difference in capital intensity between the two firms. The positive coefficient implies that the bigger the difference in capital intensity between the two firms, the more likely they are to be located in the same commune. This implies that firms entering the market are more likely to locate beside firms that are either more or less capital intensive. This result is in agreement with the results presented in column 2 using computers per employee as the proxy for the level of production technology. It once again points to technology transfers as an agglomerative force as transfers are more likely from capital intensive to labour intensive firms.

The fourth column of Table 4.4 shows the results of the baseline model when both measures of the difference in the levels of production technology between the firms in the dyad are included. The number of computers per employee is no longer statistically significant. This suggests that the results for the difference in the number of computers per employee are driven by differences in capital intensity between the firms in the dyad.

Column 5 of Table 4.4 shows the results of the regression analysis when the state and foreign-owned dummy variables are added to the baseline model. The results show that if the incumbent firm in the dyad is state-owned or foreign-owned the firms are more likely to be located in the same commune, compared with the base group of domestically owned incumbent firms. This implies that new firms entering the market are more likely to locate beside foreign-owned or state-owned firms. Foreign-owned firms and state-owned firms tend to be more technologically advanced than domestic firms in developing countries so new firms may choose to locate close to foreign-owned firms in order to maximise their chances of experiencing technology spillovers.

In column 6 the dummy variable for a large incumbent firm is added to the model specification from column 5. We define an incumbent firm as large if it is in the 90th percentile in terms of number of employees. The results show that if the incumbent firm in the dyad is large then the two firms are more likely to be in the same commune. Therefore new firms are more likely to locate beside large firms.

The last column of Table 4.4 includes all variables in the regression analysis. The results from the previous specifications of the model are robust to the inclusion of all

variables. New firms are more likely to locate beside firms that are in the same 4 or 2-digit sector, that have different levels of production technology, that are large, or are foreign-owned. High-tech entrants are more likely to locate beside other high-tech firms while low-tech firms are more likely to locate away from low-tech incumbent firms that are not in the same 2 or 4-digit sector.

As illustrated in section 4.3, the characteristics of new entrant firms differ substantially depending on whether they are state-owned, domestic privately-owned or foreign-owned firms. Therefore it is possible that the type of firms entrants choose to locate beside is influenced by their ownership structure. Additionally, state-owned enterprises may not necessarily choose their location based on profit maximisation objectives. The location choice of state-owned firms may reflect the priorities of the Vietnamese Government. For example, the Government of Vietnam (2000) explicitly states that they will encourage labour intensive firms to locate in rural areas in order to boost employment in rural regions. The location choice of state-owned firms therefore may not be influenced by the characteristics of incumbent firms. To ensure that state-owned firms are not driving the results of our analysis, and to investigate if the different ownership structures influence location choice, we repeat the analysis on dyads containing foreign and domestically owned entrants only. The results for foreign entrants are presented in Table 4.5

Table 4.5: Fixed effects dyadic regression Foreign entrants only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High-tech	0.035*** (0.014)	0.038*** (0.014)	0.033*** (0.013)	0.036*** (0.013)	0.038*** (0.012)	0.038*** (0.013)	0.040*** (0.012)
Low-tech	-0.016* (0.009)	-0.016 (0.011)	-0.013 (0.009)	-0.014 (0.009)	-0.012* (0.007)	-0.012 (0.009)	-0.010 (0.007)
Same 4-digit	0.029** (0.013)	0.027* (0.016)	0.033** (0.015)	0.032** (0.017)	0.022 (0.015)	0.021 (0.015)	0.024* (0.014)
Same 2-digit	0.036* (0.020)	0.035* (0.022)	0.035 (0.022)	0.035* (0.019)	0.019* (0.010)	0.019* (0.012)	0.019* (0.010)
Ln Comp per emp		-0.028 (0.052)		-0.077 (0.059)			-0.093** (0.044)
Ln KL Ratio			0.007*** (0.002)	0.009*** (0.003)			0.008** (0.003)
State dummy					0.006 (0.010)	0.006 (0.012)	0.003 (0.012)
Foreign dummy					0.161*** (0.053)	0.161*** (0.063)	0.159*** (0.058)
Large Firm						0.002 (0.012)	0.003 (0.012)
R-sqd	0.004	0.004	0.006	0.006	0.048	0.048	0.048
Obs	67,820	66,825	67,820	66,825	67,820	67,820	66,825

Bootstrapped and clustered (at district level) standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

There are three notable differences between these results and those shown in Table 4.4 when all dyads are included. First, in all specifications of the model the coefficient on the high-tech variable is approximately twice that found when using the full data set. This implies that high-tech foreign-owned entrant firms are twice as likely to locate beside incumbent high-tech firms as the average entrant. Second, column 7 of Table 4.5 shows that when all variables are included in the model, the computers per employee variable is negative and significant. This implies that foreign entrants are more likely to locate in the same commune as incumbents that have similar levels of production technology. Lastly, the coefficient on the foreign-owned dummy is over twice the magnitude of that found in the analysis using all dyads, in all three specifications in which it is included. This implies that a foreign entrant is over twice as likely as an average entrant firm to locate beside a foreign-owned incumbent.

Table 4.6 shows the regression results when dyads containing an incumbent firm and only domestic privately-owned entrants are included in the analysis. The only variable that is significant in all specifications of the model is the same 4-digit sector dummy variable. The coefficient is over twice the magnitude of that found when using all dyads in the analysis. This suggests that domestically owned firms are over twice as likely as the average entrant firm to locate beside a firm in the same 4-digit sector. Column 7 of

Table 4.6 illustrates that when all variables are included in the model the high-tech dummy variable is also significant and the coefficient is slightly larger in magnitude than that found when all dyads are included. This implies that domestically owned high-tech firms are slightly more likely than the average entrant to locate beside high-tech incumbent firms.

Table 4.6: Fixed effects dyadic regression Private domestic entrants only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High-tech	0.024 (0.015)	0.023* (0.014)	0.025 (0.016)	0.024* (0.013)	0.024 (0.018)	0.025 (0.016)	0.025** (0.012)
Low-tech	0.003 (0.008)	0.005 (0.008)	0.002 (0.007)	0.005 (0.008)	0.002 (0.009)	0.002 (0.008)	0.004 (0.009)
Same 4-digit	0.125*** (0.028)	0.102*** (0.029)	0.129*** (0.023)	0.106*** (0.023)	0.125*** (0.027)	0.124*** (0.028)	0.106*** (0.027)
Same 2-digit	-0.005 (0.010)	-0.004 (0.010)	-0.009 (0.010)	-0.008 (0.010)	-0.005 (0.010)	-0.005 (0.009)	-0.007 (0.008)
Ln Comp per emp		0.035 (0.044)		0.026 (0.050)			0.024 (0.042)
Ln KL Ratio			0.002 (0.002)	0.002 (0.004)			0.002 (0.003)
State dummy					-0.034 (0.010)	-0.004 (0.008)	-0.001 (0.007)
Foreign dummy					0.006 (0.019)	0.005 (0.015)	0.008 (0.018)
Large Firm						0.007 (0.014)	0.006 (0.017)
R-sqd	0.019	0.012	0.019	0.012	0.020	0.020	0.013
Obs	38,931	34,575	38,826	34,473	38,931	38,931	34,473

Bootstrapped and clustered (at district level) standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

These results suggest that technology spillovers between high-tech firms, particularly those with similar levels of production technology, are an important agglomerative force, and that foreign firms cluster together. They also indicate that for private domestic firms horizontal spillovers are the most important agglomerative force.

To further investigate the technology transfer mechanism we construct a new dummy variable to indicate if the incumbent firm has a higher level of technology than the new entrant. The variable takes the value 1 if the incumbent is more capital intensive (i.e. has a higher KL ratio) than the the new entrant firm and zero otherwise.⁴² Table 4.7 presents the results of the analysis using this indicator of the incumbent's technology for all new entrants, foreign-owned entrants, and privately-owned domestic firms separately. In columns 1, 2 and 3 the dummy variable *Incumbent Higher-Tech* is

⁴² We also conduct the analysis using the number of computers per employee as the measure of technology and find similar results.

included in the analysis. In column 2 the dummy variable *High-tech*Incumbent Higher* is also included. This indicates that both the new entrant and the incumbent are high-tech firms but that the incumbent has a higher level of technology. In column 3 the dummy variable *Low-tech*Incumbent Higher* is also included. This indicates that both the new entrant and the incumbent are low-tech firms but that the incumbent has a higher level of technology.

The results when all new entrant firms are considered suggest that new entrants locate away from incumbents with higher levels of technology. However when only foreign entrants are considered this is no longer the case; the coefficient on the *Incumbent Higher-Tech* is positive, although not significant. Column 2 of the foreign-owned new entrant results suggests that when the new entrant and the incumbent are both high-tech firms but the incumbent has higher levels of technology, the new entrant is more likely to locate in the same commune as the incumbent. This implies that possible technology transfers from higher-tech incumbents may be a motivating factor in the location decision of firms in high-tech sectors. Column 3 reveals that there is no significant effect when both firms are in low-tech sectors but the incumbent has higher levels of technology.

The results when only privately-owned domestic new entrants are included in the analysis suggest that an incumbent with higher levels of technology has no significant impact on the location choice of the firm. There is weak evidence from column 2 that high-tech new entrants are more likely to locate beside a high-tech incumbent with higher levels of technology, but the result is not statistically significant. There is also weak evidence that low-tech new entrants locate away from low-tech incumbents with higher levels of technology, although again this result is not significant. One possible explanation for this is that low-tech firms do not expect to experience technology transfers. Competition may motivate them to locate away from other low-tech firms with higher levels of technology, as these firms may be more efficient and therefore a greater competitive threat.

Table 4.7: Fixed effects dyadic regression including *Incumbent Higher-Tech Dummy*

	(1)	(2)	(3)
		All New Entrants	
High-tech	0.016* (0.008)	0.024*** (0.006)	0.015** (0.007)
Low-tech	-0.009** (0.004)	-0.009*** (0.003)	-0.018*** (0.007)
Same 4-digit sector	0.046*** (0.014)	0.046*** (0.013)	0.046*** (0.014)
Same 2-digit sector	0.020*** (0.007)	0.020*** (0.007)	0.020*** (0.007)
Incumbent Higher-Tech	-0.021** (0.009)	-0.017** (0.009)	-0.025** (0.011)
High-tech*Incumbent Higher		-0.015 (0.014)	
Low-tech*Incumbent Higher			0.015* (0.009)
R-squared	0.004	0.004	0.004
Observations	203,317	203,317	203,317
		Foreign-owned Entrants	
High-tech	0.035*** (0.013)	0.024** (0.010)	0.035*** (0.013)
Low-tech	-0.015* (0.010)	-0.016* (0.010)	-0.015 (0.010)
Same 4-digit sector	0.029** (0.015)	0.028** (0.014)	0.029** (0.014)
Same 2-digit sector	0.035* (0.020)	0.034* (0.020)	0.034* (0.018)
Incumbent Higher-Tech	0.008 (0.031)	-0.001 (0.033)	0.008 (0.033)
High-tech*Incumbent Higher		0.029* (0.015)	
Low-tech*Incumbent Higher			-0.003 (0.014)
R-squared	0.004	0.005	0.004
Observations	67,284	67,284	67,284
		Privately-owned Entrants	
High-tech	0.024 (0.018)	0.022 (0.0146)	0.024* (0.013)
Low-tech	0.003 (0.008)	0.003 (0.009)	0.004 (0.013)
Same 4-digit sector	0.124*** (0.025)	0.124*** (0.027)	0.124*** (0.027)
Same 2-digit sector	-0.006 (0.011)	-0.006 (0.010)	-0.006 (0.010)
Incumbent Higher-Tech	-0.006 (0.010)	-0.007 (0.011)	-0.006 (0.008)
High-tech*Incumbent Higher		0.004 (0.020)	
Low-tech*Incumbent Higher			-0.002 (0.012)
R-squared	0.02	0.02	0.02
Observations	38,855	38,855	38,855

Bootstrapped and clustered (at district level) standard errors are presented in parentheses. ***
p<0.01, ** p<0.05, * p<0.1

4.5 Conclusion

There are many factors that influence a firm's choice of where to locate, and many of these factors are unobservable or hard to measure empirically. In this chapter we conduct a fixed effects dyadic analysis which controls for regional characteristics and allows us to focus on agglomeration economies in the firm's choice of location. Empirical studies that attempt to measure agglomeration economies often over estimate the extent of strategic agglomeration. In order for governments to formulate effective industrial policies that foster and encourage innovation clusters they need to know to what extent firms choose to locate in a cluster because of the characteristics of the incumbent firms in that cluster. We focus on new entrants into the manufacturing sector in Vietnam and determine the type of firms they choose to locate beside. This provides a useful insight into the agglomerative forces at work in the manufacturing sector of this dynamic economy.

We find that new entrant firms are more likely to locate beside incumbent firms in the same 2 or 4-digit sectors suggesting the presence of horizontal spillovers. This is particularly the case for private domestic entrants. We also find that new entrants choose to locate close to foreign or state-owned incumbents. If the entrant is foreign-owned they are twice as likely to locate beside a foreign incumbent. If the new entrant is in a high-tech sector they are likely to locate beside a high-tech incumbent, suggesting technology transfers are an agglomerative force.

This result is even stronger for foreign-owned entrants. Finally, new entrants are more likely to locate beside incumbents that have a different level of production technology. When we disaggregate into ownership types, it appears that foreign-owned new entrants are more likely to locate beside an incumbent with higher levels of technology, particularly if both the incumbent and new entrant are in high-tech sectors. This suggests that technology transfers are an agglomerative force for foreign-owned firms as transfers are most likely to occur from a firm with a high level of technology to a firm with a lower level of technology. No significant relationship is found however between domestic privately-owned new entrants and incumbents with higher levels of technology.

Chapter 3 found evidence of between sector agglomeration in Vietnam, and that technology transfers were the most important force; the findings in this chapter support

this. The results provide evidence that technology transfers are a significant factor in the location choice of firms. Additionally, the results presented in this chapter suggest that there is also within sector agglomeration in Vietnam.

These results have important industrial policy implications. The evidence suggests that technology transfers are most likely to occur from high-tech foreign firms to high-tech foreign entrants. If policy makers wish to promote transfers to domestic firms they need not only to encourage and incentivise high-tech foreign firms to locate in clusters, but must also encourage domestic firms to do the same. Many empirical studies on the impact of foreign direct investment and technology transfers on domestic firms find significant improvements in the productivity of local firms. Future research will analyse if this is also the case in Vietnam.

An alternative interpretation of the results in this chapter is that physical proximity is more important for spillovers in high-tech sectors. Technology transfers may be occurring in low-tech sectors, but between firms that are not located in the same area. This suggests that formal or contracted transfers may be more important to firms in low-tech sectors, as these types of transfers are not necessarily correlated with physical proximity. In contrast, informal transfers or spillovers are usually associated with the geographic concentration of firms. Discovering the mechanisms through which these transfers occur is again crucial for governments in the formulation of industrial and spatial strategies. For example, if it is the case that geographic proximity is not necessary for technology transfers between low-tech firms, then encouraging low-tech firms to locate in areas with high unemployment may be preferable to encouraging them to locate in clusters. Whether spillovers occur primarily through formal or informal channels, and whether these channels are different for low and high-tech firms, is another issue that requires further research and investigation.

Chapter 5

Conclusion

The theoretical literature on the location decision and the geographic clustering of firms is extensive. Empirical evidence is limited however, particularly in a developing country context. Two issues have emerged in the empirical literature on the clustering of firms; the first is how should agglomeration be defined and measured, and the second is the identification of the agglomerative forces at work in a country, and determining their relative strength and magnitude. This thesis contributes to both strands of the literature.

Chapter 2 presents a new empirical approach to the investigation of agglomeration patterns and uses tools from network analysis to measure the extent of the clustering of firms in Vietnam. Chapter 3 proposes a new measure for between sector agglomeration and conducts an extensive analysis to test the relative importance and magnitude of Marshall's three theoretical agglomerative forces in the manufacturing industry in Vietnam. Chapter 4 focuses on new entrant firms in order to determine if, and why, firms engage in strategic clustering. The dyadic fixed effects analysis used in this chapter represents a new empirical approach that controls for unobserved regional characteristics and has the added advantage that no explicit measurement of agglomeration is required in the analysis.

The results in chapter 2 show that the extent of clustering of manufacturing firms in Vietnam is over and above that which can be attributed purely to economic zoning, population patterns or the regulatory framework. Therefore other agglomerative forces must be at work.

Chapter 3 focuses on between sector agglomeration and finds that technology transfers are the most important agglomerative force, particularly for firms in high-tech sectors. There is also evidence that high-tech firms pool workers while low-tech firms compete for workers. These results are consistent with Almazan et al.'s (2007) model which predicts more clustering in industries that spend large amounts on research and development (R&D). As R&D is a risky investment, clustering with other high R&D industries allows for labour pooling and successful innovators can hire skilled workers

from unsuccessful firms. Firms in high-tech sectors can be considered high R&D firms and so we expect them to cluster and pool workers. No evidence is found in our analysis that input-output links between sectors are an agglomerative force. This suggests that firms do not locate near domestic suppliers or customers, in other sectors, and therefore that domestic transport costs are not a factor in the location choice of the firm. This result is not surprising in the context of declining transport costs. Glaeser and Kohlase (2004) argue that today it is essentially free to move goods but expensive to move people. Input-output linkages may still be a force for *within* sector agglomeration however, and exporting firms may consider the transport costs associated with exporting when deciding where to locate. If we consider that it is more costly to move people than goods, and also that there are still some restrictions on the migration of workers in Vietnam, the result that firms in low-tech sectors compete for, rather than pool workers is also unsurprising.

The results in chapter 4 suggest that in addition to between-sector agglomeration in Vietnam, within-sector agglomeration also occurs. Identifying the reasons why firms cluster can be difficult due to reverse causality. The chapter focuses on the location choice of new entrant firms which isolates strategic clustering. The results show that new entrants are more likely to locate beside firms in the same 2 or 4-digit sectors, suggesting horizontal spillovers. The results also suggest that technology transfers are an agglomerative force, although spillovers appear most likely for foreign-owned firms. This raises some important issues for industrial policy makers.

As discussed in the introduction of this thesis, the formulation of industrial policies that foster clusters, promote growth, and reduce spatial inequality is critical, particularly in developing countries. Attracting foreign investment is often seen as crucial, but developing countries need to be cautious in attracting FDI. Rodríguez –Clare (1996) argues that the impact on developing countries of FDI depends on the type of linkages foreign firms generate. He expects multinationals to generate more linkages when they come from regions that are further away and more different in terms of their cultural, social, and legal structures. The results found in chapter 4 suggest that foreign firms are more likely to experience technology transfers than domestic firms and so this asymmetry needs to be considered, and if possible corrected with appropriate policies.

Policy makers should also be concerned with the effectiveness of policies aimed at promoting clusters. Crozet et al. (2004) find little evidence that European or national regional policies have any impact on the location decision of foreign firms in France. Deveraux et al. (2007) find that fiscal incentives are more effective when offered in areas that also have localisation benefits. Governments must consider the agglomeration economies already present in an area or industry when devising fiscal incentives to encourage clustering. Therefore the first step in the formulation of industrial policies should be to identify significant clusters and the driving forces behind them. This can be done by using the empirical approach suggested in this thesis.

Although this thesis has conducted a thorough empirical investigation into the driving forces behind the clustering of firms in Vietnam, the effect that locating within a cluster has on the firm itself has not been considered. There is empirical evidence from a number of countries that firms in clusters are more productive. Ciccone and Hall (1996) and Henderson (1986), using US data and US and Brazilian data respectively, find that productivity increases in areas with a concentration of similar activities. Siba et al. (2012) find that agglomerating firms in Ethiopia will have higher productivity, but only if they produce similar products to other firms in the cluster. Firms in clusters may also be more likely to export than firms in isolation. Koenig (2009) finds evidence of export agglomeration economies in France, and that local exporters positively influence the probability of a firm starting to export. Future research will consider whether manufacturing firms located in clusters in Vietnam have higher productivity than firms outside clusters and if export agglomeration economies are present.

Bibliography

Almazan, Andres Adolfo De Motta, and Sheridan Titman. 2007. 'Firm Location and the Creation and Utilization of Human Capital.' *Review of Economic Studies*. 74(4): 1305-1327.

Audretsch, David B. and Maryann P. Feldman. 1996. 'R&D Spillovers and the Geography of Innovation and Production.' *American Economic Review*. 86(3): 630-40.

Baptista, Rui. 2000. 'Do innovations diffuse faster within geographical clusters?' *International Journal of Industrial Organization*. 18(3): 515-535.

Bartik, Timothy J. 1985. 'Business location decisions in the United States: Estimates of the effects of unionization, taxes and other characteristics of states.' *Journal of Business and Economic Statistics*. 3: 14-22.

Bigsten, Arne, Mulu Gebreeyesus, Eyerusalem Siba and Måns Söderbom. 2011. *The effects of agglomeration and competition on prices and productivity: Evidence for Ethiopia's manufacturing sector*. Work in Progress. University of Gothenburg.

Carlton, Dennis W. 1983. 'The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables.' *Review of Economics and Statistics*. 65: 440-449.

Ciccone, Antonio and Robert E. Hall. 1996. 'Productivity and the Density of Economic Activity.' *American Economic Review*. 86(1): 54-70.

Collier, Paul and John Page. 2009. *Industrial Development Report 2009: Breaking in and Moving Up – New Industrial Challenges for the Bottom Billion and the Middle Income Countries*. Vienna: UNIDO.

Combes, Pierre-Philippe and Gilles Duranton. 2006. 'Labour Pooling, Labour Poaching, and Spatial Clustering.' *Regional Science and Urban Economics*. 36(1): 1-28.

Coughlin, Cletus C and Eran Segev. 2000. 'Location Determinants of New Foreign-Owned Manufacturing Plants.' *Journal of Regional Science*. 40(2): 323-351.

- Crozet, Matthieu, Thierry Mayer and Jean-Louis Mucchielli. 2004. 'How do firms agglomerate? A study of FDI in France.' *Regional Science and Urban Economics* 34(1): 27-54.
- Deichmann Uwe, Somik V. Lall, Stephen J. Redding and Anthony J. Venables. 2008. 'Industrial Location in Developing Countries.' *World Bank Research Observer*. 23(2): 219-246.
- Devereux, Michael P, Rachel Griffith and Helen Simpson. 2004. 'The geographic distribution of production activity in the UK.' *Regional Science and Urban Economics*. 34: 533-564.
- Devereux Michael P., Rachel Griffith and Helen Simpson. 2007. 'Firm location decisions, regional grants and agglomeration externalities.' *Journal of Public Economics*. 91(3-4): 413-435.
- De Weerd, Joachim. 2004. 'Risk-Sharing and Endogenous Network Formation.' in Dercon, Stefan (ed.) *Insurance against Poverty*. OUP Catalogue: Oxford University Press.
- Duranton, Gilles and Henry G. Overman. 2005. 'Testing for Localization Using Micro-Geographic Data.' *Review of Economic Studies*. 72(4): 1077-1106.
- Eaton, B. Curtis and Richard G. Lipsey. 1975. 'The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition.' *Review of Economic Studies*. 42: 27-49.
- Ellison, Glenn, and Edward L. Glaeser. 1997. 'Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.' *Journal of Political Economy*. 105(5): 889-927.
- Ellison, Glenn, and Edward L. Glaeser. 1999. 'The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?' *American Economic Review*. 89(2): 311-316.
- Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. 'What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.' *American Economic Review*. 100(3): 1195-1213.

- Erdoes, Paul, and Alfréd Rényi. 1960. 'On the Evolution of Random Graphs.' *Publications of the Institute of Mathematics, Hungarian Academy of Science*. 5: 17-60.
- Fafchamps, Marcel and Måns Söderbom. 2012. 'Network Proximity and Business Practices in African Manufacturing'. *World Bank Economic Review*, forthcoming.
- Fan, Cindy and Allen J. Scott. 2003. 'Industrial Agglomeration and Development: A Survey of Spatial Economic Issues in East Asia and a Statistical Analysis of Chinese Regions.' *Economic Geography*. 79(3): 295–319.
- Feldman, Maryann P. 2001. 'The Entrepreneurial Event Revisited: Firm Formation in a Regional Context.' *Industrial and Corporate Change*. 10(4): 861–91.
- Felkner, John and Robert M. Townsend. 2011. 'The Geographic Concentration of Enterprise in Developing Countries.' *Quarterly Journal of Economics*. 126(4): 2005-2061.
- Foreign Investment Agency (FIA), Ministry of Planning and Investment of Vietnam. 2009. Vietnam Investment Guide. Downloaded from:
<http://www.aseankorea.org/files/upload/board/120/6/2009%20Investment%20Guidebook%20Vietnam.pdf> on 20.1.2011.
- Freshfields Bruckhaus Deringer. 2006. Vietnam-New Investment Law. Briefing downloaded from: <http://www.freshfields.com/publications/pdfs/2006/14675.pdf> on 20.1.2011.
- Fujita, Masahisa, Paul R. Krugman and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, Mass.: MIT Press.
- Glaeser, Edward and Janet Kohlase. 2004. 'Cities, regions and the decline of transport costs.' *Papers in Regional Science*. 83(1): 197-228.
- Gorman Sean P, and Rajendra Kulkarni. 2004. 'Spatial Small Worlds: New Geographic Patterns for an Information Economy.' *Environment and Planning B: Planning and Design*. 31(2): 273-296.
- Government of Vietnam, 2000. *Vietnam Strategy for Socio-Economic Development in the Period 2001-2010*. Available at:

<http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/EASTASIAPACIFICE/XT/VIETNAMEXTN/0,,contentMDK:20209878~menuPK:441765~pagePK:1497618~piPK:217854~theSitePK:387565,00.html> [Accessed 1.3.12]

Greenstone, Michael, Richard Hornbeck and Enrico Moretti. 2010. 'Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings.' *Journal of Political Economy*. 118(3): 536-598.

Griliches, Zvi. 1998. *R&D and Productivity: The Econometric Evidence*. Chicago: University of Chicago Press.

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

Hatzichronoglou, Thomas. 1997. 'Revision of the High-Technology Sector and Product Classification.' *OECD Science, Technology and Industry Working Papers*. 1997(02): OECD Publishing

Head, Keith, John Ries and Deborah Swenson. 1995. 'Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States.' *Journal of International Economics*. 38(3-4): 223-247.

Head, Keith and Thierry Mayer. 2004. 'Market Potential and the Location of Japanese Investment in the European Union.' *The Review of Economics and Statistics*. 86(4): 959-972.

Helsley, Robert W. and William C. Strange. 1990. 'Matching and agglomeration economies in a system of cities.' *Regional Science and Urban Economics*. 20(2): 189-212.

Henderson, J. Vernon. 1986. 'Efficiency of resource usage and city size.' *Journal of Urban Economics*. 19(1): 47-70.

Henderson, J. Vernon. 1997. 'Externalities and Industrial Development.' *Journal of Urban Economics*. 42(3): 449-470.

Hotelling, Harold. 1929. 'Stability in Competition.' *Economic Journal*. 39: 41-57.

Howard, Emma, Carol Newman and Jacco Thijssen. 2011. 'Are Spatial Networks of Firms Random? Evidence from Vietnam.' UNU-WIDER Working Paper No. 87

Howard, Emma, Carol Newman and Finn Tarp. 2012. 'Measuring Industry Agglomeration and Identifying the Driving Forces.' UNU-WIDER Working Paper No. 84

Huang, Yukon and Alessandro Magnoli Bocchi. 2009. 'Lessons from experience: reshaping economic geography in East Asia' In *Reshaping Economic Geography in East Asia*. In Yukon Huang and Alessandro Magnoli Bocchi, eds. Washington, DC: World Bank, EAP Companion Volume to the WDR 2009.

Jackson, Matthew O. and Asher Wolinsky. 1996. A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*. 71: 44-74.

Jackson, Matthew O. and Brian W. Rogers. 2005. 'The Economics of Small Worlds.' *Journal of the European Economic Association*. 3(2-3): 617-627.

Jaffe, Adam B. 1986. 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value.' *American Economic Review*. 76(5): 984-1001.

Kenney, Martin and Urs von Burg. 1999. 'Technology, Entrepreneurship and Path Dependence: Industrial Clustering in Silicon Valley and Route 128.' *Industrial and Corporate Change*. 8(1): 67-103.

Kim, Sukkoo. 1999. 'Regions, resources and economic geography: Sources of U.S. regional comparative advantage, 1880-1987.' *Regional Science and Urban Economics*. 29: 1-32.

Koenig, Pamina. 2009. 'Agglomeration and the export decisions of French firms.' *Journal of Urban Economics*. 66(3): 186-195.

Krugman, Paul R. 1991. 'Increasing Returns and Economic Geography.' *Journal of Political Economy*. 99: 483-99.

Markusen, James and Anthony J. Venables. 1999. 'Foreign direct investment as a catalyst for industrial development.' *European Economic Review*. 43: 335-56.

- Marshall, Alfred. 1920. *Principles of Economics*. London: Macmillan and Co. Ltd.
- McFadden, Daniel. 1974. 'Conditional logit analysis of qualitative choice behaviour.' In: P. Zarembka, ed., *Frontiers In Econometrics*. New York: Academic Press.
- Milgram, Stanley. 1967. 'The Small-World Problem.' *Psychology Today* 2: 60-67.
- Overman, Henry G. and Diego Puga. 2010. 'Labour Pooling as a Source of Agglomeration: An Empirical Investigation.' In: Glaeser, Edward L., ed. *Agglomeration Economics*. Chicago: University of Chicago Press.
- Pack, Howard and Kamal Saggi. 2006. 'The case for industrial policy: a critical survey.' *Policy Research Working Paper Series 3839*: The World Bank.
- Pagan, Adrian. 1984. 'Econometric Issues in the Analysis of Regressions with Generated Regressors.' *International Economic Review*. 25(1): 221-47.
- Picone, Gabriel A., David B. Ridley, and Paul A. Zandbergen. 2009. 'Distance decreases with differentiation: Strategic agglomeration by retailers.' *International Journal of Industrial Organization* 27(3): 463-473.
- Prescott, Edward C. and Michael Visscher. 1977. 'Sequential Location among Firms with Foresight.' *The Bell Journal of Economics* 8: 378-93.
- Rodríguez-Clare, Andrés. 1996. 'Multinationals, Linkages, and Economic Development.' *The American Economic Review*. 86(4): 852-873
- Rosenthal, Stuart S. and William C. Strange. 2001. 'The Determinants of Agglomeration.' *Journal of Urban Economics*. 50(2): 191-229.
- Son, Dang Kim. 2009. 'Rural Development and Issues in Vietnam: Spatial Disparities and Some Recommendations.' In *Reshaping Economic Geography in East Asia*. Yukon Huang and Alessandro Magnoli Bocchi, eds. Washington, DC: World Bank, EAP Companion Volume to the WDR 2009.
- Siba, Eyerusalem, Måns Söderbom, Arne Bigsten, and Mulu Gebreeyesus. 2012. 'Enterprise Agglomeration, Output Prices and Physical Productivity: Firm-Level Evidence from Ethiopia.' Unpublished: University of Gothenburg

Siedschlag, Iulia, Xiaoheng Zhang and Donal Smith. 2009. 'What Determines the Location Choice of Multinational Firms in the ICT Sector?' ESRI Working Paper 336.

Smithies, Arthur. 1941. 'Optimum Location in Spatial Competition.' *Journal of Political Economy*. 44: 423-39.

Uchida, Hirotugu, and Andrew Nelson. 2008. 'Agglomeration Index: Towards a New Measure of Urban Concentration.' Background paper for the WDR 2009.

Wasserman, Stanley and Katherine Faust. 1995. *Social Network Analysis*. Cambridge: University Press.

Watts, Duncan and Stephen Strogatz. 1998. 'Collective Dynamics of 'Small-World' Networks. *Nature*. 393: 440-42.

Appendix A

Chapter 2 Appendix: Vietnam Maps

Figure 1: Clustering of Manufacturing Firms 2002

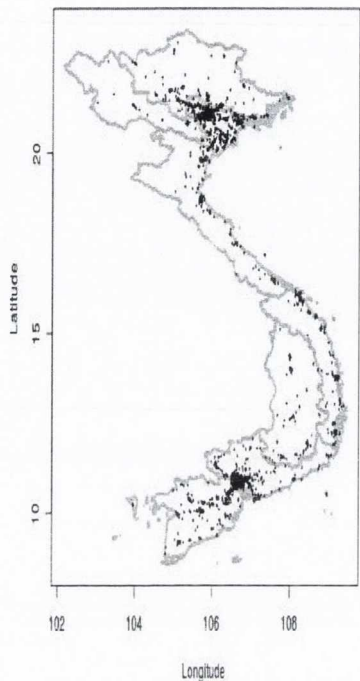


Figure 2: Clustering of Manufacturing Firms 2003

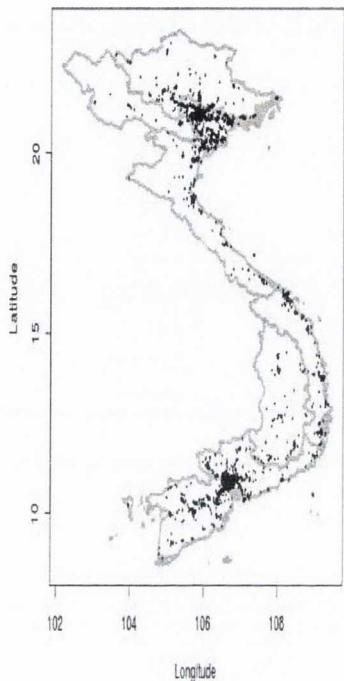


Figure 3: Clustering of Manufacturing Firms 2004

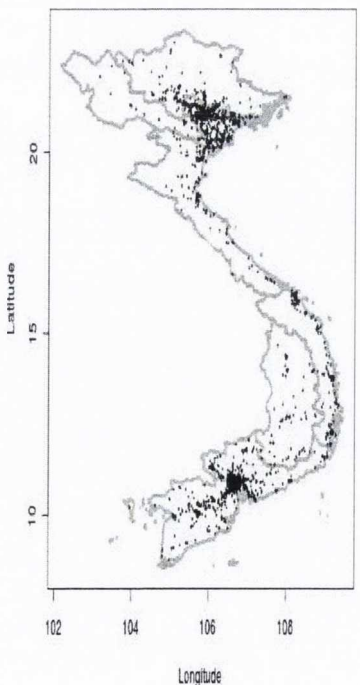


Figure 4: Clustering of Manufacturing Firms 2005

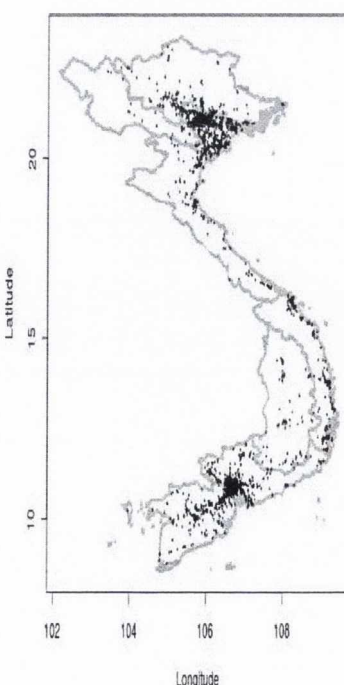


Figure 5: Clustering of Manufacturing Firms 2006

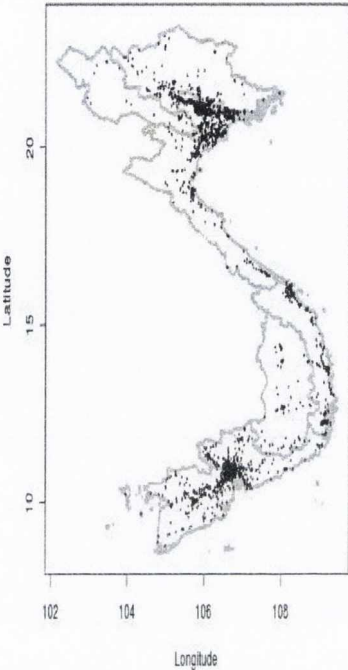


Figure 6: Clustering of Manufacturing Firms 2007

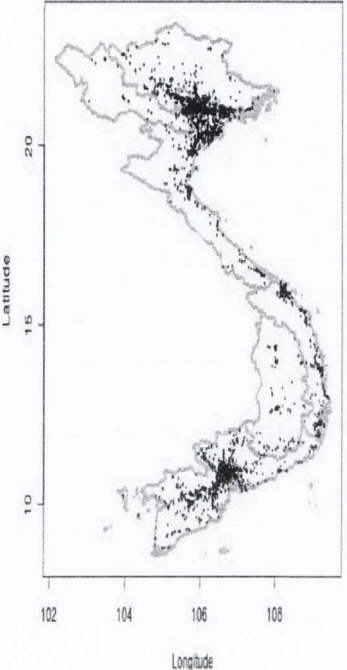


Figure 7: Clustering of Large Firms 2002

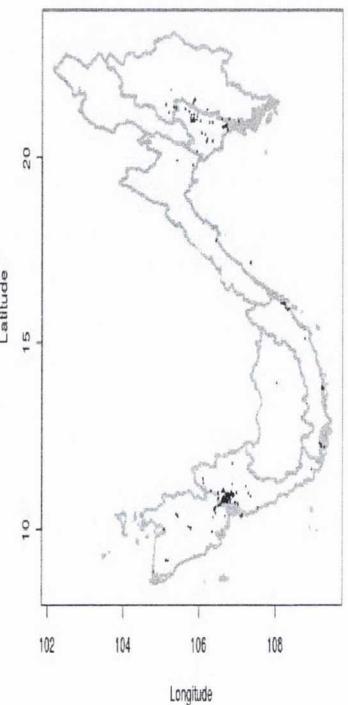


Figure 8: Clustering of Large Firms 2007

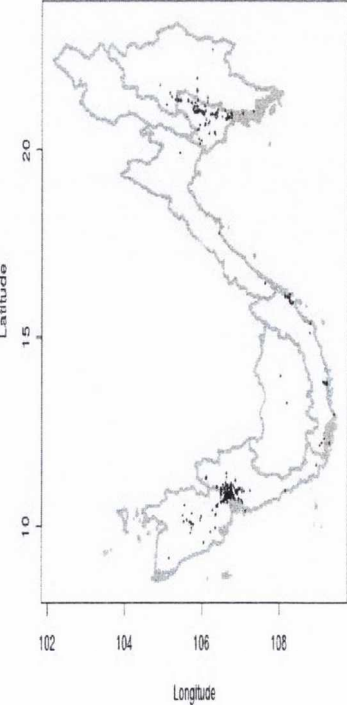


Figure 9: Clustering of Small Firms 2002

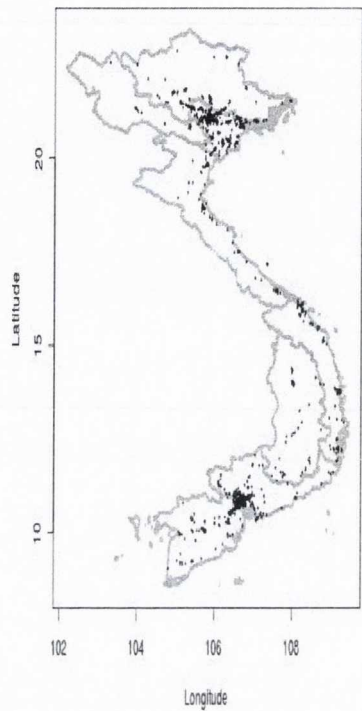


Figure 10: Clustering of Small Firms 2007

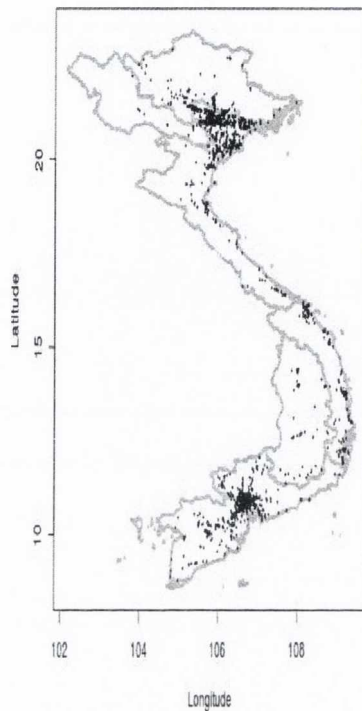


Figure 11: Clustering of Medium Firms 2002

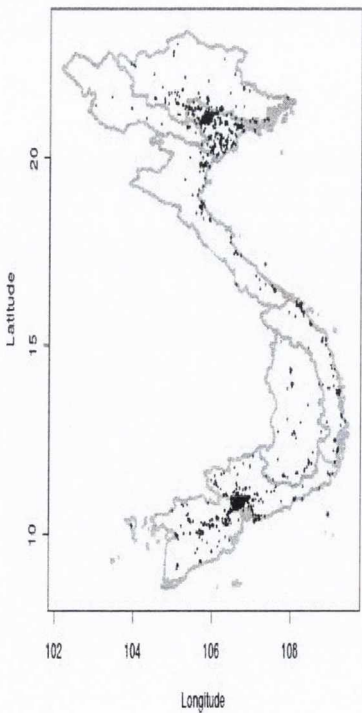


Figure 12: Clustering of Medium Firms 2007

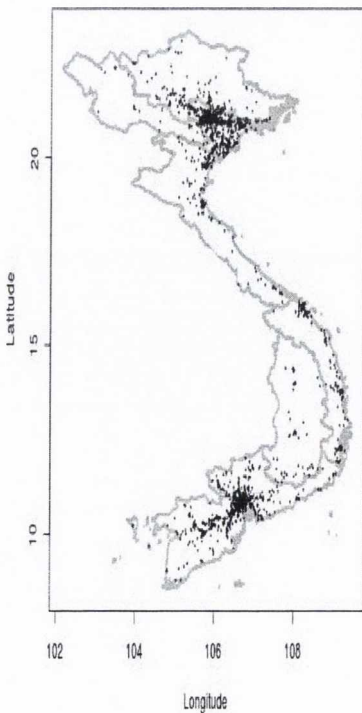


Figure 13: Clustering of State-owned Firms 2002

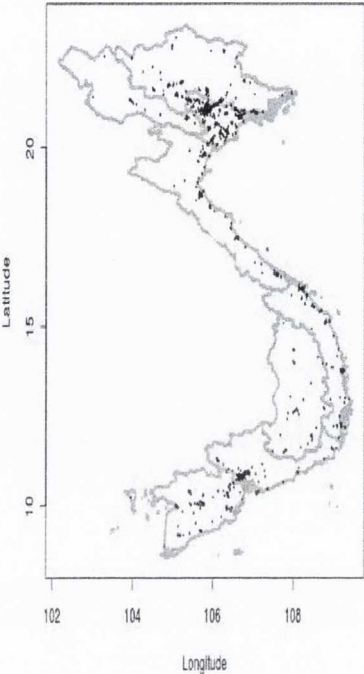


Figure 14: Clustering of State-owned Firms 2007

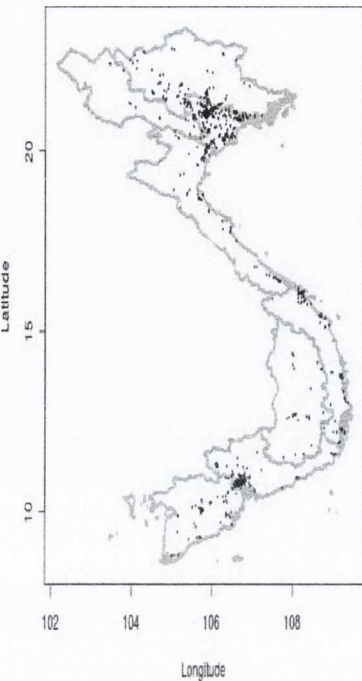


Figure 15: Clustering of Privately-owned Firms 2002

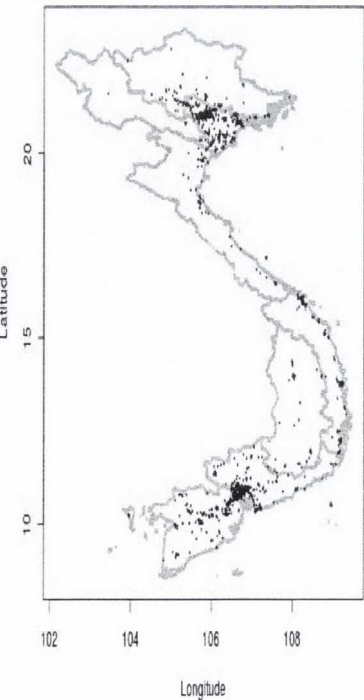


Figure 16: Clustering of Privately-owned Firms 2007

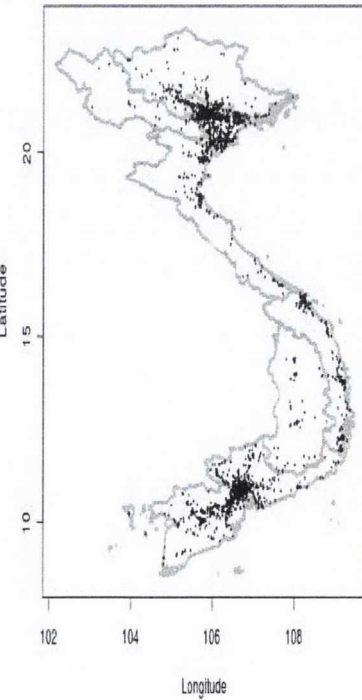


Figure 17: Clustering of Foreign-owned Firms 2002

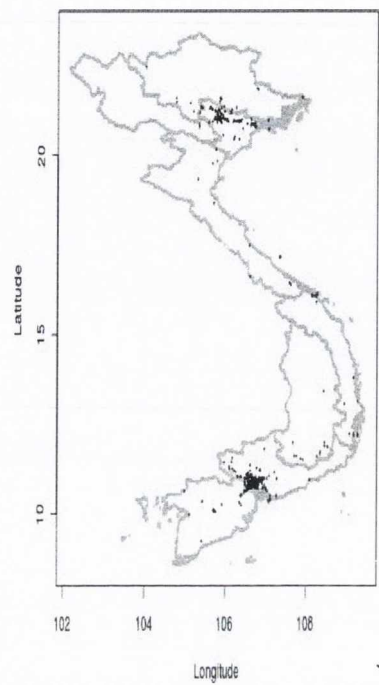
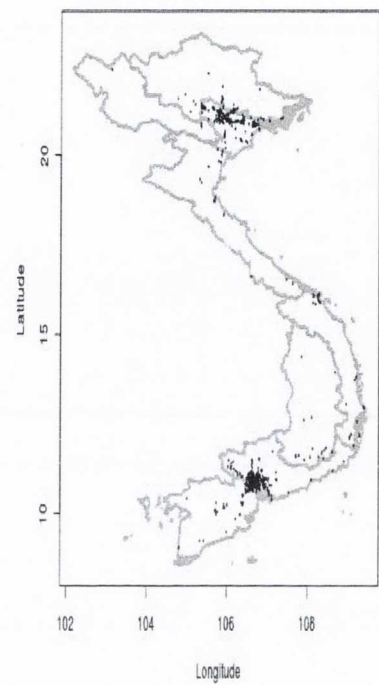


Figure 18: Clustering of Foreign-owned Firms 2007



Appendix B

Chapter 3 Appendix:

Description of Manufacturing Sector Codes

Description of manufacturing sector codes

- 1 Production, processing and preserving of meat and meat products
- 2 Processing and preserving of fish and fish products
- 3 Processing and preserving of fruit and vegetable
- 4 Manufacture of vegetable and animal oils and fats
- 5 Manufacture of milk and dairy products
- 6 Processing of rice and flour
- 7 Other food manufacturing
- 8 Manufacture of prepared feeds for animals
- 9 Manufacture of cakes, jams, candy, coca, chocolate products
- 10 Manufacture of sugar
- 11 Manufacture of alcohol and liquors
- 12 Manufacture of beer
- 13 Manufacture of alcohol-free beverages eg. soft drinks, mineral waters
- 14 Manufacture of cigarettes and other tobacco products
- 15 Manufacture of fibre (all kinds)
- 16 Manufacture of textile products (all kinds)
- 17 Manufacture of ready-made apparel (all kinds)
- 18 Manufacture of leather and leather products
- 19 Manufacture of wood and by-products
- 20 Manufacture of pulp, paper and by-products
- 21 Publishing
- 22 Printing

- 23 Manufacture of coke, coal and other by-products
- 24 Manufacture of gasoline and lubricants
- 25 Manufacture of fertilizers
- 26 Manufacture of other chemical products
- 27 Manufacture of pharmaceuticals, medicinal chemicals and botanical products
- 28 Manufacture of processed rubber and by-products
- 29 Manufacture of plastic and by-products
- 30 Manufacture of glass and by-products
- 31 Manufacture of other non-metallic mineral products
- 32 Manufacture of cement and cement products
- 33 Manufacture of metal and metal products
- 34 Manufacture of general purpose machinery
- 35 Manufacture of special purpose machinery
- 36 Manufacture of domestic appliances
- 37 Manufacture of electrical machinery
- 38 Manufacture of electrical equipment
- 39 Manufacture of machinery used for broadcasting, television and information activities
- 40 Manufacture of medical and surgical equipment
- 41 Manufacture of precision and optical equipment
- 42 Manufacture of transportation machinery and equipment
- 43 Manufacture of other goods