T.R.
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS
PhD PROGRAMME

PhD DISSERTATION

AGGLOMERATION ECONOMIES AND GEOGRAPHIC DISTRIBUTION OF MANUFACTURING INDUSTRIES IN TURKEY

OYA KENT
10710205

THESIS ADVISOR
PROF. MURAT DONDURAN

İSTANBUL
2015

TC
YILDIZ TEKNİK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT ANA BİLİM DALI
İKTİSAT DOKTORA PROGRAMI

DOKTORA TEZİ

AGGLOMERATION ECONOMIES AND GEOGRAPHIC DISTRIBUTION OF MANUFACTURING INDUSTRIES IN TURKEY

OYA KENT
10710205

Tezin Enstitüye Verildiği Tarih: 20/08/2015
Tezin Savunulduğu Tarih: 28/09/2015

Tez Oy birliği / Oy çokluğu ile başarılı bulunmuştur.

	Unvan Ad Soyad	İmza
Tez Danışmanı :	**Prof. Dr. Murat Donduran**	M. Donduran
Jüri Üyeleri :	**Prof. Dr. Fatma Doğruel**	F. D.
	Prof. Dr. Hüseyin Taştan	
	Prof. Dr. Ercan Eren	Eren
	Prof. Dr. H. Targan Ünal	

İSTANBUL
08/2015

ABSTRACT

AGGLOMERATION ECONOMIES AND GEOGRAPHIC DISTRIBUTION OF MANUFACTURING INDUSTRIES IN TURKEY

Oya Kent

August, 2015

A generally observed phenomenon is that most economic activities and industries are not uniformly distributed across space but tend to cluster in certain locations in developed and developing countries. Turkey, as a developing country, also exhibits uneven distribution of economic activity across regions, notably manufacturing employment demonstrates a prominent diversification between eastern and western regions. This study measures and evaluates the agglomeration patterns in Turkish manufacturing industries for the period 2003-2008 by exploring a confidential establishment level micro dataset compiled by TurkStat. In measuring agglomeration, it follows the methodology proposed by Ellison and Glaeser (1997) (EG) in which they present an agglomeration index of plants based on a test of comparison between the observed geographic distribution of plants and a random distribution. It also relates the geographic concentration measures to industry characteristics and agglomeration forces proposed by theory. The main finding is that Turkish manufacturing industries follow higher levels of agglomeration on average compared to developed countries. However, the stylized fact that low-technology and traditional industries exhibit higher levels of agglomeration is also valid for the case of Turkey. The study also aims to test factors behind agglomeration based on the hypotheses asserted by theory. Main finding related to that is, transportation costs do have a significant effect on industrial agglomeration, as suggested by New Economic Geography. The contribution of this study to the literature is twofold: (i) the period it analyses has not been examined within agglomeration framework before, so it will shed light on the agglomeration patterns of Turkish manufacturing for the post-2000 period and (ii) the *EG* index it employs is the first attempt to measure agglomeration in Turkish manufacturing using a micro level dataset.

Keywords: Agglomeration, geographic concentration, manufacturing

ÖZ

TÜRKİYE'DE YIĞILMA EKONOMİLERİ VE İMALAT SANAYİİNİN COĞRAFİ DAĞILIMI
Oya Kent
Ağustos, 2015

Ekonomik aktivite ve endüstrilerin büyük bir kısmının mekan içinde eşit dağılmadığı, belirli bölgelerde kümelenme gösterdiği hem gelişmiş hem de gelişmekte olan ülkelerde genel olarak gözlemlenen bir olgudur. Türkiye, gelişmekte olan bir ülke olarak bölgeleri arasında ekonomik aktivitenin dağılımı anlamında eşitliksiz bir yapı sergilmektedir, özellikle imalat sanayii istihdamı doğu ve batı bölgeleri arasında göze çarpan bir farklılaşma mevcuttur. Bu çalışma, TÜİK tarafından sağlanan mikro verileri kullanarak 2003-2008 dönemi için Türk imalat sanayiindeki yığılma örüntüsünü ölçerek değerlendirmeyi amaçlamaktadır. Yığılma ölçümünde Ellison and Glaeser (1997) (EG) tarafından önerilen, firmaların gözlemlenen dağılımı ile rassal dağılımı arasındaki karşılaştırmaya dayanarak elde edilen yığılma endeksini kullanmaktadır. Bunun yanında coğrafi yoğunlaşma ölçümlerini teori tarafından öne sürülen yığılma kuvvetleri ile ilişkilendirmektedir. Çalışmanın ana bulgusu, Türk imalat sanayiinin gelişmiş ülkelere oranla ortalamanın üzerinde bir yığılma gösterdiğidir. Ayrıca, düşük teknolojili ve geleneksel endüstrilerin daha yüksek düzeyde yığılma gösterdiği olgusu Türkiye için de geçerli olmaktadır. Çalışma aynı zamanda, yığılmanın ardındaki faktörleri teorilerin öne sürdüğü hipotezlere dayanarak test etmeyi amaçlamaktadır. Bununla ilgili temel bulgu Yeni Ekonomik Coğrafya yaklaşımının belirttiği gibi taşımacılık maliyetlerinin endüstriyel yığılmalar üzerinde önemli bir etkisinin olduğudur. Bu çalışma iki yönden var olan literatüre katkıda bulunmaktadır. Birincisi, incelenen dönem yığılma çerçevesi içinde daha önceki çalışmalar tarafından analiz edilmemiştir, bu bağlamda çalışma Türkiye imalat sanayiinin yığılma örüntüsüne 2000 sonrası dönem için ışık tutmaktadır. İkincil olarak da Türk imalat sanayiindeki yığılmayı ölçmek için mikro veri kullanarak *EG* endeksini ilk kez hesaplama girişiminde bulunmaktadır.

Anahtar kelimeler: : Yığılma, coğrafi yoğunlaşma, imalat

ACKNOWLEDGEMENT

This dissertation is a result of rigorous effort that involves a three years of research. There are a number of people whom I owe a great debt of gratitude and hereby I would like to thank and acknowledge them for their invaluable supports. Firstly, I would like to thank Professor Murat Donduran for accepting the supervision of this dissertation. I am sincerely grateful to him for his continuous support of my Ph.D and related research, for his patience, motivation, and immense knowledge. I would like to express my gratefulness to him also for sparing his time for my endless questions. I am also indebted to him for providing me a stressless environment during his supervision which enabled me to complete my dissertation smoothly. I would also like to thank my committee members, Professor Fatma Doğruel and Professor Hüseyin Taştan for accepting to serve as my committee members even at hardship.

Furthermore, I would like to use this opportunity to express my gratitude to Professor Fatma Doğruel who always kept me motivated even at tough times and helped me to go through it. I also owe her gratefulness as she never gave up discussing about my research with patience. Along these discussions, not only I am inspired by her comments and questions, but also her immense knowledge in the field of study enlightened me and broadened my viewpoint. I am thankful to her for sparing a part of her valuable and limited time to me whenever I asked for guidance. Thus, once again I would like to repeat my gratefulness to her for always encouraging me, without her this dissertation would not be as valuable for me.

The six months of this research was conducted at University of Barcelona in Spain. I would like to thank Professor Enrique López-Bazo sincerely for inviting me to Regional Quantitative Analysis Group, University of Barcelona as a research fellow and accepting my supervision there. Working with him and experiencing a fellowship in a different research environment was inspiring for me. Moreover, I need to acknowledge that Scientific and Research Council of Turkey, TUBITAK has provided the financial support for this part of research in Barcelona. I would like to thank TUBITAK for providing me the opportunity to study abroad for my dissertation research. I also would like to express my gratefulness to Professor Targan Ünal, my head of department at Okan University, for offering me infinite understanding and flexibility at working conditions. His support has always been there for motivating me to complete my PhD.

I would like to express special and deepest thanks to an important person, my significant other, Evren Bırık. I am indebted to him for his intimate support, interest and endless patience during the entire period. Without his love and understanding, it would be far difficult.

Finally my profound and endless thanks will be for my family. This is a great chance for me to express my gratefulness to my father Cengiz Kent, my mother Günnur Kent and my dear sister İpek Kent. My family has always shown respect to my choice of willing to become an academician and supported me on this tough way. I need to thank them for everything that they have done for me up until now, although it will never be possible to fully pay off their efforts. Lastly, I am extremely delighted for completing this dissertation which at the same time bears another sentimental meaning for me. I hope I would be realizing the unfilled dream of my mother by finishing my PhD. No wonder, this dissertation will be dedicated to her.

ISTANBUL OYA KENT
August 2015

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

A generally observed phenomenon is that most economic activities are not uniformly distributed across space but tend to cluster in certain locations in both developed and developing countries. This observation brings into question the reasons and effects of agglomeration phenomenon which has been an issue of considerable interest both from policy and academic perspective. In this respect, understanding the major reasons behind uneven distribution of economic activity in different geographies has long been in the research agenda of social scientists.

Although agglomeration and industrial clustering phenomena have emerged as interesting research topics in the last three decades, discussion regarding the location of economic activity dates back to 19th century and associated with names such as von Thünen, Marshall, Weber, Ohlin, Hoover, Lösch, Isard and Beckmann (Karlsson, 2008). Location theorists, economic geographers, regional scientists, urban economists, shortly researchers from several disciplines with different research traditions have employed a diverse set of theoretical frameworks and analytical approaches in examining agglomeration of economic activity. However, Krugman (1991a) has been the researcher clarifying the microeconomic underpinnings of both spatial economic agglomerations and regional imbalances at national and international levels within a full-fledged general equilibrium framework. The seminal work of Krugman (1991a) has brought forth the field called New Economic Geography and stimulated a new way of theorizing. The contribution of New Economic Geography is very significant in the sense that bringing back the notion of location to the core of economics by using the tools of mainstream economics.

While theoretical underpinnings of agglomeration have long been analysed and developed extensively during the last three decades, taking a step behind and posing the question of how to measure agglomeration correctly has also been another field of study for many researchers. The literature covers a variety of measures to evaluate the distribution of economic activity across geographic units. Many of the measures employed are atheoretical and fails to differentiate geographic concentration from randomness. Ellison and Glaeser (1997) have introduced the first model based agglomeration measure which measures the extent of geographic concentration once industrial concentration is accounted for. Following this seminal work, developments in the field has been extensive and advanced rapidly especially during the last decade.

This dissertation will be investigating the geographic concentration of economic activities in Turkey for the post-2000 period. The focus of the study is manufacturing industry by virtue of having been considered as a fundamental indicator in the development process of a country. In measuring agglomeration of industries, the methodology proposed by Ellison and Glaeser (1997) is followed. It is based on a statistical model of location choice that takes random distribution of plants across spatial units as a threshold to compare observed geographic distribution of plants. Based on this model, an index of industrial agglomeration is derived. This study makes use of Ellison-Glaeser (EG) index due to some useful properties it owns as asserted by the authors, such as being comparable across industries in which the size distribution of firms differ, comparable across countries irrespective of divergences in the level of geographic aggregation and comparable over time. Therefore, the first objective of the dissertation is to provide empirical evidence on the extent of agglomeration in Turkish manufacturing industries. Following that it aims to answer several questions; whether stylised facts observed in agglomeration patterns in developed countries also hold for the case of Turkey, whether industries belonging to different technology groups exhibit different agglomeration patterns and whether agglomeration over time is stable or there is a significant trend. In addition, having uncovered agglomeration patterns in Turkish manufacturing, the dissertation will be analysing the sources of agglomeration. It attempts to test whether relevant agglomeration theories are supported by Turkish data or not. Hereby, it is expected to give the reader a profound insight about the spatial distribution of production in Turkey from agglomeration economies perspective.

While answering the central questions of dissertation, the reader will also have essential information about the historical development of location theory. In the second chapter, these developments will be summarized in a context that allows one to gain insight about the path to New Economic Geography with a comparative view. Given that the literature on location of economic activity has a long but involved history, this study strives to discuss the subject within the boundaries of economics without getting much involved in urban economics, regional science or location theory. After presenting an overview of the literature within the defined framework, the sources of agglomeration on a theoretical basis will be discussed. This section, at the same time, will be providing the theoretical background for chapter five where sources or determinants of agglomeration will be analysed empirically. In other words, it will be telling the stories behind the explanatory variables employed in the analysis.

Having reviewed the theoretical background of agglomeration, measurement of it arises as a compelling issue. Third chapter of the dissertation revives the question of how to quantify agglomeration. In this chapter, developments regarding measurement of geographic concentration and agglomeration will be elaborated. Common measures of agglomeration employed within the literature will be investigated in detail within a hierarchical structure in order to give insight into the successive improvements. The *EG* index, to be applied in the next chapter, derived from the location choice model proposed by Ellison and Glaeser (1997) will formally be presented herein besides other measures. In a way, this chapter may be conceived as an integral part of the first chapter which quantifies the theoretically founded agglomeration phenomenon.

Chapter four will be employing the *EG* index to detect the extent of agglomeration in Turkish manufacturing industries. The index requires the use of plant level micro data and location of employment for each plant. This type of micro data is compiled by Turkish Statistics Institute (TurkStat) and available to researchers under certain conditions due to data confidentiality issues. The study makes use of this micro data set in order to examine the agglomeration behaviour of manufacturing industries. The analysis will cover the period of 2003-2008 due to the fact that after the last year of this period industrial classification system changes and does not allow for direct comparison. Manufacturing industries are classified according to NACE Rev. 1.1. industrial classification system. The analysis will be conducted in a quite fine scale such that agglomeration patterns will be examined for four-digit NACE Rev. 1.1. industries. After presenting that, agglomeration will also be handled by considering it technology-wise and over time. Co-agglomeration patterns within industries and between industry pairs will also be investigated. Finally, agglomeration patterns of Turkish manufacturing will be compared with other country studies as the index allows for that.

In chapter five, agglomeration theories discussed in the first chapter will be tested in order to examine the determinants of agglomeration in a regression framework. The factors that help to explain the causes of agglomeration will be analysed by using an empirical specification that allows one to relate the *EG* index of industry agglomeration to industry characteristics and agglomeration forces put-forward by the theory. *EG* indices obtained in previous chapter will constitute dependent variable while explanatory variables will be derived from the dataset in accordance with the theory. Agglomeration hypotheses will be tested via panel data models. Traditional static panel data models with fixed effect and random effect assumptions will be explored.

Finally, major findings regarding the central questions of the dissertation will be summarized in the conclusion part. In addition a number of topics will be debated for further research.

This dissertation will be contributing to the literature in two aspects. Previous studies examining the geographic concentration of economic activities in Turkey have employed highly aggregated data at provincial or regional level and cover a certain period of time, namely 1980-2000, due to the inconsistency of regional data from then on. Therefore, we have very limited knowledge about geographic concentration of industries for the post-2000 period. So, the first contribution will be examining the post-2000 period which has not been examined within this line of research for the case of Turkey. And secondly, it is not very common to make use of micro data in regional analysis. A wide range of studies related to Turkey exploit highly aggregated data to understand regional disparities. In this regard, exploring micro-data to reveal the agglomeration patterns of the Turkish manufacturing industry by using EG index, which is also firstly explored, will be shedding a new light on the subject.

2 THEORETICAL RETROSPECT ON THE AGGLOMERATION ECONOMIES

Over the last 25 years, the uneven distribution of economic activity across space has gained a renewed interest in economics, especially with the emergence of "new economic geography" (NEG) following the seminal work of Krugman (1991a). However, the *location* of economic activity, has a long history which dates back to 19th century starting with von Thünen, considered as the "founding God" (Samuelson, 1983, p. 1468) of economic geography and location theory, whose work has inspired urban and regional economics later on. Having had such an old history, it is not that straightforward to embed and present the *location* as a focus of analysis in a distinct and single theoretical framework. Indeed, the theoretical background of location choice phenomenon is related to a number of diverse frameworks, s.t. location theory, regional economics, urban economics, economic geography, within which the approach to *location* differs.

Nevertheless, it is well beyond the scope of this section to provide an extensive review of the history of locational analysis within the vast body of theoretical literature, which is interest of another research per se. Instead, it will provide a brief overview of the locational analysis in order to enlighten the course towards the emergence of NEG which has brought back the *location* to the core of economics.

From this perspective, one is able to comprehend how origins and the historical discussions of the location theory combined with the developments in economic theory and interactions with urban and regional economics evolves into the emergence of so-called New Economic Geography. After presenting a general overview of the location analysis in the aforesaid context, the study specifically focuses on the potential sources of localization of economic activity which have been previously mentioned in the literature. This framework basically requires to handle the issue within the scope of externalities which will establish essential links with Section 5 and help to gain insights to be reflected in Section 4.

2.1 Overview: The Path to New Economic Geography

In order to understand how analysis of location has been integrated into economic theory by 1990's, it is essential to shed light on the relationship between economic geography and other disciplines such as international economics, urban and regional economics and location theory. The way how it is related lies behind the definition of economic agglomeration or concentration in geographical space. Agglomeration of economic activities may arise at many geographical levels, as Fujita and Krugman (2004, p. 140) puts:

> "*For example, one type of agglomeration arises when small shops and restaurants are clustered in a neighbourhood. Other types of agglomeration can be found in the formation of cities, all having different sizes, ranging from New York to Little Rock; in the emergence of a variety of industrial districts; or in the existence of strong regional disparities within the same country. At the other extreme of the spectrum lies the core-periphery structure of the global economy corresponding to the North-South dualism*"

Spatial unit of reference or *spatial scale* distinguishes those various types of agglomerations, nonetheless whichever scale is executed, "the emergence of economic agglomeration is naturally associated with the emergence of inequalities across locations, regions or nations" (Fujita and Thisse, 2009, p. 109). From this point of view, it is not so hard to establish a link between economic geography and urban, regional and international economics.

The following section outlines the development of analysis of location or space within different but related disciplines in a historical context. Section 2.1.2 points to the essential links between international economics, more precisely international trade and economic geography.

As a matter of fact, Fujita, Krugman, and Venables (2001) present regional, urban systems and international models within NEG framework, in the end whatever it is called it's all about where economic activity takes place and why.

2.1.1 Locational Analysis in Retrospect

The location, as a non-negligible factor of economic activity, has long remained outside the economic analysis. Certainly, it is not because economists find economic geography, which studies where economic activity takes place and why, uninteresting or unimportant. In fact, it has always been important, but due to the fact that they have regarded it as technically intractable, it has been neglected until the emergence of New Economic Geography in early 1990's. NEG, by exploiting the new tools developed in the field of industrial organization in mid 1970's and been captivated by the increasing returns revolution, has succeeded to explain why, how and when the economic activity may be concentrated only in a few locations in a full-fledged general equilibrium framework (Fujita, Krugman, and Venables, 2001).

Without doubt, NEG did not discover *space* as an integral part of economic analysis out of nothing. The history of spatial analysis is very deep and rich such that its roots go back to the beginning of 19th century. Within the framework of *general location theory* [1], von Thünen is regarded as the "founding God" (Samuelson, 1983, p. 1468) of economic geography and location theory. Following his monumental work, "a variety of pioneering ideas have been developed periodically by great location theorists, geographers and economists such as Launhardt (1885), Marshall (1890), Weber (1909), Hotelling (1929), Ohlin (1933), Christaller (1933), Palander (1935), Kaldor (1935), Hoover (1936, 1937), Lösch (1940), and Isard (1949)" (as cited in Fujita (2010, p. 2)). And, in the second half of the 20th century, interest of researchers in the subject increased giving way to the development of regional science in 1950's, urban economics in 1960's and NEG in 1990's (Fujita and Krugman, 2004).

Given the enormous magnitude of literature on spatial economics accumulated within two hundred years and being addressed by different fields, it is quite a hard task to survey the bulk of the literature here. But, since "the historic book *The Isolated State* (1826), by von Thünen, signified the birth of spatial economics" Fujita (2010, p. 3) and the question of "how the economy organizes its use of space" redirects one to consider models pioneered by von Thünen in the early 19th century, it's worth mentioning its keystones in a nutshell.

von Thünen (1826) presumed an "isolated state" where a very large town is located at the center supplied by farmers in the surrounding countryside. It is assumed that crops differ in terms of yield and transportation costs and allowed for the possibility that each crop could be produced with different intensities of cultivation. Based on this, he asks two main questions: how to allocate land around the town to minimize the combined cost of transportation and production and

[1] Fujita (2010) considers *general location theory* as the the most fundamental theory of spatial economics and denotes NEG as representing the newest wave in the development of general location theory. What he refers as *spatial economy* is a broad term according to whom which should encompass all branches of economics dealing with the analysis of economic processes and developments in geographical space. In this respect, any field of economics dealing with space, for instance urban economics, regional economics, international trade, are counted within the realm of spatial economics.

how will the land be allocated in case of competition between self-interested farmers and landowners. Then, he showed that land rents declines from a maximum at the town to zero at the outermost limit of cultivation due to competition among the farmers. There is a trade-off between land rents and transport costs faced by each farmer. Since transport costs and yield differ among crops, a pattern of concentric rings of production emerge (Fujita, 2010).

Fujita, Krugman, and Venables (2001) and Fujita (2010) appraise von Thünen's model as ingenious and quite deep analysis of spatial economy despite it seems simple and obvious. According to Samuelson (1983, p. 1468) von Thünen "not only created *marginalism* and *managerial economics*, but also elaborated one of the first models of *general equilibrium* and did so in terms of realistic *econometric parameters*". He adds that "Modern geographers claim Thünen. That is their right. But economists like me, who are not all that taken with location theory, hail Thünen as more than a location theorist. His theory is a theory of general equilibrium" (Samuelson, 1983, p. 1482). And as a spatial economist (Fujita, 2010, p. 3) claims that "Thünen's theory is a theory of general equilibrium in space, or for short, a general location theory" [2].

With the increasing economic importance of manufacturing over the second half of the 19th century, location of industry gained a renewed interest from economists. Following the seminal work of Thünen, industrial location theory has advanced significantly in the first half of the 20th century, especially with the contributions of German scholars. Preliminary formal analysis of industrial location in late 19th and early 20th century were presented in a partial equilibrium framework where location of plants, markets, producers and prices were taken as given. Launhardt (1885)'s analysis is considered among the first of these where market area analysis and spatial price policy is studied. Another important work which later influenced development of industrial location theory is Weber (1909)'s analysis which considers optimal location of the plant that minimizes the total transport cost per unit output. Besides being a pioneering theory of industrial location, his theory was deprived of price analysis and market structure which means that prices of inputs and inputs were not determined endogenously, and due to this fact his industrial location theory was not sufficiently appreciated by economists (as cited in Fujita (2010)). It is also possible to evaluate this as an historical constraint on economic modelling in the sense that at that time neither the non-competitive theory of markets nor game theoretic approach to interactive

[2]For further discussion on von Thünen's model, the reader may refer to the mentioned articles.

behaviour was well-developed.

By the way, there has been an important contribution to the field outside the continental Europe in late 19th century. In England, Marshall (1890) presented his famous study *Principles of Economics* in which he devoted a chapter on industrial agglomeration where he examined systematically the reasons behind the concentration of specialized industries in particular locations. Marshall stressed the importance of externalities, in the formation of economic agglomerations which have been revisited by urban economics and NEG almost a hundred years later. Marshallian externalities will be discussed more in detail in the next section.

Meanwhile, during 1920's and and 1930's non-competitive theory of firms had been developed by leading scholars such as Hotelling (1929), Chamberlin (1933), Robinson (1933) (as cited in Fujita (2010)) and Kaldor (1935), which have been the precursors of new industrial organization theory based on non-competitive behaviour of firms in 1970's. But before that, non-competitive models have been applied to industrial location theory by German scholars, notably Christaller (1933) and Lösch (1940) which have been very influential thoughts in location theory [3]. The main question considered was how economies of scale and transport costs interact to produce a spatial economy. However, Fujita, Krugman, and Venables (2001) argue that these models are rather a description at best than an explanation of the economy's spatial structure due to the fact that they lack a sound economic modelling in which one finds an explanation on how a phenomenon emerges from the interaction of decisions by individuals. In their own words, "Lösch showed that a hexagonal lattice is efficient; he did not describe a decentralized process from which it might emerge. Christaller suggested the plausibility of a hierarchical structure; he gave no account of how individual actions would produce such a hierarchy (or even sustain one once it had been somehow created)" (Fujita, Krugman, and Venables, 2001, p. 27).

In the meantime, although spatial dimension of economic activity has been neglected by mainstream economics, some of the scholars have started to discuss the role of space on the distribution of economic activities from an alternative standpoint. The necessity for a general theory of location and space-economy that is fundamentally different from neoclassical general equilibrium framework based on perfect competition has been supported by Kaldor, Lösch, Isard, Koopmans and several others (Fujita, 2010). In the first place, Isard (1949) has offered

[3] For a detailed discussion on *central place theory* the reader may refer to Fujita, Krugman, and Venables (2001) and Fujita (2010).

powerful insights why competitive equilibrium paradigm could not be the right foundation for the space-economy and proposed general theory of monopolistic competition as the alternative. This insight has first been presented in a theoretical framework Koopmans and Beckmann (1957) suggesting that competitive pricing and positive transport costs are incompatible in a homogeneous spatial economy. Then, a definitive answer has been given by Starrett (1978) to competitive paradigm by extending it to a general equilibrium framework [4]. The work of Starrett has clearly shown the inability of competitive models to explain the endogenous formation of economic agglomeration [5]. Thus, Fujita (2010, p. 19) states that if one wants to get insights "about the spatial distribution of economic activities, in particular the formation of major economic agglomerations as well as regional specialization and trade, we must make at least one of the following three assumptions: (i) space is heterogeneous, (ii) externalities in production and consumption exist, or (iii) markets are imperfectly competitive". Based on these three assumptions, he classifies three modelling strategies which represent different agglomerative forces shaping economic space:

- **(A) Comparative advantage models:** These models focus on the economic outcomes of heterogeneity of space that introduces uneven distribution of immobile resources (mineral deposits, natural harbours, some production factors), amenities (climate) as well as the existence of transport nodes (transshipment points, ports). Under constant returns to scale and perfect competition, these heterogeneities generate comparative advantages among locations which in turn give rise to interregional or intercity specialization and trade (Fujita, 2010). Class A encompasses models such as monocentric city models of urban economics where central business districts arise and the Heckscher-Ohlin theory in which different endowments of production factors lead to international trade and specialization (Fujita and Thisse, 2009).

- **(B) Externality models:** Basic forces for spatial agglomeration and trade are generated endogenously through non-market interactions among firms in these models, unlike comparative advantage models. Non-market

[4] For an extensive theoretical modelling and discussion of competitive paradigm debate see chapter 2 in Fujita and Thisse (2002).

[5] Fujita and Thisse (2002) entitle this result *Spatial Impossibility Theorem*. It "implies that when space is homogeneous and transportation is costly, the only possible competitive equilibrium is the so-called backyard capitalism in which every location operates as an autarky. In turn, this is possible only when production activities are perfectly divisible. This clearly shows the limits of the competitive paradigm for studying the main features of actual spatial economies" (Fujita, 2010, p. 19).

interactions yield increasing returns external to a firm, such as knowledge spillovers, business communications, social interactions etc.. However, this approach still allows one to appeal to constant returns/perfect competition paradigm (Fujita, 2010). Moreover, "traditionally externalities have been treated in a "black box" manner that tends to hide the actual micro-interactions giving rise to such externalities" (Fujita and Thisse, 2009, p. 111). These models are mainly developed within the realm of urban economics.

- **(C) Imperfect competition models**: In these models, firms are no longer treated as price takers, rather their pricing policies depend on the spatial distribution of consumers and firms. Then, in turn some form of interdependencies arise between the location choices made by firms and households. Fujita (2010) further distinguishes these models by their approach to market competition as:
  - **(C1) Monopolistic competition models:** Firms are able to set their own prices and produce differentiated products under increasing returns to scale condition, unlike in competitive models. However, although they are price setters, strategic interaction among firms is not allowed since there are many. The models in this class is extensively developed in NEG framework.
  - **(C2) Oligopolistic competition models:** These models assume existence of a few large agents (firms, land developers, etc.) that strategically interact with each other by considering their market power. They take place within the realm of game theory and exemplified by spatial competition models in which a small number of firms compete for dispersed consumers Fujita and Thisse (2009). "Due to technical difficulties, most of the class C2 models developed so far are partial equilibrium models, leaving the advancement of class C2 general location models mostly for the future"(Fujita, 2010, p. 20).

It is interesting to note that the classification made by Fujita (2010) also follows a chronological order as well as reflecting the developments in modelling strategies. Group A models comprise traditional models of international trade in which location of economic activity is exogenously determined by differences in factor endowments and early urban economics models in 1960's where monocentricity of a city is a priori assumed. Due to the need for more general theory of urban locations, general local models of urban morphology has started to develop in early 1970's which sought to explain the geographical distribution of all agents in a given urban area without making a priori assumption about any center. And in line with the developments in industrial organization theory in late 1970, monopolistic competition models of class C1 has arisen beginning in late 1980's in urban economics and early 1990's in NEG (Fujita, 2010).

2.1.2 International Trade and Economic Geography

International economics might be expected to be treated as a special case of economic geography in which borders and the actions of governments play a crucial role in shaping the location of production. However, the analysis of international trade have not made use of insights from neither economic geography nor location theory. Traditional trade theory have treated countries as dimensionless points within which factors of production move instantly and without any cost from one activity to another, and moreover trade among countries takes place in a spaceless platform where transportation costs are zero for all traded goods (Krugman, 1991b).

By the mid 1970's, trade theory was based on the notion of comparative advantage which would result either due to technological differences (Ricardian models) or differences in factor endowments (Hecksher-Ohlin models). Countries were assumed to trade with each other over the goods in which they have comparative advantage. With this notion, it makes vague predictions for the location of economic activity in the sense that allowing for each location to specialize in the production of goods with comparative advantage. This would also hint on the idea that trade would take place between dissimilar countries in dissimilar goods. In fact, comparative advantage have explained clearly what was going on until that time. After World War II, especially after the major trade agreements of the 1950s and 1960s, the more puzzling trade patterns emerged, which Krugman (2009) calls this new phenomenon similar-similar trade. He shows evidence on the change in trade patterns by displaying the composition of British trade circa 1910 and in the 1990's. Britain, as a capital abundant country with scare land, used to export manufactured goods and import raw materials on the eve of World War I, and "the pattern of trade made perfect sense in terms of classical comparative advantage" (Krugman, 2009, p. 562). However, the situation in terms of trade patterns was not as easy as has been to explain by traditional trade theories where Britain exported and imported mainly manufactured goods in 1990's. Moreover, trade that has been restructured after World War II, has started to place between similar countries.

In fact, the case of similar-similar trade was not so incomprehensible as Balassa (1966) has given the directions of intra-industry trade in Europe as follows:

each country would produce only some part of the potential products it could produce within each industry importing those goods it did not produce because specialization in narrower ranges of machinery and intermediate products will permit the exploitation of economies of scale through the lengthening of production runs" (as cited in Krugman (2009, p. 562)). Even though these ideas were not unrealised or rejected on the ground of being incomprehensible, they had been ignored because they were associated with unexhausted economies of scale at the firm level which had a direct implication of imperfect competition. And at that time, there were not readily available *general equilibrium* models of imperfect competition where "trade theory, perhaps more than any other applied field of economics, is built around general equilibrium analysis"(Krugman, 2009, p. 563).

So the question of why from Ricardo until the 1980s there was an special emphasis on comparative advantage rather than increasing returns in explaining trade clarifies with the fact that comparative advantage could be modelled using models that assumed constant returns and competition, which were the tools at hand (Krugman, 1991b). In the meantime, mid-1970's witnessed a "new wave of theory in industrial organization which provided the economics profession with a menu of models of imperfect competition" (Krugman, 1991b, p. 6). The monopolistic competition models in the presence of increasing returns developed in the field of industrial organization, in particular the model developed by Dixit and Stiglitz (1977), has been applied to many fields in economics. The so-called *increasing returns revolution* first has been influential in international trade theory by the end of 1970's, and a few years later growth theorists applied the tool-box to economic growth theory where sustained growth arose from the presence of increasing returns. These two fields are widely known as "new trade theory" and "new growth theory" which Krugman (1998) calls the second and third wave of increasing returns revolution, and later the "new economic geography" will be named as the fourth [6].

Monopolistic competition models explored within the line of new trade theory provided a context for international trade that completely bypassed the conventional arguments based on comparative advantage. This context has been able to give way to trade between countries that were identical in resources and technology where they specialize in producing different products due to consumers' love of variety. Moreover, they were able to provide explanation for similar-similar trade

[6]The seminal and major articles related to *new trade theory* and *new growth theory* may be referred as Krugman (1979, 1980, 1981) and Romer (1986, 1987, 1990), respectively.

where "similar countries had little comparative advantage with respect to each other, so their trade was dominated by intra-industry trade caused by economies of scale" (Krugman, 2009, p. 564).

The main purpose here is far from surveying the existing literature on new trade theory, nor present its theoretical background which definitely is another research topic. Yet, it is important to bring into view what new trade theory came along with. Krugman (1979) with his path-breaking work not only clearly articulated this new revolutionary approach for international trade theory, but also planted the seeds of new economic geography where location of economic activity can be analysed within the framework of a general-equilibrium model (Committee, 2008).

In Krugman (1979)'s approach, economies of scale that are internal to the firm lies in the core which amounts to that firms are able to reduce their average costs by expanding production. The simple model in a closed economy setting produces a result such that the larger the economy the more variety of goods are produced which is the channel through which increasing returns operate. Due to consumers' taste for variety in the model, the consumers benefit from higher production in terms of increasing variety where firms exploit economies of scale by producing more. Then, one is able to compare "autarky" situation where there is no trade with the case of trade taking place between two countries. Assuming that countries have identical tastes and technologies (and factor endowment differences are ruled out by one-factor model), when trade takes place between two economies at zero transportation cost, both the scale of production and the range of goods available for consumption will increase [7]. So, welfare in both countries will increase, both due to higher real wages and increased choice. With this analysis he shows that economies of scale may give rise to trade and gains from trade even in the absence of differences in tastes, technology, or factor endowments. However, the direction of trade is indeterminate in this model, that's to say which country exports which goods is not known, it can only be said that each good will be produced in only one country (Krugman, 1979).

Building on Krugman's approach, an extensive literature has developed in this field exploring the implications of increasing returns and monopolistic competition on international trade. However, one of the important contributions has come from Krugman (1980) where he extended his 1979 model by incorporating *transportation costs* and *home market effect* into it. There has been a significant

[7] Due to the symmetry in two economies, wage rates will be equal and the price of any good produced in either country will be the same.

decline in transportation costs during 19th century which contributed largely to the growth of trade and it was not accounted for in trade models till then. Including transportation costs, then allowed for adapting home market effect into trade models analytically. Home market effect basically arises when transport costs are explicitly considered, imperfectly competitive industries tend to concentrate their production in their larger markets and to export to smaller ones (Ottaviano and Thisse, 2004). Krugman (2009, p. 565) explains the basic intuition behind transportation costs and home market effect as follows:

> *"Increasing returns provide an incentive to concentrate production of any one product in a single location; given this incentive to concentrate, transport costs are minimized by choosing a location close to the largest market, and this location then exports to other markets".*

While the work of Krugman (1979, 1980) has had an immediate impact on the trade literature and provoked further research, it would take more than ten years for his approach to have an important influence on the economic geography literature which will be mentioned in the following sections [8]. He explains in his own words how his interest evolved from international economics to economic geography as follows:

> *"As I explained in Krugman (1991), I initially thought that some interesting things about the increasing factor mobility might be said from my own perspective on international trade. As I worked on the subject, however, I found that my analysis was drifting further and further away from international economics as I knew it. In international economics, we take as our base case a world in which resources are completely immobile but in which goods can be costlessly traded. What I found myself gravitating towards was a style of model in which factors of production were perfectly mobile but in which there were costs to transporting goods. In other words, I found myself doing something closer to classical location theory than to international trade theory."*(Fujita and Krugman, 2004, p. 151).

2.2 Sources of Agglomeration

As mentioned in the previous section, agglomeration phenomenon has been in the research agenda of many scholars from different disciplines for a long time. The common purpose of these disciplines may be stated as the effort they pay to

[8]The final section of Krugman (1979) on migration and agglomeration and incorporation of transportation costs and home market effect in Krugman (1980) basically form the nucleus of, which later will be called, NEG.

understand the reasons behind the tendency of economic activities to agglomerate. Although they differ in terms of their theoretical frameworks and the way they approach to the subject, clustering phenomenon eventually derives from internal or external economies of scale (Karlsson, 2008). From this point view, one may present generally accepted sources of agglomeration as natural advantages, externalities, internal increasing returns and transportation costs.

2.2.1 Natural Advantages

Natural advantages refer to exogenously given characteristics of different locations, such as climatic conditions, availability of raw materials, proximity to natural harbours, etc. Prior to industrialization process, the distribution of economic activity has been determined by the distribution of land available for agricultural production (Kim, 1999). Besides emphasizing the importance of externalities, Marshall (1890) has also identified natural advantages as one of the main causes of geographic concentration of production.

> "*Many various causes have led to the localization of industries; but the chief causes have been physical conditions; such as the character of the climate and the soil, the existence of mines and quarries in the neighbourhood, or within easy access by land or water. Thus metallic industries have generally been either near mines or in places where fuel was cheap. The iron industries in England first sought those districts in which charcoal was plentiful, and afterwards they went to neighbourhood of collieries. Staffordshire makes many kinds of pottery, all the materials of which are imported from a long distance; but she has cheap coal and excellent clay for making the heavy saggars or boxes in which the pottery is placed while being fired. Straw plaiting has its chief home in Bedfordshire, where straw has just the right proportion of silex to give strength without brittleness; and Buckinghamshire beeches have afforded the material for the Wycombe chair making. The Sheffield cutlery trade is due chiefly to the excellent grit of which its grindstones are made*"(Marshall, 1890, p. 268-269)

Also known as 'first nature', as dubbed by Krugman (1993), is certainly "important to explain the location of heavy industries during the Industrial Revolution, because the proximity of raw materials was a critical factor"(Ottaviano and Thisse, 2004, p. 2565). However, it fails to provide reasonable explanation for other forms of agglomeration, for instance Silicon Valley type, which have nothing to do with natural advantages. Nevertheless, Ellison and Glaeser (1997) state that even if natural advantages as a reason for geographic concentration may not seem exciting, still it explains some of the observed agglomerations, for instance importance of climate for wine industry, proximity to coasts for shipbuilding industry etc. Although they don't differentiate between natural advantages and spillovers in their 1997 model, they consider only natural cost advantages as a reason for agglomeration in Ellison and Glaeser (1999) and conclude that 20 % of of measured geographic concentration in U.S. can be attributed to a few observable natural advantages. Kim (1999) also attempts to differentiate between natural advantages and spillovers causing geographic concentration by controlling for factor endowments and shows that factor endowments explain a large amount of the geographic variation in U.S. manufacturing over time. These studies show that, although natural advantages are not solely capable of capturing the incentives behind agglomeration, they do account for some clustering behaviour and better not to neglect them by deeming as obsolete.

2.2.2 Externalities

Theories of agglomeration have extensively utilised the notion of scale economies as principal factors explaining the spatial agglomeration of firms. Internal economies of scale are associated with production conditions of a single firm, while external economies are independent of a single firm but accrue to all firms located in the same area. Internal increasing returns are placed at the core of NEG which will be discussed in the next section. On the other hand, external economies generally have been used in modelling agglomeration by urban economists, regional scientists, geographers and even by management scholars. Before going deeper it is important to remark that external economies, or shortly *externalities*, are pure in the sense that which are external to an individual firm but internal to the industry. Externalities handled by this manner allows one to work within the realm of perfect competition providing a convenient framework for modelling. Furthermore, external economy models, if appropriately designed, are able to yield same agglomeration outcomes as monopolistic competition models. However, the shortcoming of these models arise as the vague description of the sources of external economies which in turn leads the spatial extent to be determined exogenously in an ad hoc manner (Fujita and Krugman, 2004).

Setting theoretical modelling discussion aside, the contribution of the concept of external economies in the form of localization economies is attributed to Marshall (1890) where identified three distinct reasons for localization: *labour market pooling*, *input sharing* and *knowledge spillovers*.

Labour market pooling: First, concentration of a number of firms in an industry in same location allows workers a pooled market with specialized skills which benefits both workers and firms:

"*When an industry has thus chosen a location for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighbourhood to one another...A localized industry gains a great advantage from the fact that it offers a constant market for skill... Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market. The owner of an isolated factory, even if he has good access to a plentiful supply of general labour, is often put to great shifts for want of some special skilled labour; and a skilled workman, when thrown out of employment in it, has no easy refuge.*"(Marshall, 1890, p. 270)

This agglomerative force defined by Marshall (1890) is known as *labour market pooling*. Krugman (1991b) clarifies the nature of the gains from labour market pooling with a trivial example. Assume that there are two locations and only two firms in an industry, each of which can produce in either of only two locations. They use the same distinctive kind of skilled labour in their production, however their demands for labour are not perfectly correlated for some reason, may be due to producing differentiated products that face uncertain demand or being subject to firm specific production shocks. In the end firms' demand for labour is imperfectly correlated and uncertain. For the sake of concreteness, now think that each firm face both good and bad economic conditions during their production where in the former case they employ more and in the latter less of specialized workers. Suppose that it employs 125 workers in good and 75 workers in bad conditions, also assume that there 200 workers with this specialized skills thus average demand for labour equals supply. At this point, the crucial question arises as: Will both firms and workers be better off if two firms choose different locations each forming a company town with a local labour force of 100 workers or if they choose the same location with a pooled labour force of 200 workers that can work in either firm? When we consider the case from firms' point of view, they will be better of by locating in the same location, such that they will be able to hire more during good conditions taking advantage of the pooled labour. Otherwise, if they had located separately, they would have had to content with the existing local labour force which does not suffice to meet labour demand in good times resulting in an excess demand for labour. If both firms located in the same location, however, then at least occasionally one firm's good conditions would coincide with other's bad conditions and additional workers would be available. Considering the case from workers' side gives the same result as in good times they have more opportunity to be employed especially when one firm's bad conditions will be offset by the other firm's good conditions.

Krugman (1991b) argues that uncertainty alone would not suffice to generate localization. In the previous example it was a necessary condition for each firm to locate in only one location to take advantage of labour market pooling. However, if divisibility is assumed in production, such that each firm would split into two identical firms, then the same pattern would be replicated in both locations and the motivation for localization would be gone. So, in order to make the assumption that both firms choose the same location, at least some form of indivisibilities should exist in production. If there are sufficient economies of scale in production a single production site emerges. In this sense, he fills the gap in Marshall's labour pooling argument by emphasizing the role of increasing returns. Yet, he claims that this is not a description of the process that might bring about concentration, rather only an argument for the advantage of concentrated production. Following that, he formalizes the labour market pooling argument with the help of a phase diagram in which horizontal axis shows the West's share of workers, vertical axis shows West's share of firms and two curves depicting which distributions of firms and workers will leave the typical firm and worker respectively indifferent between the two locations [9]. With this rough formalization he shows that three equilibria emerges, one of which is knife-edge unstable. Depending on the initial conditions there arise a converge towards concentration of both firms and workers either in East or West.

Input sharing: The second Marshallian force for agglomeration is the provision of non-traded inputs specific to an industry in greater variety and at lower cost. As Marshall (1890) verbalises:

> "*Subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material...the economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind, even though no individual capital employed in the trade be very large. For subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbours, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expenses...*"(Marshall, 1890, p. 270)

The advantage of input sharing is straightforward such that "a localized industry can support more specialized local suppliers, which in turn makes that industry more efficient and reinforces the localization"Krugman (1991b, p. 49). The notion of input sharing also crucially depends on the existence of scale economies

[9]For details about the phase diagram analysis and further thoughts on labour pooling regarding discussions about labour market clearing conditions and monopsony power for firms, the reader may refer to Krugman (1991b, p. 41-48).

in input production. In the absence of scale economies in input production even a small scale of production would be as efficient as a large one. A large center of production is able to have more efficient and more diverse suppliers than a small one only in the presence of increasing returns (Krugman, 1991b).

Similar to the example in labour pooling argument, Krugman (1991b) provides a trivial example to illustrate how input sharing stimulates agglomeration. Assume that there is a variety of products each of which is demanded as a final good and as an input in the production of other goods. For the sake of simplicity it is supposed that intermediate goods and final goods are the same. In order to concretize assume that the typical product within these varieties has total sales of 10 units, but that of 4 are sold to manufacturers of other varieties and accordingly it requires 4 units of intermediate inputs which are drawn from the same industry producing other varieties in order to produce that 10 units. Further assume that there are two locations possible for production, East and West, and each own these locations also include half of the final demand, that is 3 units of each variety of product. So given these conditions where would a firm choose to locate? The answer depends on the location choice of other firms. If all other firms are located in the West, then 7 (3 final plus 4 intermediate) of the 10 units of total demand for a particular product that the firm produces will come from the West. This will stimulate a firm to locate its production in West as well because all of the firm's supplies of intermediate goods will come from West and locating there will provide a firm transportation cost advantage. Thus both forward and backward linkages will create an incentive to concentrate production.

Based on the Marshall's insight on input sharing, Venables (1996) presents a formal general equilibrium model where agglomeration is generated by interaction between the location decisions of vertically integrated firms in industries that are linked through input-output structure. Downstream industry forms the market for upstream industry. Upstream industry considering market access are then attracted to areas where there are many downstream firms, which is a demand linkage. Apart from this, there is also a cost linkage such that firms in the downstream industry will face lower costs if they locate where there are many upstream firms since they will be saving in transportation costs in intermediate goods. So, both cost and demand linkages together will act as centripetal forces for the agglomeration of activity. On the other hand, location of immobile factors of production and final consumer demand will act as centrifugal forces working against agglomeration. Characteristics of the industry, especially the strength of vertical linkages between industries and the cost of transportation between locations will determine the balance centripetal and centrifugal forces which will in turn determine the extent of agglomeration. As a result unique or multiple equilibria may arise which comprises either dispersed production or the concentration of production at a single location [10]. It is important to highlight that linkages derive their effect from the interaction of trade cost with increasing returns to scale and imperfect competition.

Knowledge spillovers: Third and most compelling Marshallian externality explaining agglomeration is *knowledge spillovers*. He asserts the benefits of knowledge spillovers as follows:

> "*The mysteries of the trade become no mystery; but are as it were in the air.... Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas*"(Marshall, 1890, p. 270)

Knowledge spillovers emanates from non-market interactions such as imitation, business interactions, face to face communication, inter-firm circulation of skilled labour without involving any monetary transaction. These type of interactions paves the way for transmission and exchange of knowledge, ideas, information, products and processes. Since the firm that creates the new knowledge can not fully appropriate it, this knowledge spills over to other firms affecting their innovativeness positively. Also it contributes to the stock of knowledge available for each firm in the industry. Knowledge that spills over is tacit by definition, it is uncodified and can only be acquired via social interactions, hence distance matters. As they are geographically bounded to regions in which the new knowledge is created, geographic proximity becomes essential and creates an incentive for firms to locate in these regions (Beaudry and Schiffauerova, 2009). Urban economics literature widely explores knowledge spillovers as a source of urban agglomerations [11] which is commonly known as *MAR (Marshall-Arrow-Romer)*

[10]Krugman and Venables (1995) assume that the upstream and downstream industries are really the same to simplify the basic insight of input-output structure between firms. That means the same goods are consumed and used as inputs to production of other goods. In international economics framework they present a formal of industry concentration where the world is divided into two parts North and East and show that gradual process of expanding world trade as a result of falling transportation costs cause the world to divide into a high-wage, industrialised North and a low-wage, primary-producing South. Later, as transportation costs continue falling South rises again at the expense of North

[11]For a theoretical survey on the micro-foundations of urban agglomeration economies the reader may refer to Rosenthal and Strange (2004). Contrary to taxonomy made previously in Marshallian externalities, they distinguish theories by the mechanism driving them and identify

externalities as later formalized by Glaeser, Kallal, Scheinkman, and Shleifer (1992).

At this point it would be useful to make a distinction between three types of externalities asserted by Marshall. First two type of externalities discussed above (labour pooling and input sharing) as a source of agglomeration are *pecuniary externalities* which refer to "the benefits of economic interactions which take place through usual market mechanisms via the mediation of prices" (Fujita and Thisse, 1996, p. 345). On the other hand, *technological externalities* arise through non-market interactions via processes affecting the utility function of individuals or production function of firms [12]. Generally, a wide range of theoretical models in urban economics are based on technological externalities in the form of knowledge spillovers, business communications, face-to-face communication, and other spatial externalities. Treating externalities arising from non-market interactions as external to the firm but internal to the industry allows these models to work under constant returns to scale and perfect competition, which is likely to explain why these models appeared long before the development of NEG (Fujita and Thisse, 2009).

However, new economic geographers, who perceive technological externalities as *knowledge spillovers* or *pure Marshallian externalities*, have been rather shy on this topic. Theoretical NEG models are solidly based on pecuniary externalities where forward and backward linkages (through input-output structure as in Krugman and Venables (1995) or between firms and workers/customers as in Krugman (1991a)) in conjunction with increasing returns and transportation costs are able to generate agglomeration endogenously without making further a priori assumptions about location of economic activity. Fujita and Thisse (2009) describe technological externalities as "black box" that hide the actual micro-interactions causing such externalities to arise. In Fujita and Krugman (2004, p. 160) Krugman notes that: "This is not because I do not agree with the potential importance of such agglomeration forces, but because I could not find any solid micro-model of knowledge spillovers or communications". And they summarize the disadvantages of pure externality models as follows:

"*In particular, the sources of external economies are vague, at best. When norma-*

---

three types of micro-foundations, based on sharing, matching, and learning mechanisms. First presenting a core urban model under each category they discuss the literature in relation to those models.

[12]The original discussion about the nature of externalities has been provided by Scitovsky (1954) where he distinguished externalities as *pecuniary* and *technological*.

tive or policy questions are addressed, we need to know more precisely the nature of external economies. Furthermore, since the underlying mechanisms of external economies are not clear, their spatial extent can be specified only exogenously in an ad hoc manner. Even when the spatial process of external economies is well-specified, the essential details regarding the information/knowledge externalities are often missing. For example, in communication externality models of urban morphology (Fujita and Thisse 2002, Chapter 6), although the communication process is well specified, it is not clear what information is exchanged and how it is utilised by firms. Furthermore, the nature of information/knowledge externalities is essentially dynamic, and hence their full-fledged treatment requires a dynamic framework "(Fujita and Krugman, 2004, p. 160).

Now it is time to consider the fundamental pillars of NEG and its remarks on the agglomeration phenomenon.

2.2.3 Increasing Returns and Transportation Costs

As discussed in previous section agglomerations occur through both technological and pecuniary externalities. Even though Marshall has not provided the channels through which agglomeration arises, the urban economics literature put a special emphasis on external increasing returns as a requisite for agglomeration. However, on the one hand working within the framework of external scale economies enables one to deal with constant returns to scale and competitive markets, on the other hand it requires the location of economic activity to be specified exogenously since it can not be derived from the model. The main line of NEG is "how to explain the formation of a large variety of economic agglomeration in geographical space" (Fujita and Krugman, 2004, p. 140). Geographic concentration of economic activities is the outcome of two opposing forces, *centripetal* forces that tend to pull economic activity together and the *centrifugal* forces that tend to push it apart. NEG demonstrates how the geographical structure of the economy is shaped by the tension between these forces by providing them micro founded explanations in a full-fledged general equilibrium setting.

Table 2.1: **Forces Affecting Geographical Concentration**

Centripetal forces	**Centrifugal forces**
Market-size effects (linkages)	Immobile factors
Thick labour markets	Land rents
Pure external economies	Pure external dis-economies

Source: Krugman (1998)

Centripetal forces listed in the first column of Table 2.1 are in fact three Marshallian sources of external economies discussed in the previous section. A large local market creates both backward linkages (i.e. locations with better access to large markets are preferred for production subject to scale economies) and forward linkages (i.e. a large local market promotes the local production of intermediate goods by lowering the costs of downstream firms). A thick labour market, especially for specialized skills, allows firms to access workers easier and vice versa. And an industrial concentration may create pure external economies via knowledge spillovers. *Centrifugal* forces are listed on the second column. Immobile factors, which involve land, natural resources and, in an international context, workers, preclude concentration of production both from supply side (some production locates in where the workers are) and demand side (some production locates close to consumers due to dispersed markets). Land rents tend to increase due to the concentration of economic activity since demand for local land increases, and accordingly this creates a disincentive for further concentration. Also, concentration of production is likely to create pure external diseconomies such as congestion (Krugman, 1998).

Without doubt, real world agglomeration phenomenon arises out of interactions between all these forces but NEG focuses on the first item of each column in Table 2.1 to conduct analytical work on economic geography. The choice of centripetal and centrifugal forces, namely linkages and immobile factors, definitely reflects concrete modelling concerns [13]. The modelling strategy of NEG allows for an "approach that concentrates on the role of market-size effects in generating linkages that foster geographical concentration, on one side, and the opposing force of immobile factors working against such concentration on the other" (Krugman, 1998, p. 9).

The seminal work of Krugman (1991a) introduces a framework that demonstrates the interactions among increasing returns at the firm level, transport costs and factor mobility give rise to spatial economic structure to emerge and change. The existence of increasing returns at the firm level is a crucial factor in explaining agglomeration which assures economy not to give way to so called 'backyard capitalism' [14]. The indivisibilities existent in production motive firms to concentrate production in a small number of plants. It is more profitable for firms to produce on a large scale in a few places and trade its goods to dispersed con-

[13]For an elaborative discussion on modelling strategy and principles of NEG the reader may refer to Krugman (1998, 1999).

[14]The assumption of non-increasing returns has dramatic implications for economic geographers. Under this assumption, coupled with the uniform distribution of resources, the economy reduces into Robinson Crusoe economy where each individual produces for herself/himself which is called *backyard capitalism*. Without recognizing indivisibilities, each location operates as an autarky where goods are produced on a small scale sufficient to meet local demand. Trade possibly occurs between locations if the distribution of resources is not uniform (as in traditional trade theory). It is obvious that unequal distribution of resources is an insufficient explanation for understanding geographic concentration phenomenon (Fujita and Thisse, 1996).

sumers which allows them to avoid fixed costs that would arise in the case of dispersed production. So, increasing returns constitute the centripetal force for agglomeration. However, it is important to note that geographic extension of markets prevents production to concentrate only in a single place because transportation is costly. Hence, spatial dispersion of demand acts as a centrifugal force. Therefore a trade-off arises between increasing returns and transportation costs for the firms to consider whether to concentrate production or not (Fujita and Thisse, 1996). Furthermore factor mobility will have a reinforcing effect on the concentration of production. To gain the basic insight about how agglomeration emerges out of the interaction among these factors, the *core-periphery* model of Krugman (1991a) will be mentioned in general terms [15].

The basic framework of the model introduced by Krugman (1991a) can be described as follows. There are two regions, two sectors (agriculture and manufacturing) and two types of labour (farmers and workers). The manufacturing sector produces a continuum of horizontally differentiated products where each firm produces a different variety of product under increasing returns to scale using workers as the only input [16]. On the other side, agricultural sector produces a homogeneous product under constant returns with farmers used as the only input for production. Workers are freely mobile between two regions while farmers are immobile distributed equally between two regions. Finally, the agricultural good is traded costlessly between two region whereas the interregional trade of manufacturing goods involves a positive transport cost in the form of *iceberg* [17] (Fujita and Krugman, 2004).

In this model, centrifugal forces are created by immobility of farmers as they consume both goods and firms may choose to locate closer to them to meet the demand for manufactured goods. Centripetal forces are more complex and involves a circular causation as shown in Figure 2.1. First, if a large number of firms are located in a region 1, a greater number of varieties are produced. Workers (who are also consumers) will have a better access to a greater number of

[15] For a formal presentation of the model the reader may refer to the original paper.

[16] This feature derives due to employing monopolistic competition model introduced by Dixit and Stiglitz (1977) which assumes a continuum of goods that allows the modeller to "respect the integer nature of many location decisions - no fractional plants allowed - yet analyse their models in terms of the behaviour of continuous variables such as the share of manufacturing in a particular region" (Krugman, 1999, p. 146).

[17] *Iceberg* transport costs are first introduced by Samuelson (1954) which asserts that "a fraction of any good shipped simply 'melts away' in transit, so that transport costs are in effect incurred in the good shipped"(Krugman, 1999, p. 146). Modelling transportation costs this way allows one to avoid modelling transportation sector as a separate one in a general equilibrium framework.

varieties compared to workers in region 2. In addition to this love of variety, there is also price index effect at work. The equilibrium price index of manufactured goods faced by consumers will be lower in region 1 because firms do not undertake price discrimination between regions. Thus, lower price index will generate a real income effect for workers in region 1 which will induce more workers to migrate there from region 2. An increase in the number of workers, who are also consumers, in region 1 triggers the concentration process as follows. The resulting increase in the number of workers (consumers) created a larger demand for manufactured goods in region 1 which lead more firms to locate there via home market effect. Because of scale economies there is an incentive to concentrate production of each variety in only one region and because of transportation costs it is more profitable to produce in the region offering a larger market and ship to the other. This implies the production and availability of even more varieties of differentiated product in region 1, where the cycle starts again. Briefly, a circular causation for agglomeration of firms and workers is generated through *forward linkages* (workers' motive to be close to the producers of consumer goods) and *backward linkages* (producers' motive to concentrate where the market is larger) (Fujita and Thisse (1996), Fujita and Krugman (2004)).

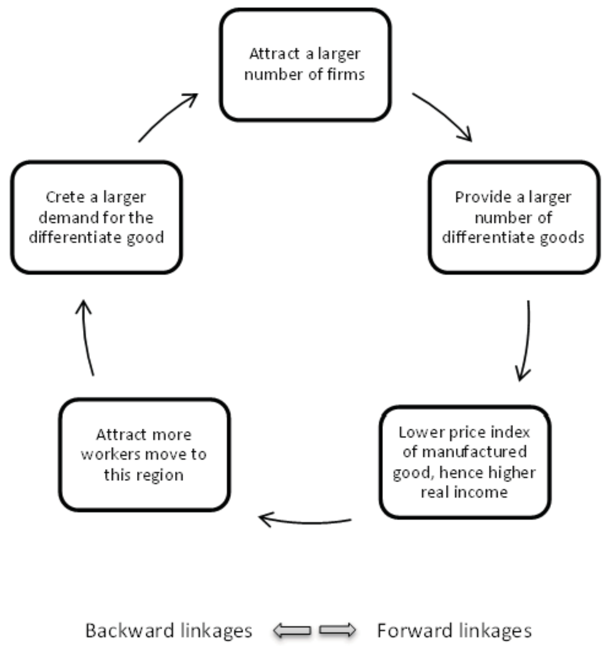


Figure 2.1: **Circular Causation Created by Centripetal Forces**
Source: Author's illustration based on Krugman (1991a)

Krugman (1991a) shows that, if forward and backward linkages are strong enough to surpass centrifugal forces generated by immobile farmers, the economy will end up with *core-periphery* pattern in which all manufacturing is concentrated in only one region. The emergence of *core-periphery* pattern is more likely (i) when transportation costs for manufacturing is low enough, (ii) when varieties of products are sufficiently differentiated and (iii) when the share of manufacturing sector in the economy is large enough.

Agglomeration may or not occur due to the parameter values of the model. However, a small change in critical parameters may move the economy from one in which two regions are symmetric to the one in which initial advantages cumulate and transform one region into an industrial core and the other into de-industrialized periphery. That means the dynamics of the model are subject to *catastrophic bifurcations* which refers to points at which their qualitative character suddenly changes (Fujita and Krugman, 2004).

Based on the same modelling architecture and modelling tricks, NEG literature have developed regional models, urban system models and international models which are discussed in detail in Fujita, Krugman, and Venables (2001). All these models show the emergence of agglomeration phenomenon in a general equilibrium framework without making a priori assumptions. Although NEG models have opened up a new way of theorizing agglomeration phenomenon, they are still subject to developments and experiencing advances progressively. This new field have been respected by other disciplines such as urban economics and regional science although they use different approaches and the search for a unifying mechanism between NEG and these disciplines began [18].

Given the short overview of theoretical foundations of agglomeration economies, it is apparent that there had been a lot to say and looks like still is. Moving away from theoretical discussions, the next chapter deals with the empirical side which reviews the quantitative literature in measuring agglomeration phenomenon.

[18] Behrens and Thisse (2007) show that concepts and tools developed in NEG may be used to revisit several problems in regional economics.

3 MEASURING DISTRIBUTION OF ECONOMIC ACTIVITY

Theoretical discussions regarding spatial distribution of economic activities are extensive, as indicated in the previous section. Beyond theoretical discussions, how to measure agglomeration at different spatial scales arises as an important question and hence another interest of research within the field. The literature offers a variety of measures to quantify the distribution of economic activity across different geographic units. This chapter reviews most common measures used in the literature within a hierarchical framework. The presentation starts with discrete measures of concentration followed by more recent continuous measures. Each concentration measure is discussed in conjunction with recent developments in the literature.

Given a wide range of concentration measures, any of them may be criticized on the grounds of its ability to quantify agglomeration correctly. Most of the early measures are not capable of disentangling agglomeration from random concentration. Say that this problem is tackled by model based indices which takes into account this, they inherently have other problems such as modifiable areal problem, etc. which will be discussed later. In this respect, Combes and Overman (2004) presented seven criteria that a satisfactory spatial concentration index should meet:

- (C1) Measures should be comparable across activities
- (C2) Measures should be comparable across spatial scales
- (C3) Measures should take a unique value under the null hypothesis of random location choice. The alternative would be identified by the theory.
- (C4) Measures should be reported by their statistical significance where the index is derived from a probabilistic model.
- (C5) Measures should be unbiased with respect to arbitrary changes to the spatial classification
- (C6) Measures should be unbiased with respect to arbitrary changes to the industrial classification

- (C7) Measures should consider as well an alternative hypothesis to the null, which is based on a theory that states the forces behind the systematic location patters.

The measures of geographic concentration to be discussed in this chapter will be assessed by referring to these criteria from time to time. They will be considered within a framework which allows one to find out developments in the field throughout time. The next section presents some common measures utilised in the literature which involves location quotient, Hoover coefficient of localization, Krugman locational Gini, Herfindahl index, Ellison-Glaeser index and Maurel-Sedillot index. Following discrete indices, most common continuous indices, namely K-function and K-density, will be mentioned.

3.1 Discrete Measures

Discrete measures are 'discrete' in the sense that they partition the geographic area in question into discrete spatial units such as, states, counties, provinces, zip-code areas, etc. to compute agglomeration measure. They range from basic shares to model based indices while any of them aggregates data into predefined geographic units. This aggregation issue, although provides ease of computation, conveys some inherent problems such as modifiable areal problem and checkboard problem as will be discussed in detail in this section. Discussion starts with the most basic measure location quotient.

3.1.1 Location Quotient

Location quotient may be regarded as one of the earlier measures often explored to quantify industrial concentration in regions which is first introduced by Florence (1939))(as cited in Guimarães, Figueiredo, and Woodward (2009)). Essentially, it computes the ratio between the regional employment share of a particular industry and the employment share of that industry in a wider area, such as country. Simply, how much industry i is concentrated in location m is computed as follows:

$$LQ_{im} = \frac{E_{im}/E_m}{E_i/E} \tag{3.1}$$

It shows the location quotient of industry i in region m where $i = 1, 2, ..., I$ and $m = 1, 2, ..., M$. Further, the terms in the fraction denote: E_{im}, regional employment of industry i in region m; E_m, regional total employment of all industries ($E_m = \sum_i^I E_{im}$); E_i, employment of industry i in the entire country ($E_i = \sum_m^M E_{im}$); and E, total employment of all industries in entire country ($E = \sum_m^M \sum_i^I E_{im}$). Rearranging the terms in equation (3.1) allows us to express it in terms of regional specialization of location m in industrial activity i.

$$LQ_{im} = \frac{E_{im}/E_i}{E_m/E} \tag{3.2}$$

If $LQ_{im} = 1$ is obtained, then the employment distribution of industry i in region m follows the same pattern with the total industrial employment distribution across country. In other words, regarding equation (3.1), a particular industry's share within a region is the same as that industry's share within the country, hence there is no room for industrial concentration. Similarly, referring equation (3.2), the region's share in a specific industry employment is observed to be the same as region's share in total industrial employment. Hence, it's concluded that there is no regional specialization on a particular industry. Values above 1, $LQ_{im} > 1$, indicate that "the location (industry) is relatively specialized (concentrated) in the industry (location), as it has relatively less employment than it would be predicted based on its aggregate employment share" (Lafourcade and Mion, 2003, p. 4).

In general, *location quotients* are widely used in cross-country empirical studies for measuring relative industrial concentration and relative regional specialisation patterns. It might be conceived as a dissimilarity indicator exactly related to the concept of comparative advantage (Cutrini, 2006). It is basically a variant of Balassa index which is widely used in international trade literature [1]. Holmes and Stevens (2004b) use LQs to display the pattern of regional specialization in North America, while Akgüngör (2003, 2006) explore them to identify regional highpoint clusters in Turkey's regions. While some studies employ LQ as it is, it is rather used as a base for constructing new indices such as Hoover coefficient of localization, Krugman locational Gini, Krugman specialization and concentration indices, etc. which will be discussed in the following section.

Although LQ index widely explored in regional studies due to computational simplicity and low data requirements, it is criticized for on several grounds. First

[1] Balassa (1965) index is first and most widely used to measure revealed comparative advantage built on export shares (as cited in De Benedictis and Tamberi (2004)). Basically it measures the sectoral relative export of a country in terms of share of world exports and expressed as $BI = \frac{X_{ij}/X_{it}}{X_{nj}/X_{nt}}$, where X denotes exports, i denotes country, j is a commodity (or industry), t is a set of commodities (or industries) and n is a set of countries. If $BI > 1$, comparative advantage of country i is 'revealed'.

one is related to the cut-off values. An industry is said to be 'over-represented' within an area if $LQ > 1$, and 'under-represented' if $LQ < 1$. "Areas with high levels of 'over-represented' industries are often held to constitute clusters because they have an above average concentration of of employment in that industry" (O'Donoghue and Gleave, 2004, p. 421). However, how large an LQ should be to give an indication of clustering remains an unanswered question. There the main limitation for using LQs is remarked as the absence of commonly agreed or theoretical LQ cut-off values for defining a cluster (Martin and Sunley, 2003). Another limitation is that results are sensitive to the chosen level of spatial aggregation. Outcomes display different cluster patterns at different spatial scales. Third but not least, it does not provide any information on whether concentration in an industry is due to the presence of large number of small plants or only one large plant hiring the same amount of labor in the region (Martin and Sunley, 2003). While the first case leaves some room for externalities arising from co-location of firms, the latter suggests internal economies of scale. In addition, LQ based indices inhere generic problems as well as LQs which will be discussed more in detail in the next section having regard to the criteria presented by Combes and Overman (2004).

A number of studies proposed the use of statistical tests for LQs to get statistically significant cut-off values rather than arbitrarily defined to identify clusters. O'Donoghue and Gleave (2004) propose an approach to compute *standardized location quotients* (SLQ) which is simply the z-statistic of the computed LQs, $SLQ_{im} = \dfrac{LQ_{im} - \overline{LQ_i}}{std(LQ_i)}$, where prior to standardization procedure they are subject to a test of normality. If the test fails to confirm the normal distribution, logarithmic transformation of LQs followed by another normality test is suggested. Having passed the normality test signifies that SLQs fit the standard normal distribution, then the critical values for the standard normal distribution 1.96 for a two-tailed test or 1.64 for a one-tailed test at 5 % significance level are used as cut-off values. However, if the normality assumption is invalid, the authors do not recommend applying this approach by denoting it as it's limitation. Billings and Johnson (2012) construct the LQ from a discrete data generating process (Binomial and Poisson)and formally test its statistical properties. They show that LQ is typically unbiased but displays a small sample bias when Poisson distribution is assumed. They also determine the accuracy of the statistical tests under these distributions and disclose that under any distribution the critical values are problematic. Findings indicate that when industrial and/or spatial aggregation is coarsened or significance level is decreased, the accuracy of the statistical tests improved. In an earlier work but in a similar fashion Moineddin, Beyene, and Boyle (2003) derive a closed-form expression for calculating the standard deviation of individual LQ and construct confidence intervals based on the estimated standard deviations with a normal distribution assumption of LQs. First they assess the performance of the proposed model by simulation, then it is illustrated by exploring health utilization data. In the exactly same manner Beyene and Moineddin (2005) maintain the previous approach by incorporating different statistical methods and show that different analytical methods produce very similar confidence limits for location quotients. One common shortcoming of these studies lies behind the assumptions on the statistical distributions of the location quotient. If distributional assumption does not hold, the test statistics are not reliable hence the results obtained are nonsense. Tian (2013) propose a bootstrap method that is independent of the statistical distribution assumption. This method asserts that "the empirical distribution function of a random variable X, which can be obtained by bootstrapping, consistently estimates the true cumulative distribution function of X" (Tian, 2013, p. 189). The test statistics obtained from the empirical distribution are consistent estimates of the true distribution as well. Bootstrap method allows one to relax distributional assumptions and obtain cut-off values independently. Another advantage of this method and hence the novelty in this study is that the cut-off value for each industry is unique as the bootstrap method is applied to each industry individually. Assuming the uniformity in data generating process of each industry is nonsense because they are likely to be determined by industry-specific characters. The study also shows that cut-off values significantly vary across industries, and this fact enables one to detect agglomeration more accurately than in the case of using conventional z-scores.

Despite the progress in the statistical properties of the location quotients made by aforementioned studies, they lacked a theoretically based statistical foundation. Guimarães, Figueiredo, and Woodward (2009) have been the first to link location quotients to an existing theoretical model of location choice. They provide a theoretical legitimacy to LQs by deriving it as an estimator from a probabilistic location model based on Ellison and Glaeser (1997) dartboard approach. Once this theoretical basis is formed, they build statistical tests which provide important information about the accuracy of the LQ.

3.1.2 Hoover (1936) Coefficient of Localization

As mentioned above, location quotients are rather used as a base for constructing other measures of industrial concentration and regional specialization.

Hoover (1936) may be regarded as one of the pioneer studies dealing with the measurement of geographic distribution of industries. He defines *localization* of industries as "the degree of dissimilarity between the geographical distribution of an industry and that of population"(Hoover, 1936, p. 162). Regional employment percentage of an industry,(E_{im}/E_{itot}), is compared with regional population percentage, (pop_m/pop_{tot}) , and it is concluded that if an industry is not localized at all it should be distributed in exactly the same pattern as population. This is precisely the same idea behind location quotients, as mentioned above, even he does not entitle so. The slope coefficient of the line obtained, $\dfrac{E_{im}/E_{itot}}{pop_m/pop_{tot}}$, when plotting regional population shares (x-axis) against regional industry employment shares (y-axis) exactly coincides with what location quotient provides. In the case of no localization at all, we expect the slope of the line overlap 45 degrees.

The measure proposed is called *localization curve* whose concavity indicates the extent of localization. Hoover's *localization curve* is analogous to Lorenz curve which measures the degree of concentration of income between individuals. This measure of dissimilarity is first applied to the geographical context by Hoover (1936). It is derived through the following steps. First, slope coefficients (location quotients) for industry i for all regions $m = 1, 2, ..., M$ are calculated. Then regions are ranked in a descending order of their slope, or in other words according to the degree of specialization for the industry in question. In the third step cumulative percentages of employment in industry i over the regions (y-axis) and the cumulative percentage of population over the regions (x-axis) are calculated. "It is as if we joined end to end all the individual sloping lines, beginning with the steepest. The final value in each of the cumulative series is of course 100 per cent; when all the local units have been considered, the entire population and all the employees of the industry are accounted for "(Hoover, 1936, p. 164). If the industry is distributed evenly across industries then the localization curve will be realized as 45-degree line. The more industry becomes regionally concentrated, the more the localization curve becomes concave [2]. Kim (1995) makes use of the Hoover *coefficient of localization* which is analogous to the Gini coefficient to examine the trends in U.S. regional manufacturing structure. It is defined as the

[2]In Hoover (1936) he derives the charts for *localization curves* by plotting a broken line of decreasing slope connecting the origin of the diagram with its opposite corner (100,100). Due to the nature of the constructed graph, the more *convexity* of the curve indicates the more localization. In this study, *localization curve* is expressed in terms of quadrant-I of the cartesian plane, which requires them to be *concave* lying above the 45-degree line. Likewise, Kim (1995) also regard them to be *concave* to indicate localization. Both characterizations refer to the same meaning in terms of localization even terms differ due to the graphical context.

area between the 45-degree line and the localization curve divided by the entire triangular area. It ranges between 0 and 1 where higher values indicate more localization.

3.1.3 Krugman's Locational Gini

Krugman (1991b) proposes the use of Gini coefficient, which are widely explored in income inequality literature, to measure the geographic disparities in economic activity. *Locational Gini* is derived from spatial Lorenz curve and considers the relative concentration of a particular industry in a region as opposed to the same sector in other regions. It is constructed in the same way as *Hoover localization coefficient*, first for each location m, it's share of employment in a particular industry i ($s_{im} = E_{im}/E_i$) and it's share of employment in total employment ($s_m = E_m/E$) is calculated, then these shares are ranked in a descending order keeping the cumulative sums of the both. And finally, cumulated values of the s_m (x-axis) are plotted against s_{im} (y-axis) to obtain *spatial Lorenz curve.* This curve is compared with a uniform distribution of employment across space characterized by a 45-degree line. Hence, the area between the spatial Lorenz curve and 45-degree line will give us the degree of industrial localization, measured by *locational Gini.* Obviously, the more the geographic distribution of a particular industry matches that of overall manufacturing, the closer the curve will lie to the 45-degree line. "An industry that was not localized at all, but simply spread out in proportion to overall employment, would have an index of 0; one that is concentrated almost entirely in a region with small overall employment would have an index dose to $0,5$" (Krugman, 1991b, p. 56).

Kim, Barkley, and Henry (2000, p. 239) bring forth a formal expression to *locational Gini* [3] which is calculated as follows for an industry i:

$$Gini_i = \frac{\Delta}{4\overline{\mu}_x} \tag{3.3}$$

where:

$$\Delta = \frac{1}{M(M-1)} \sum_{m=1}^{M} \sum_{l=1}^{M} |x_m - x_l|$$

[3] There exist different formulations of *locational Gini* within the literature, for instance Holmes and Stevens (2004b), Lafourcade and Mion (2003), Gokan (2010). However, many studies follow the formulation proposed by Kim, Barkley, and Henry (2000), to name a few ; Barkley, Kim, and Henry (2001), Sohn (2004a,b), Ruiz-Valenzuela, Moreno-Serrano, and Vaya-Valcarce (2006), Guillain and Le Gallo (2007)

m , l : indices for regions, $(m \neq l)$
M : number of regions
$x_{m(l)} = \dfrac{\text{Region } m\text{'s } (l\text{'s}) \text{ share of employment in } i}{\text{Region } m\text{'s } (l\text{'s}) \text{ share of total employment}}$
Mean of x_m: $\overline{\mu}_x = \dfrac{\sum_{m=1}^{M} x_m}{M}$

Locational Gini coefficient is favourable for it's ease of computing and lower data requirements. Nevertheless, likewise other discrete indices of the same genre, it fails to account for industrial concentration. Consider for example two extreme cases regarding region m. In the first case, there is only one plant with a large employment in industry i. Instead, in the second case, there are large number of small plants employing the same amount of labour as in the previous case. Clearly, in both cases *locational Gini* will take the same high value indicating a strong concentration pattern in industry i. However, the nature of concentration is totally different in this two cases. In the first case, concentration is at the plant level, presumably related to the factors internal to the firm such as increasing returns to scale. A high *locational Gini* would be observed in any region that it locates. By contrast, the second case displays co-location of different plants at the same place, suggesting for external factors such as labour-pooling, input-output linkages, knowledge spillovers, etc. driving this process.

This shortcoming of the index hinges on the fact that it takes employment figures as the unit of the analysis. To have some insight, we may think about employment distribution over space as the mixture of two distributions; i) plant size distribution (i.e. allocation of employment among plants) and ii) the plant location distribution (i.e. outcome of plants' location choice). The concentration of employment may be due to both sources but only in the second case spatial externalities matter (Lafourcade and Mion, 2003).

Another limitation of the index is it's a-spatial characterization such that it gives information about the concentration of a particular industry in a limited number of locations. However, geographical pattern of these locations remain unanswered, they might be spatially clustered or evenly distributed across the entire area. In both cases the value of the Gini remains the same while agglomeration levels totally differ. This shortcoming of indices alike is tackled by exploring continuous indices and/or exploratory spatial data techniques which count for spatial dependence among locations.

Besides *locational Gini*, Krugman (1991b) also introduce another measure to quantify regional divergence. Let's define s_{im} to be share of industry i in total manufacturing employment in region m and an asterisk refer to some other region.

The index proposed by Krugman (1991b) which indicates the degree of regional specialization for region m is computed as;

$$KSI_m = \sum_{i=1} |s_{im} - s_i^*| \tag{3.4}$$

So, the specialization index of any region m is obtained by the absolute differences between the employment share of region m and other region(s) summed over all industries [4]. The index lies within the range $[0, 2]$, where an observed value of 0 indicates that region m has an identical industrial structure to that of the reference region, hence no specialization in favour of m. On the other hand a value of 2 is observed when region m and reference region have totally disjoint industrial structures, indicating complete specialization for the two regions [5].

Vogiatzoglou (2006) provides a more comprehensive formulation to compare specialization of a region/country with a benchmark of a wider area average as follows:

$$SPEC_m = \sum_i \left| \frac{E_{im}}{\sum_i E_{im}} - \frac{\sum_m E_{im}}{\sum_i \sum_m E_{im}} \right| \tag{3.5}$$

The first term within absolute value is the employment share of industry i in region m and the second term is the employment share of industry i in the entire area. When this difference is summed across all industries, *specialization index* for region m is obtained.

Analogously, concentration version of the Krugman dissimilarity index for any industry i may be constructed by calculating absolute differences between the employment share of region m in industry i and the employment share of region m in the total activity by summing over all locations.

$$CONC_i = \sum_m \left| \frac{E_{im}}{\sum_m E_{im}} - \frac{\sum_i E_{im}}{\sum_i \sum_m E_{im}} \right| \tag{3.6}$$

[4]Krugman (1991b) explore this index to compare the regional specialization in US regions with one another and Europe's four big nations (Germany, U.K, France and Italy) with one another. To make results comparable with European nations he divides US into four regions. Similarly Kim (1995) examines US regional specialization patterns over the period 1860-1987 for nine census divisions using the same index.

[5]The index may well be used to compare an individual region/country with an average of a wider area. For instance, Midelfart-Kvarnik, Overman, Redding, and Venables (2004) employ *Krugman specialization index* to compare regional specialization disparities *individually* for 14 EU countries with an average value calculated for the remaining 13 countries. In a similar fashion, Vogiatzoglou (2006) compare the specialization structure of each NAFTA member with a benchmark distribution which is defined as whole NAFTA area.

Haaland, Kind, and Ulltveit-Moe (1999) argue the utilization of absolute geographic concentration indices with respect to relative ones. Two indices tend to coincide for countries of similar size, but not for the case different sizes. They also state that, which index to use depends on what you focus, for instance if your focus is comparative advantage and specialization then it has to do with relative concentration whereas if you focus on scale economies or trade then the relevant measure should be absolute concentration.

In addition to these measures Hallet (2000) proposed four other concentration indicators to measure spatial dispersion of economic activity. *Concentration* is measured by coefficient of variation which captures the spatial dispersion of economic activity. "*Clustering* measure is based on the gravity model by summing up the distance weighted activity of all pairs of regions. The *centrality* measure expresses if the economic activity is located in the centre or in the periphery. The *income* measure indicates if the economic activity is located in regions with high or low GDP per capita" (Hallet, 2000, p. 8-9) [6].

3.1.4 Herfindahl-Hirshman Index

The Herfindahl-Hirschman index (HHI) is a very common measure used to find out the degree of concentration in a market or industry, as it is calculated by summation of the squared shares of the firms within an industry. It gives an important indication about the level of competition in the respective industry. HHI has also been adopted by scholars from several disciplines (e.g. regional and urban economists) for whom the location and concentration of economic activity are of a particular concern. It may be employed in a way that enables one to measure the industrial geographic concentration or regional specialization. As stated in Lu, Flegg, and Deng (2011), let s_{im} represents the employment share of industry i in region m, where industries are indexed by $i = 1, 2, ..., I$ and regions are indexed by $m = 1, 2, ..., M$. HHI for industrial geographic concentration or industrial localization is defined by

$$H_i = \sum_{m}^{M} (s_{im})^2 \tag{3.7}$$

For a given industry i, if employment is evenly distributed across all regions, in other words if each region gets the same share of employment regarding industry i, then HHI arises as $1/M$. If industry i employment is concentrated in one

[6] Ruiz-Valenzuela, Moreno-Serrano, and Vaya-Valcarce (2006) employ these indices for the concentration of production in Catalonia (Spain) in a methodological comparison context.

region only, then HHI equals 1. In the same vein as above, we may define HHI for regional specialization as follows:

$$H_m = \sum_{i}^{I} (s_{im})^2 \tag{3.8}$$

If all industries in a given region m has the same share of employment, then HHI for regional specialization arises as $1/I$, which indicates an even distribution across industries in region m. Alternatively if region m specializes in only one industry, then HHI equals zero.

A variant of the Herfindahl-Hirschman index, so called *spatial* HHI, is also explored within the localization framework. Kim, Barkley, and Henry (2000), Barkley, Kim, and Henry (2001) and Pede, Florax, and de Groot (2011) mention *spatial* HHI [7] within geographic concentration measures in their studies and algebraically defined as follows:

$$g_i = \sum_{m=1}^{M} (s_{im} - s_m)^2 \tag{3.9}$$

where s_{im} is the share of industry i's employment in area m, and s_m is the the share of total industry employment in area m. *Spatial* HHI compares the regional distribution of industry i's employment with the regional distribution of total employment. Since it is a relative measure, as long as a particular industry i imitates the pattern of total industrial employment, it won't be considered as being concentrated. In this case, the index will take a value of 0. For values greater than 0, the index is interpreted as indicating spatial concentration of industrial activity. Put differently, higher values of the *Spatial* HHI indicate sectoral concentration in a limited number of regions.

As also mentioned by Kim, Barkley, and Henry (2000) and Barkley, Kim, and Henry (2001), one of the innate limitations of this index is that it doesn't distinguish between random and non-random distributions of plants within space. Secondly, it is very sensitive to the number of plants. If the number of regions exceed the number of plants in an industry, the *spatial* HHI tends to be calculated upward biased since s_{im} will be zero for many regions and square of the $-s_m$ will be inflating the industry's index value [8].

[7] This measure is also referred as *spatial Gini* in the literature (for instance in Lu, Flegg, and Deng (2011)), however the method of calculation is much analogous to Herfindahl-Hirschman index, therefore using *spatial* HHI term is preferred in this study.

[8] Both Kim, Barkley, and Henry (2000) and Barkley, Kim, and Henry (2001) give the same

Ellison and Glaeser (1997) recognize this issue and remark that geographic concentration is not neutral to industrial concentration. Ceteris paribus, industrially concentrated sectors will tend to exhibit a higher geographic concentration because there are small number of plants and eventually economic activity has to be concentrated in fewer places. They tackle this issue by incorporating plant size distribution into the agglomeration index built on a probabilistic location choice model. Furthermore, the agglomeration index is derived from a location choice model where geographic concentration is based on a test of randomness which is defined as the expected distribution of plants in the absence of agglomerative forces. As the index developed by them rests upon a conditional distribution of plants, it purges agglomeration from industrial concentration. Hence, in these two grounds (considering plant size distribution and building on a statistical location choice model) Ellison-Gleaser (EG)index and their approach is considered as a significant improvement over earlier concentration measures (Kim, Barkley, and Henry, 2000). Next section examines EG index more in detail.

3.1.5 Ellison-Glaeser (1997) Index

Classical measures have been criticized for their inadequacy to distinguish between "random concentration arising from industrial concentration and concentration arising from agglomerative externalities or natural advantage" (Rosenthal and Strange, 2001, p. 194). To address this problem, Ellison and Glaeser (1997) have proposed an index to purge geographic concentration from industrial structure.

Agglomeration index of plants presented by Ellison and Glaeser (1997), EG henceforth, basically compares the observed geographic distribution of plants with a random distribution. The plants are defined to be geographically randomly distributed by considering their expected distribution in the lack of agglomerative forces. In their model, plants cluster either to benefit from natural advantages or spillover externalities from other plants.

They start with a simple location model where firms of an industry, with N plants, choose locations among M geographic units in order to maximize their

example for this case. Assume that there are 10 equally sized regions in a country, such that $s_m = 0.1$ and an industry with only three plants. The *spatial* HHI for this industry has a minimum value of 0.233, a maximum value of 0.900, and a mean value of 0.354. Alternatively, when we assume another industry with four firms only, above values are estimated as 0.150, 0.900, and 0.286 respectively. Hence index calculated with three plants yields a higher mean value than the one with four plants even if plants in both industries are randomly distributed among the 10 regions.

profits. Assume that n^{th} plant chooses location r_n to maximize its profits given that it will gain a profit of π_{mn} from locating in region m. More formally, these profits are given by;

$$log\pi_{nm} = log\overline{\pi}_m + g_m(r_1, ..., r_{n-1}) + \epsilon_{nm} \tag{3.10}$$

where $\overline{\pi}_m$ is a random variable reflecting the profitability of locating in region m for a typical firm in the industry, $g_m(.)$ represent the effect of spillovers on the profitability of firm n created by the plants that have previously chosen their locations. The random component ϵ_{nm} reflect the factors that is specific to firm n.

So, basically two factors are at play in the profit equation. First one is "natural advantages" which is captured by $\overline{\pi}_m$ and determined by the nature at the start of the process. It is important to include it in the model because common factors such as natural resources, climate and landscape characteristics may render some locations more attractive for the firms belonging to the same industry. The expectation of $\overline{\pi}_m$ hence will reflect the average profitability of locating in region m. The second role is played by "local spillovers", represented by $g_m(.)$ which clearly reflects the within industry externalities arising from locating in the same area with other plants.

As the random component ϵ_{nm} is specified as an independent Weibull random variable independent of the $\overline{\pi}_m$ and no spillovers assumption ($g_m \equiv 0$ for all m) is made, then the model is assumed to be a standard logit model conditional on the realization of $\overline{\pi}_1, ..., \overline{\pi}_M$ and firms' location choices are conditionally independent random variables with probability given by;

$$\text{prob}\{r_n = m | \overline{\pi}_1, ..., \overline{\pi}_M\} = \frac{\overline{\pi}_m}{\sum_{j=1}^{M} \overline{\pi}_j} \tag{3.11}$$

EG make two key parametric restrictions on the distribution of $\overline{\pi}_m$. First, on average the model reproduces the prevalent overall distribution of manufacturing activity for all industries. In other words, in the absence of externalities, each industry would mimic the general pattern of the overall activity, such that:

$$E_{\overline{\pi}_1, ..., \overline{\pi}_M} \frac{\overline{\pi}_m}{\sum_{j=1}^{M} \overline{\pi}_j} = x_m \tag{3.12}$$

where x_m is area m's share in overall manufacturing employment.

Furthermore, the second assumption is made on the variance of the natural advantage shares:

$$\mathrm{var}\left(\frac{\overline{\pi}_m}{\sum_{j=1}^{M}\overline{\pi}_j}\right) = \gamma^{na} x_m(1-x_m) \tag{3.13}$$

The parameter γ^{na} captures the importance of natural advantages to the industry. The extreme case $\gamma^{na} = 0$ implies a model in which unobserved area characteristics has no effect on the average profitability. The other extreme $\gamma^{na} = 1$ corresponds to a case where natural advantages completely determine firm profitability hence firm location.

EG consider *spillover* theories as a second class of explanation for agglomeration. They refer broadly to "technological spillovers, gains from sharing labour markets, gains from inter-firm trade, the effect of local knowledge on the location of spin-off firms, and any other forces that might provide increased profits to firms locating near other firms in the same industry" by using the term *spillovers* (Ellison and Glaeser, 1997, p. 894). However, the spillovers they consider are "all or nothing" type, such that plants benefit fully from the externalities created by other firms if they choose identical locations and no benefits at all if they locate at separate areas. The spillovers are captured by the parameter $\gamma^s \in [0,1]$ by assuming that

$$log\pi_{nm} = log\overline{\pi}_m + \sum_{l \neq n} e_{nl}(1-u_{lm})(-\infty) + \epsilon_{nm} \tag{3.14}$$

where $\{e_{nl}\}$ are Bernoulli random variables equal to 1 with probability γ^s that indicate whether each pair of plants is likely to benefit from a potentially valuable spillover and u_{lm} is an indicator variable defining whether the plant l is located in area m. They also assume that spillovers between plants are symmetric and transitive in the sense that $e_{nl} = 1 \Rightarrow e_{ln} = 1$ and $e_{nl} = 1$ & $e_{ls} = 1 \Rightarrow e_{ns} = 1$ [9]. γ^s captures the importance of spillovers indicating the proportion of the pairs of firms between which a spillover exist.

Labelling s_{im} as the share of industry i's employment in area m, and x_m as the the share of aggregate manufacturing employment in area m, they define a raw concentration index for industry i as follows:

$$G_i = \sum_{m}^{M}(s_{im} - x_m)^2 \tag{3.15}$$

[9] This assumption signifies the rational expectations equilibrium of the model in which plants are forward looking and the nth plant considers only the the locations of the first $n-1$ plants and the order of the plants' location choices does not affect the resultant distribution of locations.

They show that the location model in which plants sequentially choose locations to maximize their profits satisfying equations (3.12),(3.13) and (3.14) yields:

$$E(G_i) = \left(1 - \sum_m x_m^2\right)[\gamma + (1-\gamma)H_i)] \tag{3.16}$$

where H_i is the Herfindahl index of the industry's plant size distribution given by $H_i = \sum_{j=1}^{N}(z_{ij})^2$, $j = 1, ..., N$ indexes number of plants in industry i, and z_{ij} denotes the employment share of jth plant in industry i and parameter $\gamma = \gamma^{na} + \gamma^s - \gamma^{na}\gamma^s$. When plants make location decisions in compliance with the location choice model built by EG, they suggest that expected value of G_i is related to the parameters qualifying the intensity of natural advantages and spillovers, plant size distribution of the industry and the size of the areas.

Equation (3.16) implies that in the absence of natural advantages or spillovers or both ($\gamma = 0$), the expected value of the raw geographic concentration of an industry, G_i, should be proportional to its industrial concentration, H_i. In this case the geographic units are equally attractive to plants and hence locations are chosen randomly.

Using the expression (3.16) above, they derive an estimator of excess concentration, γ, which is called the agglomeration index. They note that EG agglomeration index γ is an unbiased estimate of the quantity $\gamma^{na} + \gamma^s - \gamma^{na}\gamma^s$ that captures the strength of agglomerative forces and is given by the following for industry i:

$$\gamma_i = \frac{\dfrac{G_i}{\left(1 - \sum_m^M x_m^2\right)} - H_i}{(1 - H_i)} \tag{3.17}$$

For an industry with a large number of small plants, which may be regarded as perfectly competitive, H_i approaches zero and γ_i approaches $G_i/(1 - \sum_m^M x_m^2)$. In a case like that, G_i measures spatial concentration without any involvement with industrial organization. γ_i takes a value of zero if plants are distributed randomly by the dartboard model of random location choices with no natural advantages or industry-specific spillovers, while a positive value of γ_i indicates excess concentration. They also provide some value range for their index according to which they classify industries as not very concentrated ($\gamma < 0.02$), relatively concentrated ($0.02 \leq \gamma < 0.05$) and highly concentrated ($\gamma \geq 0.05$). A negative value would indicate dispersion of economic activity.

Having regard to the model developed by EG and the derived index of agglomeration, a few points worth remarking. First of all, the model developed by EG might be considered as a prominent improvement over the prevalent measures of agglomeration. Their contribution is twofold; the model disentangles geographic concentration due to randomness from agglomerative forces and it purges agglomeration from industrial concentration by incorporating the plant size distribution of industries into the computation of agglomeration index. The authors demonstrate the desirable properties of the index as; (i) is easy to compute given the available data, (ii) allows one to make comparisons with a no-agglomeration benchmark ($E(\gamma) = 0$), (iii) comparable across industries regardless of differences in both the number of plants and their distribution, (iv) comparable across industries regardless of differences in the level of geographic aggregation.

Eventhough EG index may be considered as a significant improvement over its predecessors, it still suffers from two major drawbacks such that its *a-spatial* characteristic and aggregation issues. The first one is known as *checkboard problem* and while the second one is *modifiable areal unit problem (MAUP)* which of both will be discussed in detail in section 3.1.7.

Duranton and Overman (2005) propose five major requirements for an appropriate localization measure such that (i) comparability across industries; (ii) controlling for the overall agglomeration of manufacturing; (iii) controlling for industrial concentration; (iv) unbiasedness with respect to scale and aggregation and (v) test of significance of results. According to these criteria, EG index meets first three of them. However, there has been a significant improvement in the approach undertaken by scholars to overcome the deficiencies of the index.

Feser (2000) shows that EG concentration measure is not robust to different spatial aggregation levels as proposed by the authors. He recommends the use of the index with high degree of caution and sensitivity testing with different levels of spatial aggregation. Even if he is well aware of MAUP, he does not provide any alternative to account for that. Guimarães, Figueiredo, and Woodward (2007) notice that EG index reflects localization externalities felt by larger firms more than the small ones since it is an employment weighted measure. They give an hypothetical example as shown in Table 3.1 assuming that there are 10 plants in an industry and 9 of which are located in region A (lighter gray) each employing 1 worker and the last and large one is located in region B (darker gray) employing 36 workers by itself. In this case (a), assuming each area is equally attractive with an $x_m = 0.1$, the EG index is calculated as 0. According to the authors it is nor correct to evaluate this industry as non-localized as 90 per cent of the plants seem to be affected by localization economies since they have chosen the same region to locate.

Table 3.1: **Two cases with small and large plants**

(a)

9	0	0	0	0
0	0	0	0	36

(b)

9	0	0	0	0
0	0	0	0	6

In another situation, case (b), where regional distribution of employment coincides with the expected location probabilities such that x_m for region A and B are 0.6 and 0.4 respectively and the large firm employs 6 workers. In this case, EG index is computed as -0.25 which indicates a non-localization of the industry. It is not easy to figure out why this industry is evaluated as non-localized even if small plants are concentrated in region A composing 60 percent of the industry employment. In order to circumvent this type of problems, they propose the use of plant counts rather than employment levels to measure agglomeration and modify EG index in accordance with that.

With regard to statistical properties, Cassey and Smith (2014) improve the assessment of the index by simulating confidence intervals to be used for the statistical test of significance as an alternative to the *ad hoc* treshold values defined by Ellison and Glaeser (1997). Another recent working paper by Billings and Johnson (2014) test the size and power properties of EG index through a series of simulations along with locational Gini and Duranton and Overman (2005) index.

Guimarães, Figueiredo, and Woodward (2011) address the checkboard problem which refers to the disregard of spatial proximity of regions and neighbouring effects that also EG index fails to consider. They modify the index by incorporating a spatial weight matrix into the formula to account for neighbourhood effects. Essentially, they propose inflating the ordinary EG index by spatial weights to consider the level of spatial autocorrelation across locations.

3.1.6 Maurel-Sedillot (1999) Index

Maurel and Sédillot (1999) (MS) proposes an index based on the location model suggested by Ellison and Glaeser (1997) which is however slightly different from that. Let N denote the number of industry plants and $z_1, ..., z_n$ the share of each plant in the industry. Similarly, M could be defined as the number of regions and $x_1, ..., x_m$ the fraction of each area in aggregate employment. The share of a specific industry employment located in area i is given by

$$s_i = \sum_{j=1}^{N} z_j u_{ji} \tag{3.18}$$

where $u_{ji} = 1$ if the business unit j locates in area i and otherwise 0. u_{ji} are defined as non-independent Bernoulli variables such that $P(u_{ji} = 1) = x_i$ "which means that the random location process on average lead to a pattern of employment shares matching the one that prevails in the aggregate"(Maurel and Sédillot, 1999, p. 578). Interaction between the location decisions of any pairs of plants is given by $Corr(u_{ji}u_{ki}) = \gamma$ for $j \neq i$. The parameter γ describing the strength of spillovers within the industry and ranges between $[-1, 1]$. The probability of two business units j and k locating in the same area i is independent from the business units and simply written as:

$$P(i,i) = E(u_{ji}u_{ki}) = Cov(u_{ji}u_{ki}) + E(u_{ji})E(u_{ki}) = \gamma x_i(1 - x_i) + x_i^2 \tag{3.19}$$

Anf finally, the probability of any pairs of plants locating in any same location is:

$$p = \sum_i P(i,i) = \gamma(1 - \sum_i x_i) + \sum_i x_i^2 \tag{3.20}$$

By using the linear relationship between p and γ MS derive a simple estimator of the γ from a natural estimator of the probability p. They suggest to weight the frequency estimator by the size of the plants which is discussed to be consistent with an Herfindahl measure of concentration giving more weight to larger plants. Hence, they suggest the use of weighted estimator which is written as:

$$\widehat{p} = \sum_i \frac{\sum\limits_{j,k \in i, j \neq k} z_j z_k}{\sum\limits_{j,k \in i, j \neq k} z_j z_k} \tag{3.21}$$

$j, k \in i$ denotes the event $\{$the business units j and k are located in region $i\}$. Sums of the formula gives the following by a simple calculation:

$$\widehat{p} = \frac{\sum_i s_i^2 - H}{1 - H} \tag{3.22}$$

where $H = \sum_j z_j^2$ is the Herfindahl index of the industry. Therefore Inserting equation (3.22) in (3.20) yields:

$$\widehat{\gamma} = \frac{\widehat{p} - \sum\limits_i x_i^2}{1 - \sum\limits_i x_i^2} = \frac{\dfrac{\sum\limits_i s_i^2 - \sum\limits_i x_i^2}{1 - \sum\limits_i x_i^2} - H}{1 - H} \tag{3.23}$$

The estimator $\widehat{\gamma}$ suggested by MS, which may be called γ_{MS} henceforth, differs slightly from the one suggested by EG. In EG framework, the estimator γ_{EG} mentioned in eq(3.17) is derived from the *a priori* definition of a raw geographic concentration index G_{EG} as shown in eq(3.15). From this definition EG build the estimator γ_{EG} as rewritten below:

$$\gamma_{EG} = \frac{G_{EG} - H}{1 - H} = \frac{\dfrac{\sum\limits_i (s_i - x_i)^2}{1 - \sum\limits_i x_i^2} - H}{1 - H} \tag{3.24}$$

MS note that the estimator γ_{MS} has a more natural specification than γ_{EG} because it is derived directly from the probability model without making any *a priori* definitions. "In particular, the Herfindahl index H that shows up in the expression of $\widehat{\gamma}$ comes directly from the writing of the frequency estimator $\widehat{p}$" (Maurel and Sédillot, 1999, p. 580). The difference between the estimators γ_{MS} and γ_{EG} lies in the first term of the numerator, which of both basically measures the raw geographic concentration. To compare these two estimators better, the MS estimator may be rewritten as:

$$\gamma_{MS} = \frac{G_{MS} - H}{1 - H} = \frac{\dfrac{\sum\limits_i s_i^2 - \sum\limits_i x_i^2}{1 - \sum\limits_i x_i^2} - H}{1 - H} \tag{3.25}$$

Now, as may be seen apparently the difference between MS and EG index comes from the differences in the numerator of G_{MS} and G_{EG}. Apart from that (i) both estimators are unbiased, (ii) for both estimators it can be shown that $E(G) = H + \gamma(1 - H)$, (iii) in both models γ follows the same interpretation such that a value greater than 0 shows geographic concentration in excess of random location choices. And (iv) both models do not differentiate between spillovers or natural advantages.

However, despite the slight difference between these indices, there is a nuance between them which highlights the importance they put to divergences between the industry analysed and the overall activity. If a location has a higher percentage of the industry than the overall activity ($s_i > x_i$) MS index contributes with a positive factor, while in the contrary case it's contribution is negative. For instance assume that industry share, s_i is 15 percent while region's share within overall activity, x_i is 10 percent. Also keep in mind that the first term of the numerator of the MS index and EG index can be written as $\sum_i (s_i - x_i)(s_i + x_i)$ and $\sum_i (s_i - x_i)(s_i - x_i)$, respectively. For the ratios presumed above, the contribution of the MS index, calculated as $(15-10)(15+10)$, is higher that the EG's contribution, $(15-10)(15-10)$. For the contrary case where $s_i < x_i$, MS index contributes with a negative factor while EG is positive. Moreover, if a location has a high level of overall economic activity and an even higher level of activity for the industry, its contribution to the index is very great, while if it has little overall activity, even though the weight of the industry in question is greater, its contribution is positive but small (still higher than its contribution to EG index)(Alonso-Villar, Chamorro-Rivas, and González-Cerdeira, 2004).

Therefore, it could be inferred that MS index takes higher values when the industry is located in the most industrialized regions. On the other hand, it displays a lower concentration if the industry in question is located at economically less active areas. Since EG index disregards the sign of the difference between industrial percentage and industrial aggregate, the contribution to the index is the same in both cases. So MS index may said to be "more sensitive to spatial distributions where firms are located in the most industrialized areas" (Alonso-Villar, Chamorro-Rivas, and González-Cerdeira, 2004, p. 2106).

3.1.7 Discussion

Notwithstanding that "second-generation" indices have shown great improvement over traditional measures of concentration, yet they suffer from the drawbacks of having an a-spatial character and aggregation problems, likewise traditional ones [10]. The spatial characteristic of the data is integrated out in the course of computing the index, and as a result the index derived gives information about the localization of the industry if it is above the threshold values. In other words, the computed index hints at an uneven distribution of the economic activity for a particular industry, but it does not inform about their spatial positioning. This problem is known as "checkboard problem" and was brought to attention by White (1983) and Griffith (1983) (as cited in Guimarães, Figueiredo, and Woodward (2011)).

The second problem is related to the spatial aggregation of the data to in the course of computing the index. Most apparently, aggregation restricts the analysis to a very limited number of spatial scale, because in many countries the

[10] Duranton and Overman (2005) refer to three generations of measures of spatial concentration. First generation involves Gini-type measures where in fact place has no role. EG and MS indices are considered as second generation indices which take into account plant size distribution hence industrial concentration. Third generation measures covers the ones that treat space as continuous.

number of levels of aggregation is very limited to two or three. Moreover it is difficult to compare results across different spatial scales. "For instance, questions regarding how much industries are localized at the county level after controlling for localization at the regional level cannot be precisely answered since existing indices are usually not easily additive across different levels of aggregation" (Duranton and Overman, 2005, p. 1078). Furthermore, agglomeration externalities may reach beyond the spatial boundaries which are generally defined according to administrative issues rather than economic relevance. So, the definition of the spatial boundaries, or the *size* and *shape* of the geographic units are likely to alter statistical results. In other words, the results will display different concentration patterns under different spatial zoning systems. Another important issue arising with aggregating establishments at any spatial level is spurious correlations across aggregated variables. And this becomes more problematic as higher levels of aggregations are performed (Duranton and Overman, 2005). This problem is known as "modifiable areal unit problem" and the term was first introduced by Openshaw and Taylor (1979) (as cited in Briant, Combes, and Lafourcade (2010)).

Finally and importantly, once data is aggregated, the spatial units are treated as independent in space. In effect, neighbouring spatial units and spatial units at the opposite ends of a country are treated symmetrically. This creates a downward bias in the measurement of localization where administrative boundaries are surpassed and raises the issue of spatial autocorrelation.

Arbia (2001b) notes that spatial concentration comprises two different characteristics; an a-spatial concept of concentration that is insensitive to the permutation of the data and the concept of polarization. Further he adds that traditional measures only captures the first concept and disregard the second.

To clarify the subject better, the case pointed out by Arbia (2001b, p. 271) may be referred as an illustrating example. Table 3.2 shows a hypothetical geographic area and the distribution of 12 plants over the 16 local units embodied in the cells of a 4x4 grid. In three cases through (2a) to (2c), the spatial configurations depict different levels of agglomeration [11], which is also intuitively obvious, highest in case (2a) - and higher in case (2b) than case (2c), even though concentration levels remains the same measured by different concentration indices as shown in Table 3.1 provided by Guimarães, Figueiredo, and Woodward (2011). They show that Herfindahl index (H) (computed as the sum squares of the re-

[11] Arbia (2001b) precisely uses the terminology *polarization* in the name of agglomeration to highlight the importance of considering the relative position of the data in space.

gional shares of number of firms), EG raw concentration index (G), EG index (γ) and locational Gini yield exactly the same results when applied to any of the three cases shown in Table 3.2. These concentration measures fail to recognize ranking agglomeration (2a) through (2c), just like other commonly used concentration measures i.e. entropy indices.

Table 3.2: **Three cases with the same level of concentration**

3	3	0	0
3	3	0	0
0	0	0	0
0	0	0	0

2a

3	0	0	0
3	0	0	0
3	0	0	0
3	0	0	0

2b

0	0	0	0
0	3	0	3
0	0	0	0
0	3	0	3

2c

Table 3.3: **Concentration measures for three cases**

	2a	2b	2c	4b
H	0.2500	0.2500	0.2500	0.0833
G	0.1875	0.1875	0.1875	0.0208
γ	0.1273	0.1273	0.1273	-0.0667
Gini	0.7500	0.7500	0.7500	0.2500
Moran's I	0.6111	0.4861	-0.3333	0.4861

This discussion raises the importance of considering distances between spatial units. In order to capture *spatial* dimension of agglomeration, spatial association indices are regarded as appropriate exploratory tools accounting for distances among spatial units. Among many measures of spatial association, Moran's I index of spatial autocorrelation is one of the most widely used. Spatial autocorrelation basically refers to the degree of spatial dependency among observations in a geographic space. It measures whether observations tend to be clustered in space due to a set of common spatial features (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation). As seen in Table 3.3 Moran's I distinguishes three cases of agglomeration even the concentration measures give the same value.

However, it is important to note that, spatial autocorrelation measure *per se* is not a good measure of spatial concentration. Arbia (2001b) illustrates this point by considering two alternative scenarios given in Table 3.4. In this table, it is clear that the level of dispersion is higher in case 4b than 2b. In other words, case 2b is more agglomerated than case 4b. This fact is reflected in concentration measures in Table 3.1 and show higher values for case 2b. However, if Moran's I statistic is considered as a concentration measure we come to conclude that the degree of concentration is exactly the same in both cases reflected by the same value of Moran's I statistic, 0.4861. This is basically due to the fact that Moran's I statistic is designed to take into consideration the degree of resemblance or similarity between values in contiguous areas, accounting for neighbouring effects, but fails to account for the information within each spatial unit. This stems from the definition of the spatial contiguity matrix in the calculation of the statistic.

Table 3.4: **Two cases with the same spatial autocorrelation**

3	0	0	0
3	0	0	0
3	0	0	0
3	0	0	0

2b

0	1	1	1
0	1	1	1
0	1	1	1
0	1	1	1

4b

Therefore, Arbia, Copetti, and Diggle (2009) recommends to consider spatial distribution in terms of both concentration in an a-spatial sense and *polarisation* simultaneously. He applies a descriptive plot termed GI-plot, where G accounts for locational Gini in the vertical axis and I for Moran's I statistic in the horizontal axis. This descriptive plot basically divides the graph area into four boxes which represent four extreme cases of spatial concentration. The top-right box represents high concentration in non-spatial sense ($G > 0.5$) and high polarization evidenced by positive spatial autocorrelation. On the contrary, top-left box represents cases where a-spatial concentration is low ($G < 0.5$) and it is accompanied by low polarisation reflected by negative spatial autocorrelation. However, GI-plot only provides quasi-ordering of the several geographic situations, it does not provide full ranking of the situations such that it fails to rank points falling within the top-left and bottom-right boxes. In a previous study (Arbia, 2001b), he proposes to combine Gini type location index (G), Getis Ord coefficient(GO) and Moran's I statistic in a multiplicative way in order to measure a-spatial concentration and polarization simultaneously.

Recently a number of studies make joint use of two distinct statistical indicators, one for the a-spatial concentration measure and one for polarization, to eliminate some of the deficiencies of discrete indices. Guillain and Le Gallo (2007) developed an approach in which they combine the locational Gini with the tools of Exploratory Spatial Data Analysis in order to measure the degree of spatial agglomeration and identify location patterns of economic activities in Paris. Similarly, Lafourcade and Mion (2007) show that concentration and agglomeration are different concepts where the former one is insensitive to the relative position of locations and presents spatial variability with respect to some average. In order to take into account spatial effects they apply EG index and Moran's I index together for the case of Italy. They also consider plant size in their analysis where small and large firms exhibit different concentration and agglomeration patterns. Feng and Ji (2011) investigate point patterns and inter-industrial spatial associations of Chinese manufacturing industries by applying location quotient and spatial autocorrelation measure. In a very recent study Sohn adopts four indices developed for measuring different aspects of spatial distribution. Within-industry and between-industry concentration are measured by EG concentration and co-agglomeration index, respectively. Likewise within and between industry agglomeration is captured by employing Moran's I and bivariate Moran's I, respectively.

Another problem associated with discrete indices, as mentioned above, is *modifiable areal unit problem*, henceforth referred as MAUP. Broadly, it refers to the arbitrariness in partitioning the geographic space. The problem has been realised long before by many statisticians dealing with spatial data. To mention a few significant early studies, Gehlke and Biehl (1934), Yule and Kendall (1950) and Blalock (1964) investigated the effects of MAUP, more precisely the effects of scale and data aggregation, on the correlation coefficients (as cited in Openshaw (1984)). However, "apart from the occasional mention, the MAUP seems to have been ignored until the problem was re-examined in the late 1970's" (Openshaw, 1984, p. 13). Openshaw (1977) is one of the first who draws attention to the MAUP often encountered in studies of spatially aggregated data. He asserts that MAUP manifests due to two different but related problems, namely *scale* and *aggregation* problems.

Scale problem is related to the size and number of zones used to partition the study area and it arises because of uncertainty about these units needed for a particular study. For example, when districts are aggregated into counties, the results change with increasing scale. On the other hand *aggregation* problem arises due to the alternative combinations of areal units at equal scales. In other words, the results may be sensitive to different ways of aggregating data while the number of geographical units is held constant [12]. Figure 3.1 illustrates different ways of aggregating 16 spatial zones into 8 and 4 regions. Obviously, the results which will be obtained under different aggregation schemes would yield different levels of spatial concentration measures.

[12]However, Briant, Combes, and Lafourcade (2010) investigates whether the choice of zoning systems differentiated according to size and shape of their spatial units affect the economic geography estimates. They conclude that size might matter and shape does so much less, especially when the dependent variable of interest is not aggregated in the same way as the explanatory ones and the zoning system is composed of large spatial units. Rather, specification issues are of first importance compared to this issue.

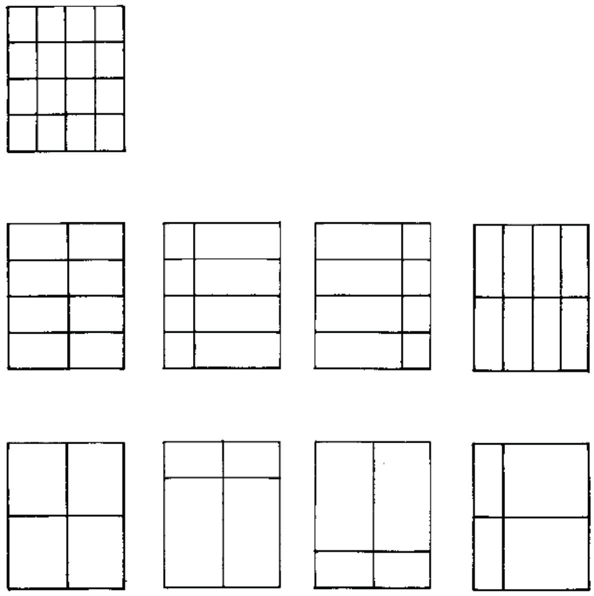


Figure 3.1: **Alternative aggregations of zones**
Source: Openshaw (1984, p. 9)

Arbia (2001a) notices the MAUP and shows in an illustrative way how scale and aggregation problems are likely to affect concentration measures. Figure 3.2 first shows a continuous space distribution of firms represented by an asterisk in panel (a) and then displays three discretized versions of it through (b) to (d). It is obvious to observe that Figure 3.2(a) presents a geographic concentration at the center of the study area. Suppose that the study area is divided into 4X4 grid as shown in Figure 3.2(b) In this situation any of the concentration measures would detect that there is no geographic concentration. However, using the same grid, if we shift the origin in the north-west direction as in Figure 3.2(c), this time we would get the opposite result indicating a high level of concentration that any concentration measure would detect. So these two figures illustrate the aggregation problem.

Contrarily, if a finer grid would have been imposed onto the study area as in Figure 3.2(d), a concentration index would range between case (b) and (c) which shows an intermediate level of concentration. Here is the description of scale problem.

Openshaw (1977) describes MAUP as one of the greatest unsolved problems faced by spatial studies. As a solution he proposes the use optimal-zone design approach where a set of zones are identified which optimizes an objective function related in some way to the performance of a model. In this algorithm, the initial data is assumed to relate to a set of zones and then these zones are aggregated

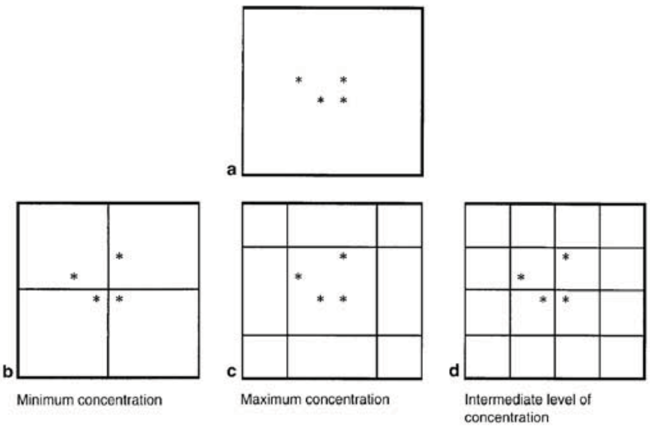


Figure 3.2: **Illustration of aggregation and scale problems**
Figures (b) and (c) illustrate the aggregation problem while (b) and (d) illustrate the scale problem. Source: Arbia (2001a, p. 414)

into a smaller number of large regions (e.g. a study area with 60 zones is aggregated into 10 regions). The aggregation is performed in a way to approximately optimize an objective function ensuring that all zones assigned to the same region are internally connected. The objective function could be any general function, for example the aim may be to maximize a correlation coefficient between two variables. Instead of using optimal zone approach, Arbia (2001a) proposes simply removing boundaries and proceed to analyse the economy on a continuous space which is the subject of the next section.

3.2 Continuous Indices

A possible solution to the problems associated with discrete indices due to aggregation and zoning issues discussed above arises as the abolition of spatial boundaries and perceiving space in a continuous plane. We may associate continuous space case with dots on a map that visually corresponds to case (a) in Figure 3.2 while in discrete case dots are transformed into units in boxes as in cases (b) to (d). Contrary to cluster-based methods, *distance-based* methods do not discretize area under consideration into spatial subunits, but instead treats as a continuous space which allows one to compare results across different scales. With respect to problems associated with MAUP-effected indices, Duranton and Overman (2005),DO henceforth, propose five criteria that a reliable localization index should rely on which are; (i) comparability across industries; (ii) controlling for the overall agglomeration of manufacturing; (iii) controlling for industrial concentration; (iv) unbiasedness with respect to scale and aggregation and the test should also be (v) giving an indication of the significance of the result. The first two requirements have long been recognized within the literature and satisfied by most of the traditional measures, like Gini coefficients. The methodology introduced by Ellison and Glaeser (1997) has been significant in the sense that fulfilling the third requirement. But still, the realm of discrete indices fail to meet last two criteria which are performed by distance-based measures.

The basic idea behind the distance-based measures is to "consider the distribution of distances between pairs of establishments in an industry and to compare it with that of hypothetical industries with the same number of establishments which are randomly distributed conditional on the distribution of aggregate manufacturing" (Duranton and Overman, 2005, p. 1079). Within the continuous modelling terminology, firms, plants, shops etc. say more generally entities, whose spatial distribution is considered are termed as *points*. Then, what they do is to identify the spatial structure of the *point* distribution. Since they are based on the distances separating pairs of entities they are called *distance-based measures*.

The best known among the existing distance-based methods is K function introduced by Ripley (1977). Its use in economic analysis was first introduced in the literature by Arbia and Espa (1996) (as cited in Arbia, Espa, and Quah (2008)), then exploited by others among which prominent ones are Marcon and Puech (2003) and Duranton and Overman (2005) [13]. K function, despite being utilised prelusively, a great variety of distance-based methods have been improved and employed by researchers, especially over the last decade. Marcon and Puech (2012) present a technically comprehensive study on a diverse set of distance-based methods (e.g. K, K_{mm}, D, g_{inhom}, K_d, M functions) which helps one to understand all properties of these new tools. It is well beyond the scope of this section to present the excessive technical details of these functions where they vary due to different theoretical assumptions, but it's worth mentioning common steps followed by them in the course of building the function as they follow the same pattern in the end.

Essentially, whichever function is employed, continuous space modelling involves three successive steps. First, after choosing relevant establishments, bilateral distances between pairs of plants are calculated. This measure is weighted by spatial density in order to refrain from spatial scale and aggregation issues. The second step is to construct a reference distribution in order to determine null hy-

[13] In fact, Duranton and Overman (2005) developed their own K function, which they call K density. The comparison of two methods may be seen in Marcon and Puech (2003)

pothesis which is usually set as randomly distributed set of locations in the area under consideration. And finally, observed distribution of distances is compared with the reference distribution to assess the significance of concentration out of randomness. For the sake of clarity illustrate these steps, K function of Ripley (1977), as being widely known and among the most frequently used measures in analysing distribution of points and secondly, K-density function developed by Duranton and Overman (2005) which has been influential in the field will be mentioned.

3.2.1 K Function

Albert, Casanova, and Orts (2012) show a brief presentation of K function, hence we follow it in an attempt to make a basic introduction without dealing with the theoretical background for which reader may refer to Ripley (1977).

What K function, $K(r)$ does basically is to measure concentration by counting the average number of neighbours each firm has within a circle of a given radius where *neighbours* should be perceived as to mean that all firms situated at a distance equal to or lower than the radius (r). It describes the characteristics of point patterns at many and different scales simultaneously which is represented as:

$$
\begin{aligned}
K(r) &= \frac{1}{\lambda N} \sum_{i=1}^{j} \sum_{j=1, j \neq i}^{j} w_{ij} I(d_{ij}) \\
I(d_{ij}) &= \begin{cases} 1, & d_{ij} \leqslant r \\ 0, & d_{ij} \geqslant r \end{cases}
\end{aligned}
\tag{3.26}
$$

where d_{ij} is the distance between i^{th} and j^{th} firm, N is the total number of points observed in the area of the study region, $\lambda = N/A$ representing its density A being rectangular area covering the study region and w_{ij} is the weighting factor to correct for border effects [14]. $I(d_{ij})$ is the indicator function taking a value of 1 or 0 depending on the magnitude of the bilateral distances between points.

Then, next step comes as determining the null hypothesis and comparing

[14] Border effect corrections should be taken into account especially when r grows in magnitude. For larger values of r there arises a gap between the study area and the boundaries of the circle. Some neighbours of points close to the border of the study area will not be counted since they are lying outside study area, thus they will be underestimated and will lead to a bias in the results.

it with the observed results. Usually, null hypothesis is built up as a sort of randomly distributed set of points in the study area. The benchmark refers to the case where firms locate with the same probability (constant density) and independently from each other at any place. This is known as *complete spatial randomness* (CSR). Supposing a completely random distribution for each point i within the study area, expected number of points in a circle of radius r is $\lambda \pi r^2$. As points located inside a circle around a firm are termed *neighbours* and $K(r)$ is defined as the average number of firms divided by λ, as long as CSR assumption is made the K function will be equal to πr^2 (Marcon and Puech, 2003).

The difference between the empirical K value of the real point pattern of each sector and the theoretical K value (benchmark) is calculated. If the empirical K value, $K(r)$, is higher than the theoretical one, πr^2 , it demonstrates the concentration of the observed point pattern distribution, since the real density is greater than that of the benchmark. Lower values indicate dispersion and if $K(r)$ equals πr^2, then it means that observed points are independently distributed.

And finally, in order to assess statistical significance of departures from randomness robustly, confidence intervals are constructed. Since the distribution of K is unknown, the variance can not be evaluated and hence used in constructing confidence intervals. Alternatively, Monte Carlo method is exploited, which comprises generating a large number of independent random simulations of homogeneous spatial distributions with the same number of points and same density as in the sample (Marcon and Puech, 2003).

Marcon and Puech (2003) apply K function to measure the concentration patterns of French manufacturing industries. To normalize K function and compare concentration with respect to 0 reference value, they employ Besag's L function. Their study area covers Paris and its 40X40 km surrounding. Their results show significant concentration for all distances from 0 to 25 km for all manufacturing sectors. This method not only displays the geographic concentration of firms but at the same time detects differences between geographic concentration scales according to the industry under consideration. Each sector exhibits spatial concentration peaks at different distances. However, their measure fails to satisfy one of the five criteria defined by DO, which is controlling for overall agglomeration. They set their benchmark as CSR for each sector in considering concentration. Albert, Casanova, and Orts (2012) conduct the same line of research for Spanish manufacturing firms. In their analysis they set another second benchmark to control for overall agglomeration and thus fulfils all five of DO criteria. As an extension to the K function method, Arbia, Espa, and Quah (2008) introduce *bivariate* K function in order to uncover co-agglomeration and repulsion phenomena between the different sectors. In another study (Arbia, Espa, Giuliani, and Mazzitelli, 2009), they take it a step further and consider temporal dynamics in clusters of economic activities. The process of firm demography is examined by studying the dynamics of localization through space-time K functions.

3.2.2 K Density

Duranton and Overman (2005)'s K density is considered one of the leading functions in spatial economics and their methodology has been widely applied in the field since their seminal paper (Marcon and Puech, 2012). In order to be able to be acquainted with, the main lines of the DO approach will be presented. The methodology may be reviewed in three major steps however for exhaustive details the reader may refer to the original article.

At the first step, having selected the relevant observations, they calculate the Euclidean distance between every pair of establishment, for an industry A with n establishments. One gets $\frac{n(n-1)}{2}$ unique bilateral distances in consequence. Then, they build the observed distribution of bilateral distances in an industry by using Gaussian kernel-smoothing estimator with a rule-of-thumb bandwidth á la Silverman (1986) (as cited in Duranton and Overman (2005)) by the following:

$$\widehat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d - d_{ij}}{h}\right) \tag{3.27}$$

where h is the bandwidth and f is the kernel function. In fact, this step fullfills their fourth requirement in terms of being unbiased with respect to spatial scale and aggregation.

As a second step, they construct counterfactuals where they consider hypothetical industries with the same number of establishments as have been observed. This step satisfies the first and third criteria about comparability across industries and the need to control for industrial concentration. They assume that any existing establishment, regardless of its industry, occupies one site and establishments are randomly allocated across these existing sites. So, since active sites are determined so as to regard all industries regardless of industry differentiation, it controls for overall manufacturing which satisfies their second criteria.

The counterfactuals are constructed by drawing locations from the population of sites and calculating the set of bilateral distances. The simulation is run by 1000 times for each industry and for each simulation sampling is made in accordance with the number of establishments in the industry under consideration.

Finally at the third step, they build local confidence intervals based on the simulation trials obtained at the second step in order to compare the actual distribution of distances to randomly generated counterfactuals. This step enables them to assess the significance of departures from randomness which also meets their fifth criteria. They use the the median value of bilateral distances between firms (180 km) as the upper limit of interval [0,180] for which they restrict their analysis in constructing confidence bands.

For each industry, for each kilometre in this interval simulations are ranked in ascending order and they select the 5-th and 95-th percentile to obtain a lower 5 % and an upper 5% confidence interval that are denoted $\overline{K}_A(d)$ and KA $\underline{K}_A(d)$, respectively (Duranton and Overman, 2005, p. 1086). For industry A, if $\widehat{K}_A(d) > \overline{K}_A(d)$, the industry is said to exhibit *localization at distance* d (at a 5% confidence level). For the reverse case where $\widehat{K}_A(d) < \underline{K}_A(d)$, the industry is said to exhibit *dispersion at distance* d (at a 5% confidence level). They also define an index of localization and dispersion as shown below, respectively:

$$\gamma_A(d) \equiv max(\widehat{K}_A(d) - \overline{K}_A(d), 0) \tag{3.28}$$

$$\psi_A(d) \equiv max(\underline{K}_A(d) - \widehat{K}_A(d), 0) \tag{3.29}$$

To reject the null hypothesis of randomness at distance d in favour of *localization* requires to have $\gamma_A(d) > 0$ while for *dispersion* it is $\psi_A(d) > 0$. Even though the exact value of these indices do not matter they indicate how much localization or dispersion occurs at different distance levels. However, while constructing local confidence intervals, bottom and top 5 % of the simulated densities are disregarded at each distance. "Taken together, large fraction (well above 5%) of the simulated densities is then dropped from the sample over the distance range" (Barlet, Briant, and Crusson, 2013, p. 348). They state that local confidence intervals are too restrictive to make any statements about the industry's global location patterns. To handle this issue they define global confidence bands in such a way that they "choose identical local confidence intervals at all levels of distance such that the global confidence level is 5%"(Duranton and Overman, 2005, p. 1087) [15].

[15] The precise construction of global confidence intervals requires a somehow detailed and complicated step-by-step procedure which is beyond the scope of this section. For details please refer to the original article

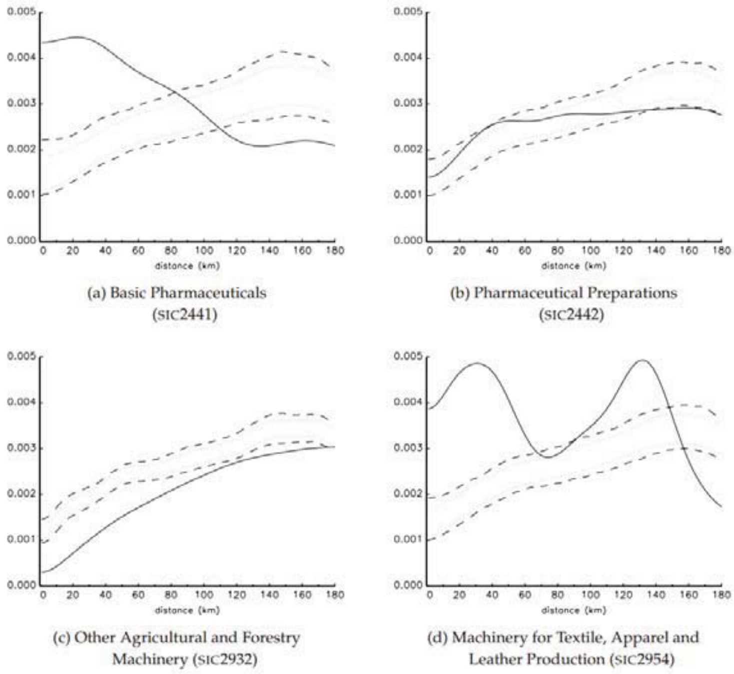


Figure 3.3: **K-density, local confidence intervals and global confidence band for four UK industries**

Figure 3.3 is borrowed from Duranton and Overman (2005, p. 1084) so as to provide a better understanding with the help of graphical illustration. DO state that, *global localization* is graphically detected whenever K-density of one particular industry lies above its upper confidence band for at least one distance over the range. On the contrary, in order to detect *global dispersion* K-density should lie below the lower confidence band and never above the upper confidence band. In terms of Figure 3.3(a)-(d) global confidence bands are represented by two dashed lines while local confidence intervals are shown by dotted lines.

For instance, case(d) exhibits global localization whereas (c) exhibits dispersion. On the other hand, case (b) shows neither global localization nor dispersion while case (a) shows global localization. The case (a) shows global localization by definition, not dispersion although its K-density does go beneath the lower confidence band after 110 km.

In a later study, Duranton and Overman (2008) extended their methodology to explore detailed patterns of manufacturing locations by studying location patterns of different groups of firms such as entrants vs. exiters, affiliated and non-affiliated plants, domestic and foreign plants, small vs. large plants etc. Following that, their methodology has been applied to many countries, for instance US auto industry (Klier and McMillen, 2008), Japan manufacturing and service industries (Nakajima, Saito, and Uesugi, 2012), German manufacturing and service industries (Koh and Riedel, 2014), Canadian manufacturing industries (Behrens and Bougna, 2015), etc. Significant contributions have also been made by recent studies to the DO methodology as well. Standing on the shoulders of the giants, Barlet, Briant, and Crusson (2013) have made some improvements over the DO methodology while Scholl and Brenner (2012) have developed a firm-level cluster index by criticizing enormous computational requirements of the DO index.

3.2.3 Discussion

Beyond doubt, distance-based measures are considered as more developed tools compared to their discrete counterparts in measuring the extent of geographic concentration phenomenon. They surpass discrete measures by means of defeating the aggregation and scale problems therein which has been the major criticism point for them. Especially, with the increasing availability of geo-referenced data in developed countries, distance-based methods have become to be utilised further in the last decade. The literature on localization using micro-geographic data, despite growing recently, is still very limited. Major difficulty with employing continuous measures arise as the data being too demanding. The case has been more desperate, particularly with the developing countries. For instance, almost every survey related to business activity in Turkey compiles the data of location information of an entity in a very fine scale. However, due to confidentiality issues they are not made available to researchers. Thus, data unavailability arises as the most significant drag on the application of continuous measures. Once data is available, continuous measures have also been criticized for requiring enormous computations. Bearing in mind today's developed computer technology this point may well be treated as outdated.

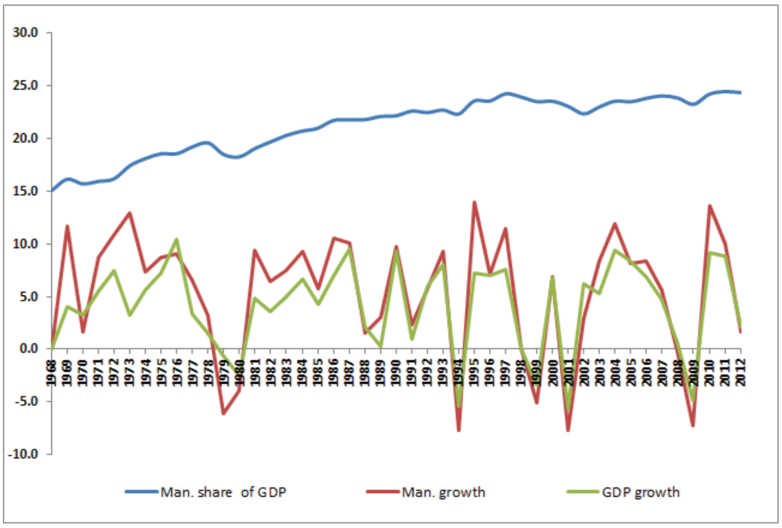
4 THE EXTENT OF AGGLOMERATION IN TURKISH MANUFACTURING INDUSTRY

This chapter analyses geographic distribution of manufacturing industries by employing *EG* index discussed in the previous chapter. Before analysing industrial agglomeration, it discusses the concentration of total production in order to display some evidence on the regional disparities in total production activity as a whole. Then it describes the dataset used in the study briefly. Section 4.3, which is the main part of the analysis, examines the extent of industrial agglomeration rigorously. Firstly, industry level agglomeration patterns at four digit industry classification according to NACE Rev. 1.1 are analysed. Following that, technology-wise agglomeration patterns are studied in order to see whether industries belonging to different technology groups exhibit different patterns. Next, industrial agglomeration over six years is examined to find out whether there is a trend in agglomeration or stable over time. Having examined agglomeration of four digit industries, the co-agglomeration behaviour within two-digit industries and between industry pairs are examined as well. Lastly, agglomeration patterns in Turkey are compared with other countries, as the index is designed to allow for comparison between industries, countries and over time,

4.1 The Concentration of Total Production

Manufacturing industry is considered as a fundamental indicator in the development process of a country. Rodrik (2007) remarks as a stylized fact that rapidly growing countries are those with large manufacturing sectors. Doğruel and Doğruel (2008, p. 65) indicate that there has been a significant increase in the share of Turkish manufacturing industry in total employment and GDP within the period 1970-2006. As Figure 4.1 shows the share of manufacturing sector within GDP has steadily been increasing over the last four decades, starting from 15.1 % in 1968 and reaching to 24.3 % in 2012.

As Doğruel and Doğruel (2008) mentions, 1970-2006 period of Turkish economy has been a period that witnessed many economic crises, important structural changes and diverse set of economic policies. Even after 2006, the last year of the period they considered, Turkish economy has experienced the negative impacts of the global financial crisis in 2008. The crisis periods resulted in economic downturns have also found its reflections in the manufacturing sector such that the growth of the sector has also taken negative values as shown in Figure 4.1. Out of recession periods, manufacturing growth is observed to be above GDP growth, while recessions have been more severe. They deduce that manufacturing industry has continually preserved its position as being the driving force of the economy except the crisis periods.



Source: TURKSTAT

Figure 4.1: **Share and growth of manufacturing sector (% of GDP)**

Despite the strength the manufacturing industry gained within this period, the distribution of manufacturing activity across the regions of Turkey has been far from being even. Doğruel, Doğruel, and Karahasan (2011) shows that regional distribution of manufacturing employment has remained same to a great extent between years 1985 and 2000 demonstrating a prominent diversification between eastern and western regions.In this regard, investigating the dispersion of manufacturing activity gives hints about understanding the regional disparities.

Regional specialization patterns might be used as a preliminary tool in order to observe regional variation in economic activity. As discussed in section 3.1.1, location quotient is the simplest measure of geographic concentration which basically compares the industry share in an area relative to its share in a wider area, i.e. country. Thus, location quotient above one (LQ>1) indicates that industry's share in regional economic activity exceeds industry's share in national economic activity and leaves room for the importance of industry to the region.

Location quotient for location i is defined as $LQ_i = \frac{s_i}{x_i}$, where s_i denotes location i's share of industry employment, and x_i denotes its share of total employment. Figure 4.2 shows LQ values for broadly defined Nace Rev. 1.1 sectors by NUTS-2 local units of Turkey [1]. There are 26 subregions according to NUTS-2 classification and the names of subregions are as shown in the map.

The picture of geographic concentration exhibited by broadly defined sectors exhibits some striking points. First consider manufacturing industry. In 2003, the shares of employment in manufacturing in Bursa and Tekirdağ regions are 58 and 38 percent greater than in the nation as a whole, respectively; such that these regions have LQs of 1.58 and 1.38. In contrast, Mardin region, has a manufacturing share of employment 69 per cent lower than the national share, with an LQ of 0.31. Moreover, it is observed that manufacturing activity is highly concentrated in western regions (İstanbul, Tekirdağ, İzmir, Manisa, Bursa, Kocaeli) plus Gaziantep region in 2003. However, İstanbul has experienced 17 percentage points, İzmir, Manisa and Gaziantep have experienced 2 percentage points decline in their LQ values when reached to 2008 indicating a decreasing level of regional specialization through the period. In addition to that, as Figure 4.3 illustrates, some regions increased specialization in manufacturing compared to 2003, like Tekirdağ, Bursa and Kocaeli. And, two regions, namely Kayseri and Zonguldak became geographically concentrated areas in terms of manufacturing activity. Holmes and Stevens (2004a) find a similar finding for the case of U.S that manufacturing activity is concentrated in a few divisions. They relate this to the fact that manufacturing sector has substantial possibilities for trade across regions.

[1] This classification pertains to Level 1 of NACE Rev. 1.1 defined by alphabetical characters from A to Q comprising of 17 parts. The missing parts, A-Agriculture, B-Fishing, J-Financial intermediation, L-Public administration and defence; compulsory social security, P-Activities of households and Q-Extra-territorial organizations and bodies, are not included due to data limitations.

	Nace Rev. 1.1.																					
	C		D		E		F		G		H		I		K		M		N		O	
	Mining and quarrying		Manufacturing		Electricity, gas and water supply		Construction		Wholesale and retail trade, repair of motorvehicles, etc.		Hotels and restaurants		Transport, storage, communication		Real estate, renting and buss. act.		Education		Health and social work		Other community, social and pers. serv. act.	
NUTS2	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008
TR10-Istanbul	0.16		1.22	1.05	0.51		0.95	0.88	0.87	0.98	0.72	0.83	0.91	0.92	1.27	1.28	1.11	1.08	1.09	1.13	0.93	1.11
TR21-Tekirdag	0.97	1.34	1.38	1.67	0.77		0.48	0.76	0.85	0.77	1.08	0.78	0.83	0.70	0.64	0.57		0.78	0.76	0.64	0.79	0.57
TR22-Balikesir	1.94	2.01	0.77	0.92	0.72		0.76	0.94	1.16	1.08	1.66	1.50	1.04	1.03	0.75	0.55	0.97	0.99	0.73	0.71	0.86	1.08
TR31-Izmir	0.50		1.10	1.08	0.73		1.09	0.99	0.99	1.02	0.90	0.82	0.89	0.85	1.05	1.14	0.86	0.90	0.77	1.01	0.78	0.90
TR32-Aydin	1.58	1.96	0.97	0.87	1.14		0.79	0.71	0.88	0.97	2.24	2.32	0.82	1.02	0.57	0.65		1.10	1.09	0.94	1.21	0.86
TR33-Manisa	4.84		1.02	1.00	1.42		0.71	0.60	0.97	0.83	1.05	0.97	0.78	0.95	0.72	0.58	0.92	0.79	0.84		1.23	0.86
TR41-Bursa	0.75	0.95	1.58	1.63	0.71		0.51	0.73	0.75	0.75	0.69	0.83	0.68	0.65	0.73	0.78	0.81	0.72	0.94	0.59	0.85	0.88
TR42-Kocaeli	0.39		1.21	1.46	0.86		0.93	0.93	0.97	0.81	0.94	0.66	0.81	1.13	0.71	0.66	0.84	0.59	0.67	0.64	0.92	0.79
TR51-Ankara	1.15	1.15	0.65	0.66	1.57		2.47	1.88	0.96	0.87	0.84	1.05	1.06	0.94	1.75	1.64	1.50	1.48	1.34	1.04	1.03	1.13
TR52-Konya	1.06	1.12	0.98	1.01	1.24		0.55	0.58	1.14	1.22	0.78	0.81	0.99	1.00	0.96	0.69	0.63	0.98	0.90	1.09	1.13	0.87
TR61-Antalya		0.35	0.41	0.44	0.74		1.22	1.23	1.13	1.11	2.82	2.89	1.05	0.91	1.08	0.97	0.97	0.79	1.23	1.10	1.40	1.20
TR62-Adana	0.36	0.74	0.72	0.71	0.81		0.97	0.92	1.28	1.21	0.78	0.77	0.99	1.23	1.16	1.01	1.03	1.27	1.38	1.55	1.49	1.03
TR63-Hatay	0.14	0.37	0.84	0.98	2.44		0.86	0.85	1.17	1.18	0.82	0.82	1.17	1.18	0.56	0.54	1.92	0.73	1.18	0.85	0.95	0.91
TR71-Kirikkale	0.47		0.71	0.88	1.96		0.68	1.29	1.20	1.13	1.29	1.13	1.37	1.03	0.70	0.62	0.70	0.95	0.90	0.60	0.99	1.04
TR72-Kayseri	1.05	1.87	0.87	1.02	1.22		0.76	1.81	1.00	0.98	0.52	0.69	2.25	0.79	0.57	0.73	0.55	1.05	0.82	0.77	0.61	0.91
TR81-Zonguldak	12.12	11.93	0.75	1.06	1.14		0.54	0.70	1.00	0.89	1.43	1.03	0.00	0.00	0.36	0.39		1.00	0.73	0.96	1.03	0.87
TR82-Kastamonu	1.33	1.39	0.65	0.99	1.36		0.66	1.32	1.16	0.95	0.98	0.65	2.01	1.27	0.66	0.82	0.49	0.62	0.54	1.06	1.09	1.20
TR83-Samsun	1.15		0.77	0.86	1.42		0.53	0.76	1.36	1.25	1.01	0.87	0.93	1.23	0.66	0.70	0.99	0.76	0.96	0.95	1.01	0.94
TR90-Trabzon	1.42		0.68	0.91	1.19		0.91	1.47	1.39	1.55	0.98	1.35	1.19	1.71	0.82	0.91	0.87	1.31	0.95	0.85	1.12	1.02
TRA1-Erzurum	0.44	0.98	0.40	0.39	2.73		0.91	0.73	1.50	1.43	1.22	0.64	1.48	1.96	0.43	0.71	0.49	1.04	0.70	1.22	1.24	0.53
TRA2-Agri			0.35	0.46	2.19		0.31	0.27	1.49	1.74	1.82	1.20	1.82	1.42	0.45	0.30	0.57	0.52				0.99
TRB1-Malatya	1.52		0.65	0.78	2.12		0.96	1.34	1.22	1.16	0.76	0.72	1.31	1.15	1.00	0.71	1.67	1.24		1.18	1.11	0.86
TRB2-Van	0.11	0.20	0.42	0.51	2.13		1.15	1.26	1.40	1.27	1.42	0.61	1.48	1.86	0.70	0.53	0.96	1.30	0.23	1.77	1.12	0.81
TRC1-Gaziantep	0.77		1.19	1.17	1.20		0.60	0.79	1.06	1.09	0.82	0.80	0.70	0.92	0.62	0.51		0.55	1.18	1.45	1.25	1.40
TRC2-Sanliurfa	0.19	0.92	0.51	0.45	1.72		1.62	1.25	1.34	1.20	0.99	0.54	1.45	1.73	0.72	1.27	0.91	0.80	0.83	1.22	1.03	0.98
TRC3-Mardin	4.93	3.55	0.31	0.33	3.11		1.18	0.70	1.34	1.13	0.70	0.97	2.04	2.63	0.45	0.61	0.99	1.58	0.75	0.97		0.86

Figure 4.2: **Employment Location Quotients by broadly defined sectors**

Source: Author's calculations based on employment data from Annual Industry and Service Statistics provided by Turkish Statistical Institute (TURKSTAT). Empty cells are due to the data unavailability. Coloured cells show: yellow if 1<LQ<1.25, green if 1.25<LQ<2 and red if LQ>2

On the other hand, possibilities for trade regarding service sectors are relatively limited. For example, many regions have LQs above 1 in wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods sector (G) in line with transport, storage and communication (I) sector and not incidentally, these regions have very low levels of LQ, in other words they are lagging behind the national share of manufacturing. It is straightforward to see that regional specialization in mining and quarrying (C) industry is majorly driven by "first nature", termed by Krugman (1993), referring to given spatial distributions of natural endowments, technologies and/or factors. Red highlighted cells in this sector correspond to regions which are quite rich in terms of mineral reserves, i.e. Balıkesir for boron, Manisa for lignite, Zonguldak for mineral coal and Mardin for petroleum extraction are pretty well known regions. Hotels and restaurants (H) industry, not surprisingly, has concentrated in Aydin and Antalya region to a vast degree as they arise two major tourism centers of the country. İstanbul and Ankara subregions (which are only composed of themselves as İstanbul and Ankara province) are specialized in real estate, renting and business activities (K) industry as they stand out as the most populated and developed (in terms of urbanization) regions among others.

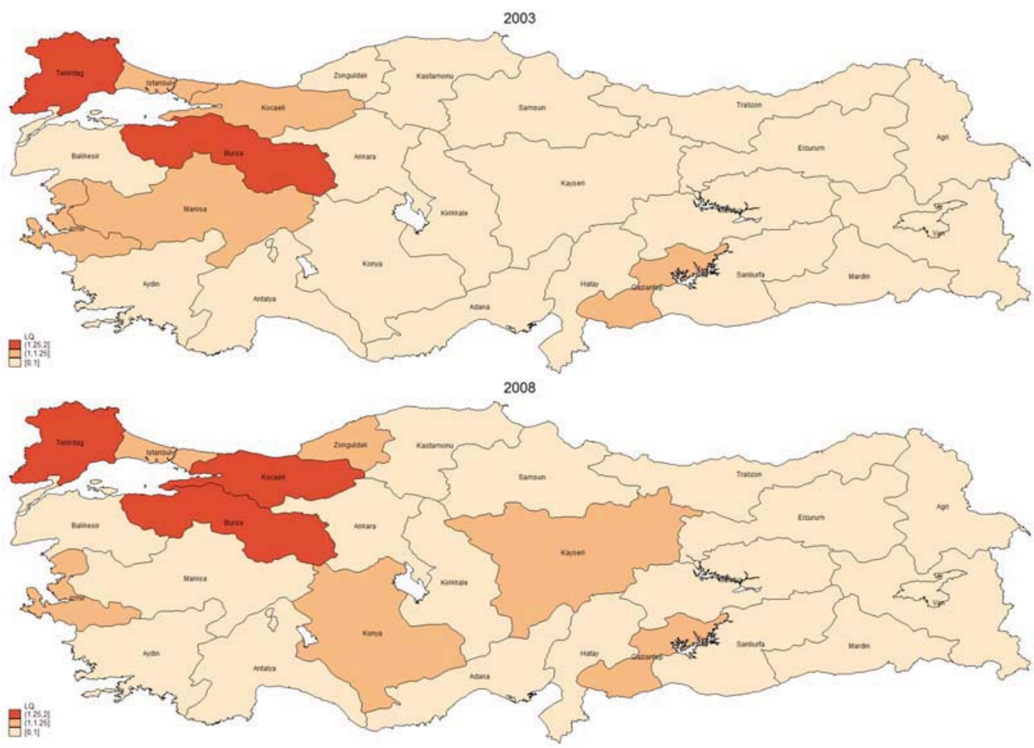


Figure 4.3: **Manufacturing employment LQs by NUTS-2 regions**

Shifting our focus specifically to manufacturing industry, concentration of total production figures verify the inferences derived from regional specialization patterns regarding manufacturing. Figures 4.4 and 4.5 show the distribution of total manufacturing employment by Turkey's NUTS-2 regions for 2003 and 2008. As figure 4.4 illustrates, İstanbul (TR10), İzmir (TR31), Bursa (TR41) and Ankara (TR51) are the four subregions with the highest proportion of total manufacturing employment in 2003. Considering 2008, Kocaeli (TR42) emerges as an additional region to those mentioned in 2003. These four subregions together represent around 58.5 % of the total production employment where İstanbul (TR10) constituting the largest portion with 35.3 % in 2003. And in 2008, this share is met by five regions and reported as 63.1 %, where again İstanbul (TR10) dominates with a share of 32.8 %.

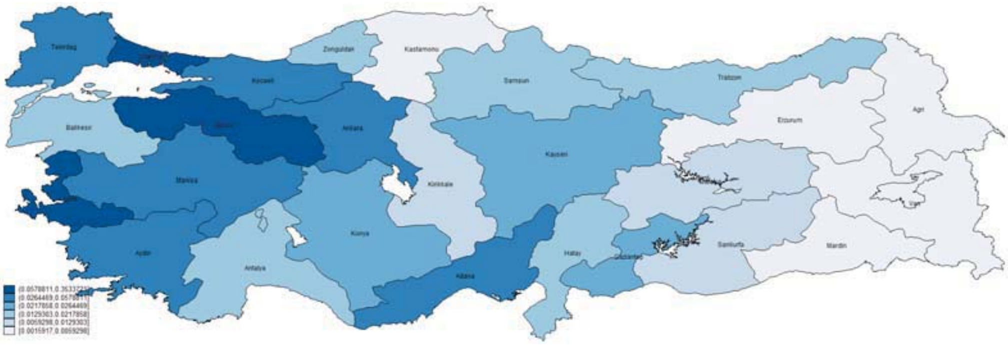


Figure 4.4: **Regional share of manufacturing employment - 2003**

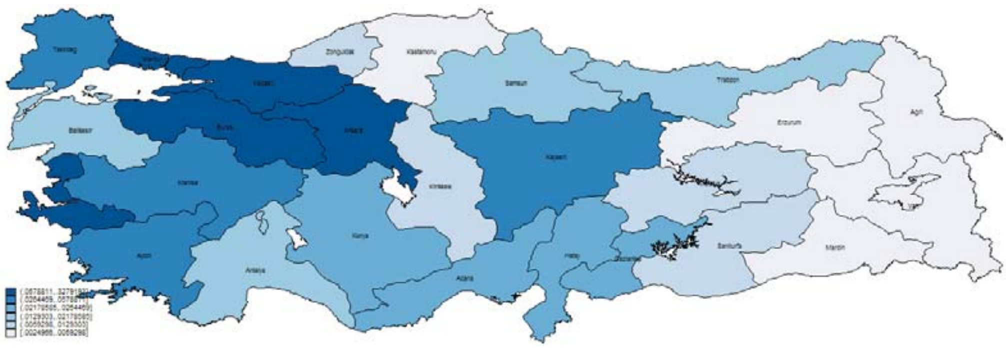


Figure 4.5: **Regional share of manufacturing employment - 2008**

Referring to Table 3 in Devereux, Griffith, and Simpson (2004, p. 541), Table 4.1 shows different measures calculated for total manufacturing activity - γ is EG agglomeration index defined in equation (3.17), G is measure of raw geographic concentration defined in equation (3.15) and H stands for the industrial concentration measure. Along with these, locational Gini coefficient [equation(3.3)] and Moran's I index of spatial autocorrelation are shown. The measures are constructed relative to a uniform distribution with the purpose of examining the aggregate distribution. The uniform distribution is set as $x_m = 1/M$ where M is the total number of regions. Raw geographic concentration measure G and locational Gini indicate that the total production is geographically concentrated. The value of γ, index of agglomeration, also reflect the fact that overall production activity is highly agglomerated. However, it is observed that there is a decline in all three measures throughout the period which means that the concentration of total production activity slightly lowered. On the other hand, an interesting point arises when it comes to consider Moran's I statistic. It reveals is that total production activity exhibits spatial dependence in either year (both significant at $\alpha = 0.05$). In other words, total production activity tends to be clustered in space and in 2008 it became even more clustered. What all these measures tell is fairly approved by the maps shown in Figure 4.4 and 4.5. Total production activity obviously exhibits a clustering phenomenon on the western part of the country.

Table 4.1: **Concentration measures for total production**

	2003	2008
H	0.0002	0.0002
G	0.1174	0.1023
γ	0.1218	0.1062
Gini	0.3179	0.3081
Moran's I	0.090	0.132

To have an idea about the geographic concentration of narrowly defined industries before analysing deeper, location quotients are employed once more specific to 4-digit industries. Likewise in broadly defined sectors, regional share of a specific industry is compared with its national share, indicating a concentration in the respective industry if it takes a value of greater than 1. In 2003, there are 234 four-digit manufacturing industries operating actively. Keeping in mind that there are 26 subregions, we should have 6084 (234X26) unique location quotients. However, as may be expected, many regions have only a limited number of industries, which means that regional share of a specific industry would be observed as zero in many regions, hence LQ will be calculated as 0. Almost 42 % of the location quotients are calculated as zero. Furthermore, roughly 36 % of them are computed to be smaller than 1 (LQ<1), leaving no room for geographic concentration. Thus, remaining 22 % of them are to be considered for localization.

Table 4.2: **Number of industries by location quotient quartiles and NUTS-2 regions, 2003**

	LQ>1					LQ<1	Total
	1st	2nd	3rd	4th	Num.ind.	Num.ind.	
TR10-İstanbul	40	48	23	0	111	108	219
TR21-Tekirdağ	13	13	12	11	49	101	150
TR22-Balıkesir	10	8	15	19	52	84	136
TR31-İzmir	39	19	19	16	93	105	198
TR32-Aydin	14	11	12	5	42	107	149
TR33-Manisa	17	21	16	18	72	93	165
TR41-Bursa	18	27	16	8	69	124	193
TR42-Kocaeli	19	23	25	22	89	96	185
TR51-Ankara	29	18	25	21	93	99	192
TR52-Konya	10	11	19	20	60	97	157
TR61-Antalya	12	16	20	17	65	87	152
TR62-Adana	14	13	18	13	58	106	164
TR63-Hatay	4	16	9	7	36	90	126
TR71-Kırıkkale	9	8	13	15	45	78	123
TR72-Kayseri	6	13	16	10	45	102	147
TR81-Zonguldak	3	7	10	6	26	81	107
TR82-Kastamonu	9	6	5	20	40	54	94
TR83-Samsun	15	12	10	15	52	100	152
TR90-Trabzon	9	9	8	8	34	86	120
TRA1-Erzurum	6	7	7	13	33	50	83
TRA2-Agri	2	1	6	9	18	32	50
TRB1-Malatya	4	3	10	13	30	71	101
TRB2-Van	2	7	1	11	21	36	57
TRC1-Gaziantep	17	8	8	8	41	111	152
TRC2-Sanliurfa	6	5	4	16	31	70	101
TRC3-Mardin	7	3	6	12	28	37	65
Total	334	333	333	333	1,333	2205	3538

Note:Quartile boundries are by LQ and defined as; 1:(1.0007, 1.3828), 2:(1.3831, 2.0134), 3:(2.0139, 3.5999), 4:(3.6103, 149.5721).

Table 4.2 shows the distribution of the number of industries according to LQ values and NUTS-2 regions for the year 2003 [2]. LQ is grouped into four quartiles whenever observed to be greater than the reference value, 1. The fifth column displays the number of industries in which LQ>1 and sixth column shows the ones with LQ<1. As last column shows, İstanbul, İzmir, Bursa and Ankara regions have the greatest number of industries displaying an industrial diversity, however roughly half of them are below the national shares. Adhering to quartile boundaries defined in the table note, there are 666 industry-location pairs in the third and fourth quartile signalling for geographic concentration. However, one should be very cautious while evaluating location quotients due to its inefficacy in purging industrial concentration from geographic concentration as discussed in section

[2]The table for 2008 is not reported here but available upon request.

3.1.7. For instance industry 1592 has only one firm in the entire country and it's located in İzmir with a Herfindahl of 1 and LQ of 12.75. Without knowing the industrial concentration measure and solely focusing in LQ, we would conclude that the sector is geographically concentrated, which is really not the case. Moreover, since LQ or its variants are not derived from a location choice model, the results revealed by them does not distinguish whether geographic concentration arises due to a random process or not. Thus, in order to get accurate inferences about geographic concentration of industries, more theoretically grounded measures should be employed, which will be held in the following sections.

4.2 Data

This study explores establishment-level micro-data called "Annual Industry and Service Statistics" (AISS) provided by TurkStat. The sectoral coverage of the dataset encompasses all the manufacturing and service industries with a few exceptions [3]. The number of establishments covered in the survey ranges from 125000 to 144850 as shown in Table 4.3 throughout the sample period. However the focus of this study is the sub-sample of manufacturing firms as the purpose is examining the trends of industrial agglomeration. The number of manufacturing firms covered in the sample ranges from 39700 to 48000. We observe an upward trend in the dataset as the number of manufacturing firms have been growing over the sample period, but there is a decrease in the number of establishments from 2004 to 2005 due to a change in the sampling procedure. The number of 4-digit industries included in the dataset are roughly around 240, but for the sake of computing the EG index properly industries with less than three employment are dropped and we are left with the resulting number of industries shown in the third row. NACE Rev. 1.1 classification is used as a statistical classification of economic activities for the years 2003-2008 which is listed in Table AII.1 in Appendix II [4]. Since 2009 NACE Rev. 2 is in use. Due to the comparability problems that are likely to arise as a result of switching from NACE Rev. 1.1 to NACE Rev. 2, the analysis is restricted to consider only NACE Rev. 1.1 classification and hence to the period 2003-2008.

[3] The sector codes of non-covered sectors according to NACE Rev. 2 are A, K, O, T, U, a division of J ("Programming and broadcasting activities") , a division of S ("Activities of membership organisations") and two classes of L, namely "Buying and selling of own real estate" and "Renting and operating of own or leased real estate".

[4] NACE stands for **N**omenclature statistique des **A**ctivités économiques dans la **C**ommunauté **E**uropéenne in French. It means Statistical Classification of Economic Activities in The European Community.

Table 4.3: **Sample size of the dataset**

Number of establishments / year	2003	2004	2005	2006	2007	2008
The original data including all manufacturing and service	125003	125652	109397	137481	144057	144849
The data set of manufacturing firms only	39714	43958	36051	45908	47532	48024
Number of 4-digit industries	228	228	229	227	228	229
Average employment per establishment	54.5	54.4	71.8	58.4	58.4	59.6

Studies of economic geography requires to have information on firm location. However, the choice of geographic unit is likely to affect the measure of industrial agglomeration - the so called modifiable area unit problem (MAUP). The extent of geographic breakdown chosen in studies of agglomeration ranges from state, region, county, province, district and postcode level to even more precise location data forming the base for continuous space modelling (e.g. Arbia (2001a), Duranton and Overman (2005)) to minimize the MAUP. Unfortunately, there is no available data regarding Turkish manufacturing firms in such fine locational units, but instead we have information about the location information of establishments on a very rough scale.

In fact TurkStat provide establishment level data in two bases. One of them provides quiet detailed information about the employment, expenditure, income, stocks and investment structure of establishments based on the responses collected by the questionnaires. This dataset covers numerous variables broadly such as: (i) employment, hours worked and payments (ii) expenditure and its items in detail (iii) income and its items in detail (iv) stocks (excluding VAT) (v) investment (tangible, intangible, VAT paid for fixed capital, value of fixed capital sales, depreciation) (vi) capital shares of the enterprise (public, private, foreign) and if foreign it's share, (vii) 4-digit NACE activity area and (viii) number of local units.

In this dataset, each enterprise is identified by a registration number. However, it does not provide information on the *location* of the enterprise because an enterprise may have more than one plant and each having only one identifier, then it becomes a data management issue which plant's location to display. Thus, this dataset provide detailed information on the aggregated plants of an enterprise as a whole. Additionally, it provides information on the *number of local units* of an enterprise. Number of local units shows how many different plants a particular enterprise has and that's all. If one wants to get more information on the individual plant records of a particular enterprise, then should resort to the second dataset provided by TurkStat, which conveys information about the *local units* . In *local units* dataset, now there is another identifier, besides registration number of an enterprise, which identifies the plants of an enterprise. For instance, if an enterprise has five local units displayed in the first dataset, we may reach information on that five plants in the *local units* dataset. It is the *local units* dataset which provides information on the location of the plants in the NUTS-2 geographical level which covers 26 regions in Turkey. NUTS of Turkey is listed in Table AII.2 in Appendix II [5]. One drawback of this dataset is that it provides very limited information relative to the first one. There are only 6 key variables shown in the dataset, namely: (i) NACE 4-digit economic activity area, (ii) employment, (iii) wages and salaries, (iv) turnover, (v) gross investment in tangible goods and (vi) location in NUTS-2 level, as already mentioned.

So, *local units* dataset arises as an appropriate dataset for calculating *EG* index Turkish manufacturing industry, providing all the necessary information on employment at plant, industry and geographical levels.

Table 4.4 shows the distribution of manufacturing firms in terms of number of establishments and employment by 2-digit sectors for years 2003 and 2008. As observed in the table, the sectors food products and beverages (15), textiles (17) and wearing apparel, dressing & dyeing of fur (18) comprise almost 40 and 38 per cent of all establishments for 2003 and 2008, respectively. In terms of employment, these sectors comprise 46 and 38.5 per cent of the overall employment in manufacturing, again for the years 2003 and 2008, respectively. It's striking to note that these sectors take place within the low-technology group according to OECD technology classification.

Figure 4.6 graphically illustrates the distribution of establishments and employment by sectors shown in Table 4.4. Figure 4.6a shows that number of establishments has increased in all 2-digit sectors through 2003 to 2008, except the sector tobacco products (16). Although the number of firms has absolutely increased across sectors over the period, namely from 39,714 to 48,024 which means almost 21 % growth, the share of firms across industries exhibit different patterns. For instance the share of textile (17) and food and bevarage (15) firms has declined while the share of wearing apparel, dressing & dyeing of fur (18) firms has increased as seen in 4.6b. Employment has also increased in absolute terms across 2-digit industries as 4.6c clearly displays. Overall employment of manufacturing industries in the dataset has increased from 2,163,663 in 2003 to 2,863,616 in 2008, which corresponds to an almost 32 % of employment growth. Accordingly, figure 4.6d shows the change in the shares of employment by sectors throughout the period. A salient feature of this graph arises as the decline in

[5]NUTS in French abbreviated for Nomenclature des Unités Territoriales Statistiques is a geocode standard for referencing the subdivisions of countries for statistical purposes. It means Nomenclature of Territorial Units for Statistics.

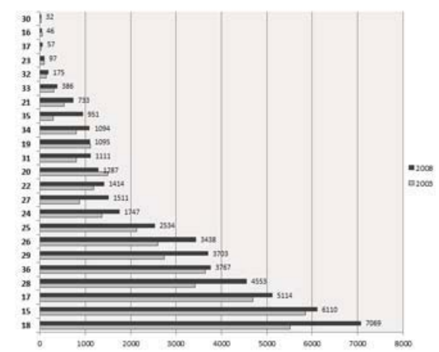
Table 4.4: **Establishment and employment distribution by sector(%)**

Industry	Establishments 2003	Establishments 2008	Employment 2003	Employment 2008
15 Food products and beverages	14.74	12.72	12.50	11.44
16 Tobacco products	0.14	0.10	1.02	0.65
17 Textiles	11.81	10.65	18.36	13.13
18 Wearing apparel, dressing & dyeing of fur	13.87	14.72	15.75	14
19 Tanning and dressing of leather	2.79	2.28	1.91	1.77
20 Wood products, except furniture	3.77	2.68	2.43	2.50
21 Pulp, paper and paper products	1.36	1.53	1.39	1.51
22 Publishing, printing & rep. of recorded media	2.97	2.94	2.03	2.48
23 Coke, refined petroleum products & nuclear fuel	0.24	0.20	0.29	0.22
24 Chemicals and chemical products	3.44	3.64	3.60	2.92
25 Rubber and plastic products	5.39	5.28	4.33	5.22
26 Other non-metallic mineral products	6.54	7.16	5.45	6.27
27 Basic metals	2.2	3.15	3.37	3.56
28 Fabricated metal products, except machinery & equipment	8.61	9.48	6.08	8.16
29 Machinery and equipment	6.91	7.71	6.52	8.17
30 Office machinery and computers	0.06	0.07	0.03	0.04
31 Electrical machinery and apparatus	2.04	2.31	2.34	2.83
32 Radio, TV and communication equipment & apparatus	0.34	0.36	0.89	0.65
33 Medical, precision and optical instruments, watches & clocks	0.77	0.80	0.55	0.87
34 Motor vehicles, trailers & semi-trailers	2.01	2.28	3.86	4.77
35 Other transport equipment	0.74	1.98	1.05	2.15
36 Furniture; manufacturing n.e.c.	9.2	7.84	6.21	6.61
37 Recycling	0.06	0.12	0.02	0.04

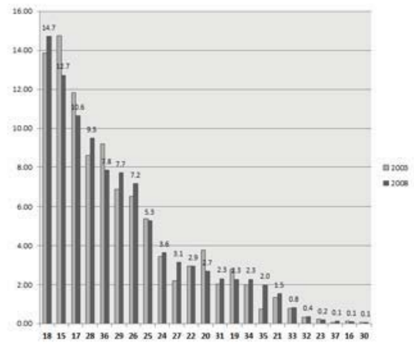
the employment shares of industries (18), (17) and (15) despite still being among the three mostly employed sectors across 2-digit industries. It is also clear that to a great extent the share of other sectors in manufacturing employment has increased over this period.

As we are dealing with micro data, it's natural to have a look at the size distribution of firms. Tables 4.5 and 4.6 show the size distribution of firms by sectors based on the new definition of the Small and Medium Sized Enterprises (SMEs)[6] for the years 2003 and 2008, respectively. According to this definition establishments employing 0-9 workers are called *micro*, 10-49 workers are called *small*, 50-249 workers are called *medium* and employing over 250 workers are called *large* sized enterprises. Based on this definition, 32 % of the establishments are micro sized within the universe of plants where small and medium sized plants

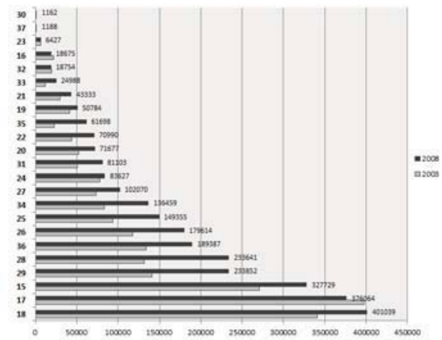
[6] "Directive regarding the amendment of the Directive on the definition, qualification and classification of the Small and Medium Sized Enterprises" came into force as of November 4, 2012 and broadcasted in official journal no.28457 http://www.resmigazete.gov.tr/eskiler/2012/11/20121104-11.htm



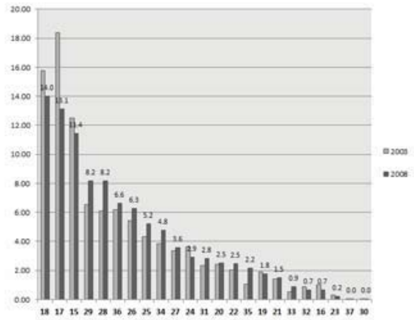
(a) Number of establishments



(b) Per cent of establishments



(c) Number of employed



(d) Per cent of employment

Figure 4.6: **Establishments and employment by sectors**

constitute 41.6 % and 23.1 %, respectively. In total, SMEs compose a significant portion of manufacturing industries, namely 96.7 %. Large enterprises constitutes only 3.3 % of them. Regarding year 2008, the share of micro sized establishments decreased by almost 2 percentage points where small sized establishments' share increased by 4 percentage points and realized as 29.9 and 45.6 %, respectively. The share of medium sized enterprises dropped by almost 3 percentage points, to 20.4 % and large sized ones' share observed as 4.1 % with a small increase.

If we were to consider industry-wise size distribution of establishments, roughly the same pattern emerges with the overall distribution. However, sector (16)-manufacture of tobacco products, stands out as a outlier among other sectors in both years. In this sector, 57.9 % of the establishments are large in 2003, employing more than 250 workers. Nevertheless, in 2008 this share has decreased to 32.6 % in favor of micro and small sized establishments where their shares increased significantly by 27.3 and 12.6 percentage points respectively. In addition, the number of firms in this sector is quite low in both years, even lower in 2008 than 2003, which hints on a high degree of industrial concentration. Sectors (15) - manufacture of food products and beverages, (17) - textiles and (18) - manufacture of wearing apparel; dressing and dyeing of fur have the greatest

Table 4.5: **Size distribution by sector - 2003**

	Size of the firm									
NACE 2-digit	**micro**		**small**		**medium**		**large**		**Total**	
	N	Row %	No.	Row %	N	Row %	N	Row %	N	Row %
15	2445	41.8	2014	34.4	1237	21.1	158	2.7	5854	100.0
16	3	5.3	4	7.0	17	29.8	33	57.9	57	100.0
17	1203	25.6	1820	38.8	1335	28.5	334	7.1	4692	100.0
18	1451	26.3	2355	42.7	1486	27.0	217	3.9	5509	100.0
19	405	36.5	442	39.9	254	22.9	8	0.7	1109	100.0
20	346	23.1	889	59.3	255	17.0	8	0.5	1498	100.0
21	155	28.8	216	40.1	147	27.3	21	3.9	539	100.0
22	426	36.1	512	43.4	218	18.5	23	2.0	1179	100.0
23	42	43.3	41	42.3	9	9.3	5	5.2	97	100.0
24	565	41.4	495	36.2	244	17.9	62	4.5	1366	100.0
25	791	37.0	830	38.8	480	22.4	39	1.8	2140	100.0
26	882	34.0	1110	42.8	530	20.4	74	2.9	2596	100.0
27	227	26.0	395	45.2	206	23.6	46	5.3	874	100.0
28	982	28.7	1645	48.1	749	21.9	42	1.2	3418	100.0
29	809	29.5	1214	44.2	645	23.5	76	2.8	2744	100.0
30	14	56.0	7	28.0	4	16.0	0	0.0	25	100.0
31	208	25.7	356	44.0	215	26.5	31	3.8	810	100.0
32	35	25.9	58	43.0	33	24.4	9	6.7	135	100.0
33	102	33.6	132	43.4	67	22.0	3	1.0	304	100.0
34	152	19.0	329	41.2	264	33.1	53	6.6	798	100.0
35	78	26.4	116	39.3	88	29.8	13	4.4	295	100.0
36	1378	37.7	1535	42.0	697	19.1	42	1.2	3652	100.0
37	9	39.1	12	52.2	2	8.7	0	0.0	23	100.0
Total	12708	32.0	16527	41.6	9182	23.1	1297	3.3	39714	100.0

number of establishments within 2-digit sectors and the number of firms in these sectors increased by 4, 9 and 28 %, respectively where sector (18) experienced the highest increase. In there of these, what is observed significantly is that the share of medium sized enterprises increased over the period. Another striking point arising from the tables is that sectors (30) - manufacture of office machinery and computers and (37) - recycling do have very limited number of firms in both years, even increased by 28% and 100% over the period respectively. In 2003, 56 % of the establishments are micro sized in sector (30) while at the end of the period 65.6% of the establishments were small sized. Regarding sector (37), 52.2 % of the firms were small sized and dominant in group in 2003 while this figure changed in favour of micro sized enterprises with a share of 57.9%. As well as the transition between size categories within each 2-digit sector may be detected by simply comparing shares, an analysis based on firm dynamics would reveal far more exhaustive information which is beyond the scope of this research.

Table 4.6: **Size distribution by sector - 2008**

NACE 2-digit	Size of the firm									
	micro		**small**		**medium**		**large**		**Total**	
	N	Row %	No.	Row %	N	Row %	N	Row %	N	Row %
15	2167	35.5	2571	42.1	1119	18.3	253	4.1	6110	100.0
16	15	32.6	9	19.6	7	15.2	15	32.6	46	100.0
17	1280	25.0	2319	45.3	1198	23.4	317	6.2	5114	100.0
18	2273	32.2	3139	44.4	1362	19.3	295	4.2	7069	100.0
19	368	33.6	503	45.9	195	17.8	29	2.6	1095	100.0
20	223	17.3	720	55.9	313	24.3	31	2.4	1287	100.0
21	167	22.8	364	49.7	178	24.3	24	3.3	733	100.0
22	482	34.1	612	43.3	265	18.7	55	3.9	1414	100.0
23	36	37.1	47	48.5	10	10.3	4	4.1	97	100.0
24	648	37.1	766	43.8	269	15.4	64	3.7	1747	100.0
25	655	25.8	1212	47.8	570	22.5	97	3.8	2534	100.0
26	988	28.7	1605	46.7	738	21.5	107	3.1	3438	100.0
27	589	39.0	605	40.0	246	16.3	71	4.7	1511	100.0
28	1199	26.3	2230	49.0	990	21.7	134	2.9	4553	100.0
29	1048	28.3	1781	48.1	710	19.2	164	4.4	3703	100.0
30	7	21.9	21	65.6	3	9.4	1	3.1	32	100.0
31	263	23.7	534	48.1	267	24.0	47	4.2	1111	100.0
32	51	29.1	77	44.0	39	22.3	8	4.6	175	100.0
33	119	30.8	176	45.6	75	19.4	16	4.1	386	100.0
34	191	17.5	516	47.2	292	26.7	95	8.7	1094	100.0
35	188	19.8	494	51.9	228	24.0	41	4.3	951	100.0
36	1372	36.4	1581	42.0	707	18.8	107	2.8	3767	100.0
37	33	57.9	19	33.3	4	7.0	1	1.8	57	100.0
Total	14362	29.9	21901	45.6	9785	20.4	1976	4.1	48024	100.0

4.3 How Much Are Industries Agglomerated?

The patterns of geographic concentration in Turkish production activity is examined in this section. The section mainly attempts to address the following points:

- What is the extent and general trend of agglomeration in Turkish manufacturing industries in the post-2000 period? The focus of the analysis is to identify agglomeration before attempting to answer further questions such as the main determinants of agglomeration.
- The stylized fact that low-tech industries tend to be more agglomerated than high-tech ones also hold for the case of Turkish manufacturing industries?
- Is agglomeration stable over time or is there a significant trend?
- Does agglomeration differ between different technology groups?

By making use of AISS local units dataset, EG index is calculated for Turkish manufacturing industries for the year 2003. In our sample there are 228 and 229 4-digit industries in 2003 and 2008 respectively in terms of NACE Rev. 1.1. Industries with less than three plants are dropped due to the fact that EG index tends to be biased upwards, as mentioned before.

In the simple dartboard model of EG in which plants choose their location in a random manner, in the absence of natural advantages or spillovers, expected value of the raw geographic concentration should be proportional to the industrial concentration. To state algebraically, $E(G) = \left(1 - \sum_m x_m^2\right)H$.

So, as a first step it should be tested whether observed geographic concentration G, is statistically significantly different from $\left(1 - \sum_m x_m^2\right)H$. For the year 2003, the mean values for the above expressions across 228 manufacturing industries are calculated to be 0.17 and 0.09, respectively and the difference between these measures is highly significant. Regarding 2008, these values are calculated to be 0.16 and 0.08, respectively across 229 manufacturing industries and the difference is again highly significant [7]. To be more precise, for the year 2003, in 191 out of 228 and for the year 2008, in 188 out of 229 4-digit manufacturing industries, the level of raw geographic concentration is found to be exceeding the value which would be obtained in the case of a random location choice. Regarding year 2003, the difference between G and $\left(1 - \sum_m x_m^2\right)H$ is positive and larger than twice its standard deviation in 176 of the 191 4-digit industries, hence the difference statistically significant, while 15 of them are not significant despite being positive. For the year 2008, the difference between G and $\left(1 - \sum_m x_m^2\right)H$ is positive and larger than twice its standard deviation in 182 of the 188 4-digit industries, confirming statistical significance, while 6 of them are not significant despite being positive. In 37 and 41 of the 4-digit industries, the difference is found to be negative, for 2003 and 2008 respectively. It's worth mentioning as a striking point that these industries exhibit negative values of EG as well. This would indicate that negative agglomeration indices are far from being statisti-

[7]Ellison and Glaeser (1997) provide a lengthy formula for the variance of G under the null hypothesis of no natural advantages and spillovers, as follows:

$$var(G) = 2\left\{H^2\left(\sum x_m^2 - 2\sum x_m^3 + (\sum x_m^2)^2\right) - \sum_j z_j^4\left(\sum x_m^2 - 4\sum x_m^3 + 3(\sum x_m^2)^2\right)\right\}$$

Under this formula, the standard deviation of the sample mean under the null is calculated as 0.004 and 0.003, as regards 2003 and 2008 respectively.

cally significant, hence displaying a random distribution across space rather than a dispersion. Hence, the null hypothesis of a random location choice can not be rejected for those industries.

4.3.1 Industry Level Agglomeration

Figure 4.7 shows the distribution of γ over 228 and 229 4-digit industries for years 2003 and 2008, respectively. Regarding year 2003, the mean value of γ is 0.1060 with a median value of 0.0675 for Turkish manufacturing industries. As for 2008, the mean and the median values are calculated as 0.1046 and 0.0602, respectively. The two distributions in the figure appear to be slightly right skewed depicting the agglomeration at higher levels. Ellison and Glaeser (1997) calculate γ across 459 US manufacturing industries and they also found the distribution of the index to be right skewed with a mean value of 0.051 and median value of 0.026.

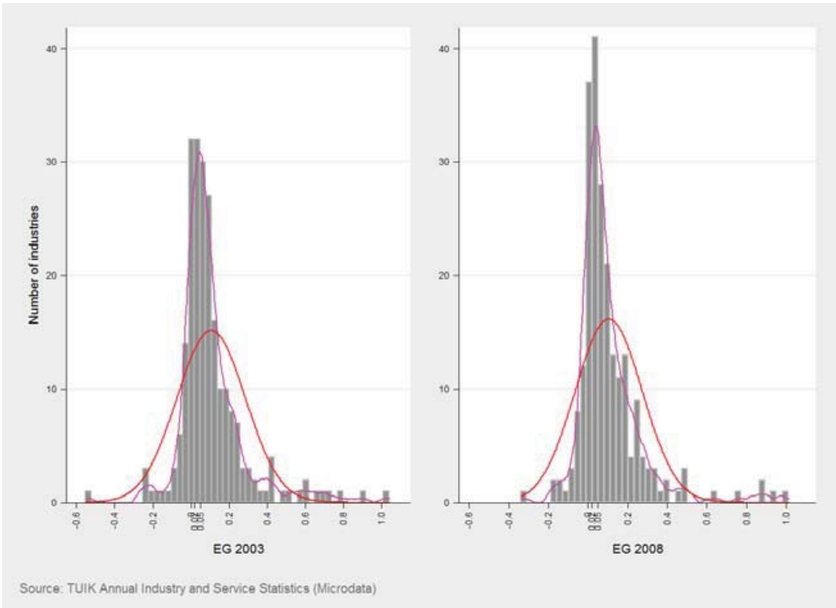


Figure 4.7: **Histogram of γ (4-digit industries)**

Figure 4.8 and 4.9 show the distributions of H, and G across four-digit industries. Geographic concentration is considerably less positively skewed than industrial concentration.

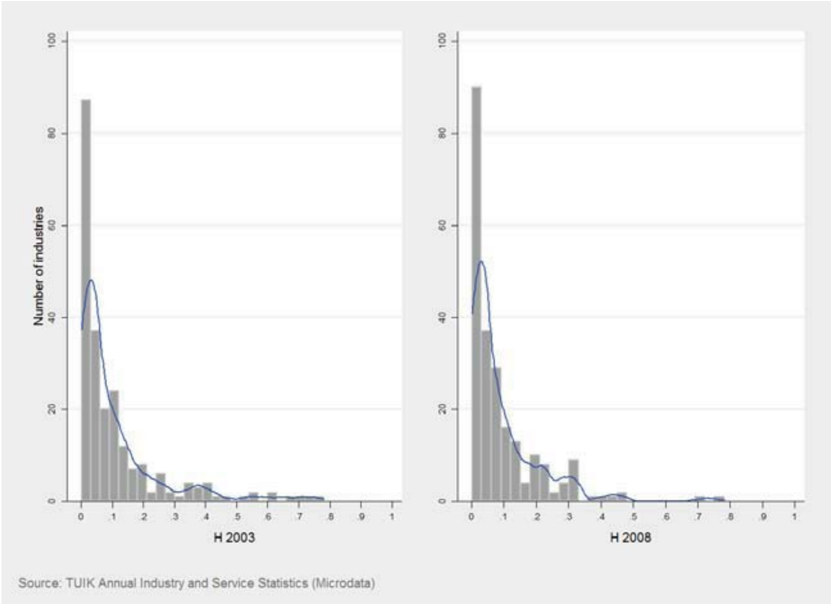


Figure 4.8: **Histogram of H (4-digit industries)**

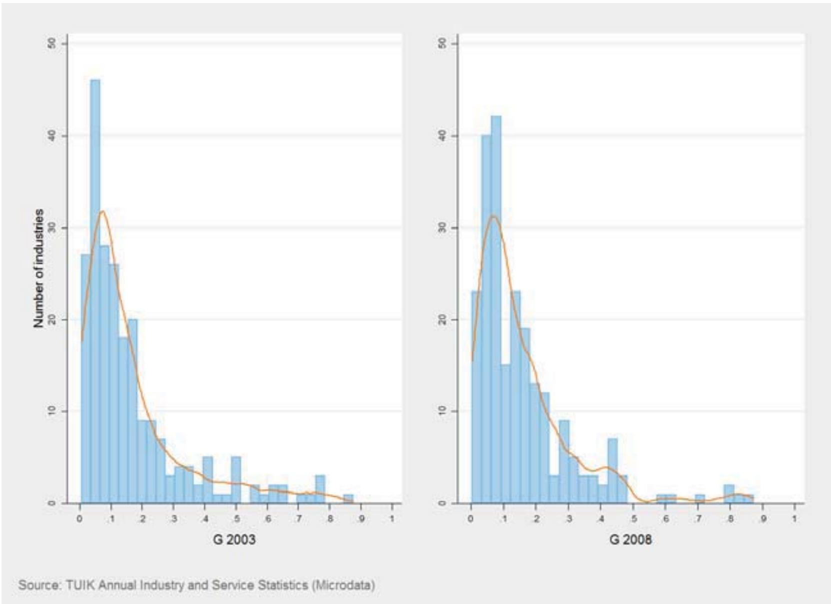


Figure 4.9: **Histogram of G (4-digit industries)**

Tables AI.1 and AI.2 in Appendix present correlations between each measure and the number of firms observed in each industry. The correlation between concentration measures G and γ and between G and H is positive and high in both years. In addition the correlation between Gini and G and between Gini and H is also positive and around 0.60. However, the correlation with the number of firms in the industry is negative in each case in either year. The correlation between the number of firms and γ is low, on the other hand correlations with G and Gini is higher where both indices do not condition on industrial concentration. The negative correlation with the Herfindahl is so expected, as the number of firms increase H decreases. The same findings takes place in Devereux, Griffith, and Simpson (2004) where number of firms are negatively correlated with all other concentration measures and they state that this is an expected result.

Table 4.7 shows the extent of agglomeration by the number and per cent of industries according to the ranges that γ lies in. In 2003 57.9 % of the industries are highly agglomerated displaying a value $\gamma \geqslant 0.05$, 14.5 % are moderately concentrated with $0.02 < \gamma \leqslant 0.05$, and 11.5 % of them have a low degree of concentration, $0 < \gamma \leqslant 0.02$. 16.2 % of the industries take a negative index value which implies plants choosing to locate more diffusely than expected by randomness (Cassey and Smith, 2014). Regarding year 2008, 55.9 % of the industries are highly agglomerated, while 17.5 % of them are moderately concentrated and 8.7 % of them have a low degree of concentration.

Using these definitions, if we were to compare these results with a few developed country examples, in the US 25 % of the industries are classified as highly agglomerated (Ellison and Glaeser, 1997), in France 27 % (Maurel and Sédillot, 1999) and in UK 16 % (Devereux, Griffith, and Simpson, 2004). Regarding the intermediate range, the findings indicate 65 % for the US, 23% for France and 19% for the UK. For the low agglomeration range, US industries display a value of 10 % of, France 50% and the UK 65%. The findings of EG index calculations on Turkish industries show that agglomeration of the 4-digit manufacturing industries is widespread in Turkey compared to developed countries while a small number of industries fall in the category of low concentration.

Table 4.7: **Extent of agglomeration by years**

	# of industries		%	
Range	2003	2008	2003	2008
$\gamma \leqslant 0$	37	41	16.2	17.9
$0 < \gamma \leqslant 0.02$	26	20	11.4	8.7
$0.02 < \gamma \leqslant 0.05$	33	40	14.5	17.5
$0.05 < \gamma \leqslant 0.1$	47	48	20.6	21
$0.1 < \gamma \leqslant 0.2$	45	41	19.7	17.9
$0.2 < \gamma \leqslant 0.3$	18	21	7.9	9.2
$\gamma > 0.3$	22	18	9.7	7.8
Total	228	229	100	100

Table 4.8 summarises the pattern of agglomeration in 2003 by showing the mean of γ calculated at the four-digit industry level by two-digit industries and

the percentage of four-digit industries in each quartile of γ across all four-digit industries. The first quartile comprises the least agglomerated industries while the fourth quartile contains the most agglomerated industries [8].

Recycling (37) and manufacture of coke, refined petroleum products and nuclear fuel (23) industries have two 4-digit sub-industries and manufacture of office machinery and computers (30) and tobacco products (16) industries have only one 4-digit industry. They show the highest level of agglomeration on average where the index ranges in the fourth quartile in each case. These industries share a common point such that they comprise of a small number of firms as confirmed by Table 4.5 with the lowest numbers across two-digit industries. So, high levels of agglomeration may be due to the fact that agglomeration index tends to be upward biased when the number of firms are relatively small in an industry.

Textile (17) and other non-metallic mineral products (26) industries show significant level of agglomeration on average both comprising a large number of sub-industries. 42 % of the 4-digit industries in textiles (17) fall in the fourth quartile which indicates high levels of agglomeration in its 4-digit sub-industries. Regarding other non-metallic mineral products (26) industry, 37 % of sub-industries fall into fourth quartile range. Food products and beverages (15) industry has 30 sub-industries and on average display a value of 0.1277 where 30 % of them lie within the fourth quartile.

To have a closer look, Table 4.9 list the 15 most localized industries in terms of index γ for 2003. The table also shows the number of firms in each industry, the geographic concentration measure G and the industrial concentration measure H.

Regarding specific 4-digit industries the most highly concentrated industry is found to be manufacture of processing of tea and coffee (1586) with an EG index of 1.029 and raw geographic concentration of 0.868. This result is expected as a very high proportion of tea production takes place in Trabzon (TR90) NUTS-2 region including six provinces. Herfindahl index of this industry is 0.013 which implies that industry is quite competitive and the employment is distributed across many plants, so localization may be attributed to raw geographic concentration. This may be broadly ascribed to natural advantages. Recycling of metal waste and scrap (3710) industry is the second most localized industry with a γ value of 0.890. Herfindahl index measuring industry concentration is 0.120 implying that the industry is somewhat concentrated, however raw geographic concentration

[8] The template of the table refers to Table 4 in Devereux, Griffith, and Simpson (2004, p. 543).

Table 4.8: **Agglomeration in 4-digit industries, by 2-digit industry (2003)**

NACE 2-digit	**Mean γ**	**Percent of 4-digit industries in quartile**				**Number of 4-digit industries**
		1st (least)	2nd	3rd	4th (most)	
37 Recycling	0.5599	0	0	0	100	2
23 Coke, refined petroleum & nuclear fuel	0.4726	0	0	0	100	2
30 Office machinery and computers	0.2118	0	0	0	100	1
16 Tobacco products	0.2045	0	0	0	100	1
17 Textiles	0.1715	16	26	16	42	19
26 Other non-metallic mineral products	0.1501	25	13	25	37	24
15 Food products and beverages	0.1277	30	13	27	30	30
24 Chemicals and chemical products	0.1246	22	22	28	28	18
27 Basic metals	0.1197	19	31	25	25	16
18 Wearing apparel; dressing and dyeing of fur	0.1192	17	33	17	33	6
32 Radio, TV & communication equip.	0.1150	33	0	33	33	3
35 Other transport equipment	0.0999	25	13	37	25	8
34 Motor vehicles, trailers and semi-trailers	0.0963	34	0	33	33	3
29 Machinery and equipment n.e.c.	0.0873	23	32	27	18	22
36 Furniture; manufacturing n.e.c.	0.0598	25	33	33	9	12
28 Metal products (exc. machinery & equip.)	0.0586	38	37	12	13	16
19 Leather; man. of luggage, handbags, etc.	0.0577	0	67	33	0	3
21 Pulp, paper and paper products	0.0527	17	50	33	0	6
20 Wood products (exc. furniture)	0.0398	33	0	67	0	6
22 Publish., print.& reprod. of recorded media	0.0389	33	33	9	25	12
25 Rubber and plastic products	0.0245	43	28	29	0	7
31 Electrical machinery and apparatus n.e.c.	0.0199	28	43	29	0	7
33 Medical, precision and optical instruments, etc.	-0.1008	25	50	25	0	4

Quartile boundries are by index γ 1:(-0.5503, 0.0152), 2:(0.0165, 0.0673), 3:(0.0677, 0.1446), 4:(0.1458, 1.0292). ,

is 0.763 which shows a high level of localization in certain regions. We observe that this industry is geographically concentrated in İstanbul (TR10) and İzmir (TR31).

Table 4.9: **15 Most Localized Industries (2003)**

4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
1586 Processing of tea and coffee	0.013	0.868	1.029	215
3710 Recycling of metal waste and scrap	0.120	0.763	0.890	14
2310 Coke oven products	0.360	0.733	0.794	6
2741 Precious metals production	0.736	0.782	0.722	9
1572 Prepared pet foods	0.776	0.786	0.692	2
2666 Other art. of concrete, plaster and cement	0.197	0.614	0.660	16
1717 Prep. and spinn. of other textile fibres	0.566	0.701	0.609	4
2461 Explosives	0.212	0.580	0.603	6
2625 Other ceramic products	0.249	0.568	0.565	8
1724 Silk-type weaving	0.116	0.492	0.528	22
2653 Plaster	0.147	0.483	0.498	20
2233 Reproduction of computer media	0.548	0.625	0.427	3
2951 Machinery for metallurgy	0.306	0.502	0.416	9
2441 Basic pharmaceutical products	0.625	0.659	0.414	10
1753 Non-wovens & related articles (exc. apparel)	0.405	0.548	0.411	4

Except the first ranked industry (1586), a striking point draws the attention in table which is the small number of firms, hence high values of industrial concentration, H. Although EG index accounts for industrial concentration in the course of computing the index, still it may produce biased results in the case of industries with small number of firms. So in order to interpret agglomeration more reliably, Table 4.10 list 20 most agglomerated four-digit industries comprising of at least 20 establishments. First column displays the ranking of the industry in terms of agglomeration within all four-digit industries where the last column shows the number of firms in the related sector. As seen in the last column, the number of firms are quiet high which directs us to more sound results. The first three industries (1586, 1724 and 2653) has also taken place among the 15 most agglomerated industries in Table 4.9 as we also confirm by their ranks. Industries followed by them, again with more than 20 establishments, rank within top 40 as seen in column one and γ ranges within values 0.4 and 0.2, which indicates pretty high level of agglomeration. It is also interesting to note the variation in industrial concentration while all of these industries display high geographic concentration (G). For example building and repair of ships (3511), has high geographic concentration and low industrial concentration, whereas manufacture of knitted and crocheted hosiery (1771) has quite high geographic concentration coupled with high industrial concentration. Textile and related industries (17) have 19 four-digit sub-industries (Table 4.8) of which 6 of them are ranked in the list of 20 most agglomerated industries in Table 4.10. Also three sub-industries of food products and beverages (15) and other non-metallic mineral products (26) take place in this list.

Table 4.10: **20 Most Localized Industries with Num. of firms > 20 (2003)**

Rank	4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
1	1586 Processing of tea and coffee	0.013	0.869	1.029	215
10	1724 Silk-type weaving	0.116	0.492	0.528	22
11	2653 Man. of plaster	0.147	0.483	0.499	20
16	2614 Man. of glass fibres	0.346	0.513	0.399	21
17^{mh}	2960 Man. of weapons and ammunition	0.085	0.365	0.379	163
18	1830 Dressing and dyeing of fur; man. of articles of fur	0.116	0.357	0.347	26
19^{h}	3220 Man. of TV and radio transmitters and apparatus	0.395	0.507	0.341	35
20	2863 Man. of locks and hinges	0.034	0.286	0.315	129
21	3511 Building and repairing of ships	0.014	0.273	0.313	158
23	1725 Other textile weaving	0.023	0.240	0.267	236
27	1751 Man. of carpets and rugs	0.031	0.231	0.250	197
28	1771 Man. of knitted and croch. hosiery	0.108	0.274	0.243	217
29	2211 Publishing of books	0.034	0.223	0.238	68
31	1715 Throwing and preparation of silk	0.022	0.213	0.236	166
32	1512 Production and preserving of poultrymeat	0.049	0.230	0.235	158
34^{h}	2442 Man. of pharmaceutical preparations	0.024	0.208	0.228	190
35	1587 Manufacture of condiments and seasonings	0.055	0.218	0.216	47
36^{h}	3002 Man. of computers & other info. processing equip.	0.103	0.247	0.212	23
37	2640 Man. of bricks, tiles and construction products	0.007	0.183	0.211	382
38	1772 Man. of knitted and croch. pullovers, cardigans, etc.	0.007	0.182	0.209	517

Table 4.11 [9] provide more detailed information on the location characteristics of most agglomerated 20 industries listed in Table 4.10. Columns 2 and 3 display the two regions with the highest proportion of industry employment. In many of the industries first and second regions are contiguous to each other and hence may signal a larger agglomeration. Manufacture of weapons and ammunition (2960) in Kırıkkale and Ankara, building and repairing of ships (3511) in İstanbul and Kocaeli, other textile weaving (1725) in Bursa and İstanbul may be given as examples to agglomerations in neighbouring regions. Fourth and fifth columns show the percentage of industry employment in these two regions. The ratio of industry employment in the top region ranges from 92.9 % in dressing and dying of fur (1830) industry to 22.3 % in manufacture of bricks, tiles and construction products (2640). Column 6 shows the total number of firms in the industry, and columns 7 and 8 show the proportion of firms in the two regions. This ratio ranges from 76.3 % in processing of tea and coffee (1586) to 5.5 % in manufacture of weapons and ammunition (2960) industry. Finally, last two columns display the average firm size in the first and the second region.

The table enables us to gather some basic and significant inferences. Saliently, İstanbul appears to be the most agglomerated region as it takes place in 14 out of 20 cases and in 8 of those it is the first most agglomerated region while in the other 6 cases it is the second one. In four-digit industries where İstanbul is listed as the most agglomerated region, we observe that the second regions in those industries have ratios far below İstanbul, both in terms of employment and firm percentage. It hints about the dominance of İstanbul region in those industries, for instance in industries coded 1830, 1771, 3511.

If we zoom out of İstanbul region and adopt a broader look at the picture, two types of industries arise. First type may be called single agglomeration industries that comprise large proportions of both employment and firms. Examples include processing of tea and coffee (1586) in which Trabzon region involves 88.7 % of employment and 76.3 % of firms, manufacture of locks and hinges (2863) in which İstanbul region involves 87 % of employment and 67.5 % of firms, manufacture of knitted and crocheted hosiery (1771) in which again İstanbul comprises 85.4 % of employment and 68.2 % of firms. These industries have large number of firms and average firm size of these industries is around 65.

[9] The table is constructed with reference to Table 6 in Devereux, Griffith, and Simpson (2004, p. 546).

Table 4.11: **Most agglomerated regions, 2003**

NACE 4-digit	**1st region**	**2nd region**	**Percent of emp. in region**		**Number of firms**	**Percent of firms in region**		**Av. firm size (employment)**	
			1st	2nd		1st	2nd	1st	2nd
1586 Processing of tea and coffee	TR90 - Trabzon	TR10 - İstanbul	88.7	6.3	215	76.3	13	63	26
1724 Silk-type weaving	TR41 - Bursa	TR10 - İstanbul	77.1	20.9	22	68.2	22.8	25	20
2653 Man. of plaster	TR51 - Ankara	TR41 - Bursa	62.8	12.4	20	40	5	26	41
2614 Man. of glass fibres	TR42 - Kocaeli	TR10 - İstanbul	70.5	9.9	21	19	14.3	223	42
2960 Man. of weapons and ammunition	TR71 - Kırıkkale	TR51 - Ankara	41.9	34.8	163	5.5	10.5	355	156
1830 Dressing and dyeing of fur; man. of articles of fur	TR10 - İstanbul	TR21 - Tekirdağ	92.9	3	26	65.4	3.8	33	18
3220 Man. of TV and radio transmitters and apparatus	TR51 - Ankara	TR10 - İstanbul	73.2	19.9	35	28.6	28.6	348	95
2863 Man. of locks and hinges	TR10 - İstanbul	TR41 - Bursa	87	4.7	129	67.5	3.1	67	78
3511 Building and repairing of ships	TR10 - İstanbul	TR42 - Kocaeli	84.9	7.5	158	78.5	7	73	73
1725 Other textile weaving	TR41 - Bursa	TR10 - İstanbul	54.6	21.6	236	68.6	14.5	40	75
1751 Man. of carpets and rugs	TRC1 - Gaziantep	TR72 - Kayseri	38.8	20.1	197	32	9.1	65	117
1771 Man. of knitted and croch. hosiery	TR10 - İstanbul	TR21 - Tekirdağ	85.4	4.7	217	68.2	2.8	127	171
2211 Publishing of books	TR10 - İstanbul	TR51 - Ankara	78.6	15.3	68	58.8	20.6	42	23
1715 Throwing and preparation of silk	TRC1 - Gaziantep	TR41 - Bursa	42.8	19.7	166	38	10.2	157	267
1512 Production and preserving of poultrymeat	TR42 - Kocaeli	TR22 - Balıkesir	31.7	21	158	19.6	7.6	124	211
2442 Man. of pharmaceutical preparations	TR10 - İstanbul	TR42 - Kocaeli	78.7	6	190	52.6	7.9	196	100
1587 Manufacture of condiments and seasonings	TR31 - İzmir	TR10 - İstanbul	49.4	16.3	47	29.8	12.8	44	34
3002 Man. of computers & other info. processing equip.	TR10 - İstanbul	TR31 - İzmir	83.2	5.9	23	43.5	13	44	10
2640 Man. of bricks, tiles and construction products	TR33 - Manisa	TR83 - Samsun	22.3	18.5	382	17.3	16	49	45
1772 Man. of knitted and croch. pullovers, cardigans, etc.	TR10 - İstanbul	TRC1 - Gaziantep	75.5	6.5	517	68.6	7.5	60	48

However, there are a few industries in the single agglomeration type which are characterized by high percentage of employment with low percentage of firms. For instance manufacture of glass fibres (2614) industry has 70.5 % of its employment in Kocaeli where only 19 % of the related industry firms are located. However average employment size in this region is quite high, namely 223, which signals the existence of a few large firms in the region. Number of firms in this industry is also pretty low which amounts to 21. Another similar example is manufacture of TV and radio transmitters and apparatus (3220) industry. This industry also has small number of firms, and 73.2 % of its employment and 28.6 % of firms are located in Ankara. Average firm size in this industry is pretty high at a level of 348 which shows the dominance of only a few large firms in the industry.

It is also worth noting that even this first type is called single agglomeration industries, we observe that when the share of employment in the first and second region are summed up, total share of employment in these regions reaches to ratios between 80 to 95 %. That much high shares of employment only in two regions actually mean that production is concentrated in these two regions where the first region far above the second one.

On the other hand, second type of industries comprise the ones with two agglomerations where relatively similar sized regions in terms of employment take place. Examples to this type include production and preserving of poultrymeat (1512) and manufacture of bricks, tiles and construction products (2640) industries which of both have very close shares of employment. Regarding industry 1512, 31.7 % of employment is concentrated in Kocaeli and 21 % is concentrated in Balıkesir region. As for the industry 2640, 22.3 % of employment is concentrated in Manisa and 18.5 % is concentrated in Samsun region.

Another example to second type industries is manufacture of weapons and ammunition (2960) industry with employment shares 41.9 % and 34.8 % concentrated in Kırıkkale and Ankara, respectively. However a striking point in this industry is that only a small portion of firms takes place in the first region, to name it 5.5 % and average firm size in the first region is 355, which is quite high. This is the highest average firm size in the list of most agglomerated regions. It also shows that there are small number of pretty large firms in the first region.

Table 4.12: **15 Least Localized Industries-2003**

4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
3350 Watches and clocks	0.680	0.425	-0.550	4
2214 Publishing of sound recordings	0.625	0.449	-0.247	3
2052 Articles of cork, straw and plaiting materials	0.459	0.276	-0.243	6
2232 Reproduction of video recording	0.531	0.354	-0.239	2
2664 Manufacture of mortars	0.243	0.067	-0.217	9
2215 Other publishing	0.208	0.053	-0.182	12
1595 Other non-distilled fermented beverages	0.280	0.140	-0.159	5
2931 Agricultural tractors	0.140	0.037	-0.111	95
3230 TV and radio receivers, sound/video recording/...	0.203	0.113	-0.086	54
2955 Machinery for paper and paper-board production	0.135	0.053	-0.084	13
2511 Rubber tyres and tubes	0.239	0.151	-0.080	14
1552 Ice cream	0.112	0.044	-0.068	74
2465 Prepared unrecorded media	0.257	0.180	-0.059	8
2626 Refractory ceramic products	0.115	0.056	-0.055	26
3161 Electrical equipment for engines and vehicles n.e.c	0.142	0.082	-0.053	40

Table 4.12 lists 15 least agglomerated industries in 2003. As mentioned before the index of least agglomerated, along with the rest of them with negative values of γ, are found to be statistically insignificant. This has been also the case for least agglomerated industries in US (Ellison and Glaeser, 1997) and Germany (Alecke, Alsleben, Scharr, and Untiedt, 2006). These industries are characterized by high levels of industrial concentration represented by high Herfindahl indices and generally high levels of geographic concentration represented by G. This means that despite the unequal distribution of production activity in the related sector, compared to the distribution of total production employment, the geographic dispersion of employment is largely explained by industrial concentration. In other words employment ends up being concentrated in a few very large and randomly scattered plants. However, manufacture of mortars (2664), agricultural tractors (2931), other publishing (2215) machinery for paper and paper-board production (2955) and refractory ceramic products (2626) industries are characterized by low levels of geographic concentration. As Krugman (1999) states, centrifugal forces may be at work for the dispersion of economic activity due to the factors both from supply side (firms may choose production location according to the place of workers) and the demand size (some firms may have a motivation for locating close to consumers). However, great majority of the industries do not arise as being those in which scattering in order to be close to final consumers is important. Dispersion with the incentive of proximity to final consumers is expected to arise in industries producing perishable goods, especially in the sub-industries of food products and beverages (15) industry, especially in three-digit industries such as; production, processing and preserving of meat and meat products (15.1), processing and preserving of fish and fish products (15.2) and manufacture of dairy products (15.3). Only ice cream (1552) industry arises as a perishable good in-

dustry among the least agglomerated ones. As Ellison and Glaeser (1997) name, least agglomerated industries are "something of a mixed bag" and not so easy to explain forces beyond dispersion.

Now, if we were to consider extent of agglomeration in 2008, Table 4.13 summarises the pattern of agglomeration in 2008 in the same vein as Table 4.8. Recycling (37) and manufacture of coke and refined petroleum products and nuclear fuel (23) industries have only two 4-digit sub-industries and again rank in the top five as in 2003 , but compared to 2003 the mean value of γ for theses industries declined definitely which is also observed by the percentage changes of the index in the quartiles. In 2003 both sub-industries for these sectors only have taken place in the forth quartile while in 2008 one of them switched to the third quartile apparently. Two industries ranking in top five in 2003, namely tobacco products (16) and manufacture of office machinery and computers (30), are not among the top ranked industries in terms of average γ's in 2008. Especially tobacco products (16) industry is ranked as the last in 2008 contrary to the case in 2003 where it was the fourth. This may be due to the fact that the number of establishments has decreased and firms changed structure in terms of size as confirmed by Table 4.6, thus these two facts together may have had an influence of the de-agglomeration of the industry.

Table 4.13: **Agglomeration in 4-digit industries, by 2-digit industry (2008)**

NACE 2-digit	**Mean γ**	**Percent of 4-digit industries in quartile**				**Number of 4-digit industries**
		1st (least)	2nd	3rd	4th (most)	
17 Textiles	0.2583	10	16	21	53	19
37 Recycling	0.1723	0	0	50	50	2
23 Coke, refined petroleum products & nuclear fuel	0.1554	0	0	50	50	2
32 Radio, TV & communication equipment and apparatus	0.1476	0	0	33	67	3
27 Basic metals	0.1431	38	18	19	25	16
15 Food products and beverages	0.1326	13	20	33	33	30
18 Wearing apparel; dressing and dyeing of fur	0.1277	0	17	50	33	6
36 Furniture; manufacturing n.e.c.	0.1242	17	25	25	33	12
34 Motor vehicles, trailers and semi-trailers	0.0978	0	0	100	0	3
29 Machinery and equipment n.e.c.	0.0962	36	27	18	18	22
24 Chemicals and chemical products	0.0943	22	33	22	22	18
26 Other non-metallic mineral products	0.0908	38	13	29	21	24
30 Office machinery and computers	0.0908	50	0	0	50	2
19 Leather; man. of luggage, handbags, etc.	0.0856	0	67	0	33	3
20 Wood products (exc. furniture)	0.0742	17	50	17	17	6
25 Rubber and plastic products	0.068	29	43	0	29	7
21 Pulp, paper and paper products	0.0475	17	67	0	17	6
28 Metal products (exc. machinery & equip.)	0.0439	44	31	19	6	16
33 Medical, precision & optical instruments, etc.	0.0423	50	0	50	0	4
35 Other transport equipment	0.0412	38	13	38	13	8
31 Electrical machinery and apparatus n.e.c.	0.0269	14	57	29	0	7
22 Publish., print. & reprod. of recorded media	0.0219	33	33	17	17	12
16 Tobacco products	-0.1505	100	0	0	0	1

Quartile boundaries are by index γ , 1:(-0.3364, 0.0139), 2:(0.0155, 0.0602), 3:(0.0631, 0.1457), 4:(0.1459, 1.0013).

On the other hand, textile (17) industry which was ranked fifth in 2003, is ranked as the top in terms of average values of 4-digit γ's in 2008. It had an average value of 0.1715 in 2003 and increased to 0.2583 in 2008. Comparing the values in Tables 4.8 and 4.13 in the quartile columns for this industry also shows that the percentage of four-digit industries in the first two quartiles has decreased and accordingly increased in the third and fourth quartiles during the period (53 % of the 4-digit industries in textiles (17) fall in the fourth quartile). It demonstrates that many of the 4-digit sub-industries of textiles (17) industry have absolutely experienced increases in their index values.

Radio, television & communication equipment (32) industry also displays a significant change with respect to two years. The average γ value of the industry has been observed as 0.1150 and the percentages of the γ's were evenly distributed among first, third and fourth quartiles in 2003. However, in 2008 average value increased to 0.1431 and the distribution of the index values in terms of quartiles concentrated in the third and fourth quartile. Also note that, there are only three sub-industries in this sector, and one of these has switched from the first quartile in 2003 to fourth quartile in 2008, hence pulling the average value up. Basic metals (27) industry is ranked fifth among the two-digit industries in 2008. It's average γ increased from 0.1197 to 0.1431 even though the share of the first quartile increased (38 %) to the detriment of the second and third quartile and the share of the fourth quartile remained same. So, despite many industries lying in the first quartile, the values observed in the fourth quartile in 2008 should be higher than the ones in 2003 in order to contribute as an increase to the average value.

Table 4.14 list the 15 most localized 4-digit industries in terms of index γ for 2008. Five out of 15 industries in this year has also been ranked within the top 15 list in 2003 (1717, 1586, 2741, 1753 and 1724). Most localized industry is observed as preparation and spinning of other textile fibres (1717) with a γ of 1.001. Herfindahl measure and the number of firms in this industry demonstrate that even though the industry is geographically concentrated, this is largely due to the industrial concentration. Processing of tea and coffee (1586) industry is again among the most localized industries as has been in year 2003. Number of firms in this industry declined a little but the pattern of agglomeration did not change, it is still highly concentrated with an index of 0.944. The low measure of H attained by the industry also shows that industry is quite competitive and agglomeration is mainly driven by raw geographic concentration.

A similar point as observed in top listed industries in 2003 can be remarked also for the case in 2008. Fewness of firms as a numerical amount and highness in the measure of industrial concentration γ arise as a significant feature of the table, though to a lesser extent compared to 2003. Only five industries, which are 1586, 1751, 1725, 1715 and 3615, have numerous firms. Thus, in order to

Table 4.14: **15 Most Localized Industries (2008)**

4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
1717 Preparation and spinning of other textile fibres	0.109	0.860	1.001	14
1586 Processing of tea and coffee	0.013	0.812	0.944	184
2741 Precious metals production	0.752	0.830	0.864	4
2941 Portable hand held power tools	0.440	0.794	0.864	3
1753 Non-wovens and related articles, except apparel	0.176	0.699	0.774	10
2754 Casting of other non-ferrous metals	0.142	0.589	0.633	18
1751 Carpets and rugs	0.019	0.433	0.495	261
1725 Other textile weaving	0.012	0.429	0.493	259
1724 Silk-type weaving	0.088	0.454	0.483	26
2621 Ceramic household and ornamental articles	0.115	0.440	0.449	73
1715 Throwing and preparation of silk, ...	0.019	0.353	0.399	209
3615 Manufacture of mattresses	0.049	0.364	0.394	114
1595 Other non-distilled fermented beverages	0.209	0.429	0.367	12
2465 Prepared unrecorded media	0.309	0.465	0.336	13
2611 Flat glass	0.247	0.426	0.331	5

have a more reasonable analysis of geographic concentration purged from industrial concentration, Table 4.15 list top 20 four-digit industries with at least 20 establishments. Additionally, the first column shows the ranking of the industry in terms of γ within all industries. As last column shows, the number of firms are quite high which enables us to make more coherent comments about agglomeration. At first glance at Table 4.15 we observe that first 7 industries have also been listed among the top 15 in 4.14 as also confirmed by their rankings. Industries followed by them take place within the top thirty two industries and range between 0.24 and 0.32, which indicates agglomeration at pretty high levels. Another point to be mentioned is that textile (17) sub-industries constitute 6 out of 20 industries which is an expected outcome as Table 4.13 displays this two-digit sector in the top of the list with the highest percentage in the fourth quartile.

It is worth noting that 11 industries that are marked with an asterisk in Table 4.15 have also been listed within the top 20 regarding year 2003. In other words, we may say that 55 % of the industries remained in the top list during the period. Moreover, five textile (17) related industries in this group have all enhanced their rankings relative to 2003. Again this is not a surprise as Table 4.13 signalled that on average textile (17) industries have increased their index values. Manufacture of weapons and ammunition (2960) was ranked the same in both years, however the rest of the industries are ranked lower relative to their 2003 values.

A striking feature of the most localized industries is that it largely encompasses low-tech industries. Within this technology group textile and traditional industries are observed to be dominant. In 2003, one medium tech and three high-tech industries appear in the list of 20 most localized industries as shown as a superscript in the first column of Table 4.10. Regarding 2008, there are

Table 4.15: **20 Most Localized Industries with Num. of firms > 20 (2008)**

Rank	4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
2*	1586 Processing of tea and coffee	0.013	0.812	0.944	184
7*	1751 Manufacture of carpets and rugs	0.019	0.433	0.495	261
8*	1725 Other textile weaving	0.012	0.429	0.494	259
9*	1724 Silk-type weaving	0.088	0.454	0.483	26
10	2621 Manufacture of ceramic household	0.115	0.440	0.449	73
11*	1715 Throwing and preparation of silk	0.019	0.353	0.399	209
12	3615 Manufacture of mattresses	0.049	0.364	0.394	114
17^{*mh}	2960 Manufacture of weapons and ammunition	0.038	0.298	0.321	209
18*	2653 Manufacture of plaster	0.076	0.317	0.317	31
19^{mh}	2461 Manufacture of explosives	0.189	0.371	0.299	22
20*	1830 Dressing and dyeing of fur; man. of articles of fur	0.019	0.263	0.293	113
22	3663 Other manufacturing n.e.c.	0.051	0.267	0.274	218
23*	1772 Man. of knitted and croch. pullovers, cardigans, etc.	0.011	0.240	0.272	540
24	1712 Preparation and spinning of woollen-type fibres	0.118	0.297	0.259	69
26^{mh}	2463 Manufacture of essential oils	0.143	0.305	0.248	28
27	3710 Recycling of metal waste and scrap	0.309	0.411	0.246	23
28*	2863 Manufacture of locks and hinges	0.041	0.238	0.246	162
29*	3511 Building and repairing of ships	0.005	0.212	0.243	715
30	3622 Manufacture of jewellery and related articles n.e.c.	0.013	0.216	0.242	829
32	2123 Manufacture of paper stationery	0.074	0.255	0.240	50

three medium-high tech industries within the list of most localized industries, namely manufacture of weapons and ammunition (2960), manufacture of explosives (2461) and manufacture of essential oils (2463). The agglomeration in these sectors are presumably driven by spillovers rather than natural advantages. However, it should be noted that EG index solely is not capable of revealing the sources or determinants of agglomeration as it does not make differentiation between the natural advantage or spillovers. A thorough analysis is essential to detect the drivers of agglomeration by also considering agglomeration theories. Hence it's better taking the information revealed by EG index as a useful tool for detecting the extent of agglomeration.

Table 4.16 give detailed information on the location characteristics of most agglomerated 20 industries in 2008 in the same way as in Table 4.11. As the case has been in 2003, İstanbul appears to be the most agglomerated region as it is shown in 12 out of 20 cases. In the cases where İstanbul is listed as the first region, we see that the shares of second regions are far below İstanbul in terms of employment and firm percentage, e.g. 3622, 2863, 3663 etc.

Also in 2008 two types of industries arise, as has been called before, single agglomeration type and two agglomeration type industries. The first type encompasses industries with large percentage shares of both employment and firm number. Examples include other manufacturing (3663) industry in which İstanbul region involves 82.2 % of employment and 77.1 % of firms, manufacture of paper stationery (2123) industry again İstanbul involves 80.5 % of employment and 84 % of firms, other textile weaving (1725) industry in which Bursa region involves 70.2 % of employment and 65.6 % of firms, processing of tea and coffee (1586) industry in which Trabzon region involves 86.3 % of employment and 59.2 % of firms, etc. What is most striking in this type is that industries taking place therein are extremely concentrated in İstanbul region and roughly constitute more than 75 % of employment and at least 70 % of firms.

On the other hand, second type of industries include industries with two agglomeration where relatively similar sized regions in terms of employment take place. For instance, preparation and spinning of woollen-type fibres (1712) industry and manufacture of essential oils (2463) industry where in the former one employment is concentrated in Tekirdağ (%39.9) and Manisa (%31.7) and in the latter one in Antalya (%53.3) and İstanbul (%42.6). However a striking point is observed regarding industry 1712. Although almost 40 % of the employment is concentrated in Tekirdağ, the region holds only 5.8 % of the firms in the industry and average firm size in this region is 295 which is the highest value in the list of most agglomerated regions. These together indicate that there are small number of quite large firms in the region. Another example to this type may be given as industry 2960 in which Kırıkkale is the second most agglomerated region with an employment share of 30.5 % and holds only 5.7 % of firms with an average firm size of 284.

Comparing location characteristics of industries common in both years also reveals important information. Processing of tea and coffee (1586) industry has maintained the shares of employment in two regions in 2008, however the share of firms in the first region has decreased from 76.3 % to 59.2 % where average firms size increased from 63 to 113. This shows that firms in Trabzon has become fewer but larger throughout the period. Manufacture of carpets and rugs (1751) industry is again mostly agglomerated in Gaziantep and Kayseri. However the share of Gaziantep has increased significantly both in terms of employment and number of firms throughout the period while Kayseri experienced a notable decline in the respective shares reflecting the dominance of Gaziantep in the industry over the years. Other textile weaving (1725) industry has experienced increase in the employment share of Bursa as the first region while İstanbul has lost more than half of its share in terms of employment. Contrarily Bursa had a slight decline in regional firm share while İstanbul experienced 5 percentage points increase where average firm size has increased in the former and decreased in the latter. Thus we may infer that Bursa has had fewer large firms and İstanbul had more small firms according to 2003.

Both regions in silk-type weaving (1724) industry has experienced slight declines in both employment and firm shares, which indicates that Bursa and İstanbul are not as dominant as has been in 2003 in the sector but still most agglomerated regions. Manufacture of plaster (2653) industry has undergone a different process than other industries. The first most agglomerated region in both years has been observed as Ankara, however in 2008 the second most agglomerated region switched from Bursa to Mardin. Also the industry has small number of firms (31) and the index tends to be biased in such cases as mentioned before. A similar case happened in manufacture of weapons and ammunition (2960) industry where first two regions switched, each one took the other's place in 2008. Kırıkkale and Ankara have been the first and second agglomerated regions in 2003 and the case has been vice versa in 2008. Gaziantep has experienced a notable increase, 10 percentage points, in its employment share in throwing and preparation of silk (1715) industry where Bursa had a slight increase in this share. Nonetheless, firm share of Gaziantep declined almost by 8 % while in Bursa increased by the same amount. Also having the information on average firm sizes of the regions we may conclude that Gaziantep has held fewer large firms and Bursa held more small firms according to 2003.

Employment share of İstanbul in dressing and dyeing of fur (1830) industry has declined while its firm share declined in 2008. Also the number of firms has significantly increased in this industry. Manufacture of knitted and crochet pullovers, cardigans, etc. (1772) industry has experienced an increase in employment share in the second region Gaziantep while firm share declined. Also the significant increase in the average firm size in this region demonstrates that large firms has been dominant when it comes to year 2008.

Manufacture of locks and hinges (2863) industry was mostly agglomerated in İstanbul and Bursa in 2003 and in 2008 the second most agglomerated region came out to be İzmir.

In building and repairing of ships (3511) industry, İstanbul still being the most agglomerated regin in 2008 has experienced declines in both employment and firm shares. On the other hand, Kocaeli has enhanced both its employment and firm share within the industry.

Table 4.16: **Most agglomerated regions, 2008**

NACE 4-digit	1st region	2nd region	Percent of emp. in region		Number of firms	Percent of firms in region		Av. firm size (employment)	
			1st	2nd		1st	2nd	1st	2nd
1586* Processing of tea and coffee	TR90 - Trabzon	TR10 -İstanbul	86.3	6.6	184	59.2	16.3	113	32
1751* Manufacture of carpets and rugs	TRC1 - Gaziantep	TR72 - Kayseri	61.4	8.8	261	34.5	4.2	111	130
1725* Other textile weaving	TR41 - Bursa	TR10 - İstanbul	70.2	9.7	259	65.6	19.7	61	28
1724* Silk-type weaving	TR41 - Bursa	TR10 - İstanbul	72.4	11.4	26	50	15.4	26	13
2621 Manufacture of ceramic household & ornm. artic.	TR33 - Manisa	TR41 - Bursa	64.3	16.5	73	32.9	9.6	160	140
1715* Throwing and preparation of silk	TRC1 - Gaziantep	TR41 - Bursa	52.2	20.5	209	29.2	18.2	194	122
3615 Manufacture of mattresses	TR72 - Kayseri	TR10 - İstanbul	57.9	10.9	114	27.2	16.7	115	35
2960* Manufacture of weapons and ammunition	TR51 - Ankara	TR71 - Kırıkkale	41.3	30.5	209	15.3	5.7	144	284
2653* Manufacture of plaster	TR51 - Ankara	TRC3 - Mardin	48.3	12.4	31	32.3	6.5	39	50
2461 Manufacture of explosives	TR42 - Kocaeli	TR51 - Ankara	48.9	31.3	22	13.6	36.4	129	31
1830* Dressing and dyeing of fur; man.of articles of fur	TR10 - İstanbul	TR21 - Tekirdağ	79	17.7	113	85	8	17	40
3663 Other manufacturing n.e.c.	TR10 - İstanbul	TR21 - Tekirdağ	82.2	5	218	77.1	2.8	53	90
1772* Man. of knitted and croch. pullovers, cardigans, etc.	TR10 - İstanbul	TRC1 - Gaziantep	78.6	11.2	540	78.3	3	44	167
1712 Preparation and spinning of woollen-type fibres	TR21 - Tekirdağ	TR33 - Manisa	39.9	31.7	69	5.8	29	295	47
2463 Manufacture of essential oils	TR61 - Antalya	TR10 - İstanbul	53.3	42.6	28	71.4	17.9	9	30
3710 Recycling of metal waste and scrap	TR31 - İzmir	TR42 - Kocaeli	60	13.4	23	21.8	17.4	82	23
2863* Manufacture of locks and hinges	TR10 - İstanbul	TR31 - İzmir	79.8	3.6	162	58	5.5	81	38
3511* Building and repairing of ships	TR10 - İstanbul	TR42 - Kocaeli	75	10.7	715	71.3	14	62	45
3622 Manufacture of jewellery and related articles n.e.c.	TR10 - İstanbul	TR31 - İzmir	77	5.5	829	74.5	4.3	27	34
2123 Manufacture of paper stationery	TR10 - İstanbul	TR21 - Tekirdağ	80.5	8.6	50	84	2	32	143

Notes: An asterisk (*) indicates that the industry is also listed within the top industries in 2003

Table 4.17 lists least agglomerated industries in 2008. Industries with an asterisk (*) indicates that the industry is also listed within the least agglomerated industries in 2003. Again as has been in 2003, all of the indices with negative values are statistically insignificant. These industries are characterized by high Herfindahl indices and geographic concentration indices and geographic dispersion. However, it is not straightforward to provide a sound explanation such as centrifugal forces defined by Krugman (1999) as a source for dispersion of industries.

Table 4.17: **15 Least Localized Industries (2008)**

4-digit NACE Rev. 1.1 code	H	G	γ	Num. of firms
2233 Reproduction of computer media	0.463	0.243	-0.336	3
2214* Publishing of sound recordings	0.291	0.147	-0.169	7
2664* Manufacture of mortars	0.207	0.067	-0.163	12
3530 Aircraft and spacecraft	0.278	0.143	-0.154	44
1600 Tobacco products	0.309	0.176	-0.150	46
3140 Accumulators, primary cells & primary batteries	0.209	0.094	-0.125	34
3543 Invalid carriages	0.307	0.211	-0.088	6
3630 Musical instruments	0.218	0.129	-0.086	33
2124 Manufacture of wallpaper	0.235	0.148	-0.081	8
2830 Steam generators, except central heating...	0.152	0.086	-0.061	43
3350* Watches and clocks	0.298	0.221	-0.059	11
2623 Ceramic insulators & insulating fittings	0.412	0.328	-0.052	7
3001 Office machinery	0.193	0.132	-0.049	8
2441 Basic pharmaceutical products	0.388	0.309	-0.045	3
2955 Machinery for paper & paperboard prod.	0.077	0.034	-0.040	25

4.3.1.1 Technology-wise agglomeration

Table 4.18 presents the extent of agglomeration by technology groups. Each row shows the percentage of the technology-wise agglomeration within the agglomeration range defined in the first column and the last column shows the number of 4-digit industries in that range.

Low and medium-low technology sectors show a higher degree of agglomeration. Specifically, above the 0.05 threshold, the share of low-technology sectors are much higher than other technology levels for both years. We also observe that from 2003 to 2008, the share of low-technology group at medium and high agglomeration levels increased.

Regarding medium-level technology group, the picture is somewhat different. The share of this group at no agglomeration and low agglomeration levels have increased while for higher levels decreased over the period. It shows that agglomeration in this group has undergone a moderation process. On the contrary, throughout the period the share of medium-high sectors at each level of agglomeration has increased while a significant decline has been observed at no concentration level. This indicates an enhancement in agglomeration levels in the medium-high technology group.

The share of high technology group in each range is substantially low compared to others. The decline in agglomeration in this group is pretty obvious as its share in no agglomeration level has increased, in medium agglomeration level has dropped to zero and in high agglomeration level has been stable around 5.4 %. Consistent with the existing literature on agglomeration, also in Turkey high-technology sectors display agglomeration to a very limited degree.

Table 4.18: **Extent of agglomeration by technology groups (%)**

	Low	**Med-low**	**Med-high**	**High**	**Total**	**N**
			2003			
$\gamma \leqslant 0$	37.9	24.3	32.4	5.4	100	37
$0 < \gamma \leqslant 0.02$	38.5	38.5	23.0	0	100	26
$0.02 < \gamma \leqslant 0.05$	33.3	33.3	27.3	6.1	100	33
$\gamma > 0.05$	47.0	28.0	19.7	5.3	100	132
			2008			
$\gamma \leqslant 0$	36.6	34.1	17.1	12.2	100	41
$0 < \gamma \leqslant 0.02$	14.3	47.6	38.1	0	100	21
$0.02 < \gamma \leqslant 0.05$	42.0	29.0	29.0	0	100	38
$\gamma > 0.05$	48.8	24.8	21.0	5.4	100	129

In fact, inferences we obtained regarding technology-wise agglomeration above are supported by summary statistics shown in Table 4.19. As mentioned above,on average agglomeration has increased in two technology levels, namely low and medium-high group. Mean of agglomeration has increased from 0.114 to 0.130 in the former group and from 0.081 to 0.084 in the latter one. On the other hand, in the med-low and high technology sectors agglomeration on average has declined. To be precise, average agglomeration index in aforesaid sectors has decreased from 0.118 to 0.094 and 0.080 to 0.055, respectively. Taking a glance at the maximum values displayed in each technology level reveals that the highest agglomeration is observed in the low technology level and declines as we move towards high technology. There are two extreme cases in this table. First one is the maximum value displayed in the medium-low group for γ_{2007} which is 1.758 and the other is the maximum values displayed in the high technology group for γ_{2004}, that is 1.161. Industry in the medium-low technology group is precious metals production (2741) and the other in the high technology group is manufacture of industrial process control equipment (3330).These industries are characterised by very few firms, namely 5 and 4 firms respectively, and one of

Table 4.19: **Summary statistics by technology level**

	γ_{2003}	γ_{2004}	γ_{2005}	γ_{2006}	γ_{2007}	γ_{2008}
			Low			
mean	0.114	0.141	0.110	0.148	0.157	0.130
sd	0.194	0.215	0.331	0.197	0.211	0.193
min	-0.247	-0.146	-1.601	-0.048	-0.115	-0.336
max	1.029	1.140	1.442	1.32	1.326	1.001
N	97	97	97	96	98	97
			Med-low			
mean	0.118	0.086	0.085	0.075	0.113	0.094
sd	0.183	0.117	0.149	0.106	0.247	0.158
min	-0.217	-0.174	-0.211	-0.140	-0.107	-0.163
max	0.795	0.529	0.768	0.559	1.758	0.864
N	67	66	66	66	66	67
			Med-high			
mean	0.081	0.080	0.081	0.059	0.067	0.084
sd	0.127	0.120	0.194	0.098	0.087	0.144
min	-0.111	-0.140	-0.819	-0.104	-0.078	-0.125
max	0.603	0.594	0.881	0.441	0.404	0.864
N	53	53	53	52	52	53
			High			
mean	0.080	0.148	0.051	0.077	0.070	0.055
sd	0.255	0.396	0.143	0.175	0.159	0.119
min	-0.550	-0.567	-0.318	-0.171	-0.180	-0.154
max	0.414	1.161	0.284	0.444	0.443	0.230
N	11	12	13	13	12	12

these firms provide a significant share of employment within industry and located in one region. Hence EG index favours large firms, an entry of a large firm into the industry makes one to obtain high index values.

4.3.1.2 Agglomeration over time

The evolution of agglomeration throughout the period is another important issue that should be addressed. In section 4.3.1 we investigated whether the most concentrated industries in 2008 were among the most concentrated ones in also 2003. One may be interested in analysing whether there exists a tendency to a lower or higher agglomeration.

Descriptive statistics of the γ index of 4-digit industries by years are shown in Table 4.20. The degree of overall agglomeration is observed to be stable as both the mean and median display quite similar values throughout the period. Nonetheless, a value of γ around 0.1 still represents a high level of localization on average. The stability in agglomeration levels in most of the Turkish manufacturing industries is a common pattern observed in other countries, for instance Dumais, Ellison, and Glaeser (2002) for US (1972-92) and Alonso-Villar, Chamorro-Rivas, and González-Cerdeira (2004) for Spain (1993-99) have findings on the stability of agglomeration. However some industries experienced remarkable changes in their agglomeration levels.

Table 4.20: **Summary statistics of γ by years**

γ	μ	median	σ	min.	max.	N
2003	0.106	0.067	0.180	-0.55	1.029	228
2004	0.112	0.062	0.188	-0.567	1.161	228
2005	0.093	0.058	0.250	-1.601	1.442	229
2006	0.102	0.064	0.158	-0.171	1.32	227
2007	0.119	0.065	0.202	-0.18	1.758	228
2008	0.105	0.060	0.170	-0.336	1.001	229

Figure 4.10 shows the evolution of EG index γ between 2003 and 2008 for each manufacturing industry. The x-axis represents the index value as a deviation from the median and y-axis the average rate of change in the index over the 5 years. It is straightforward to interpret the figure as follows: industries above the horizontal red line have experienced increase in their geographic concentration levels over the period while the ones below experienced a decline. On the other hand, vertical red line stands for a threshold for industries that are below and above the median value in 2003 which is 0.067. Accordingly, industries to the right of the line depict high levels of agglomeration in 2003 and low levels to the left. However, it should be kept in mind that 0.067 is already a high value indicating a high level of geographic concentration. So, some of the industries to the left of red vertical line still experience high level of geographic concentration. In order to distinguish those industries and to make more sound inferences, we place three more reference lines on the x-axis at values -0.017, -0.047 and -0.067 represented by the blue, green and purple lines, respectively. These values correspond to γ values of 0.05, 0.02 and 0 respectively, which are the lower thresholds for high, medium and low level of agglomeration defined byEllison and Glaeser (1997).

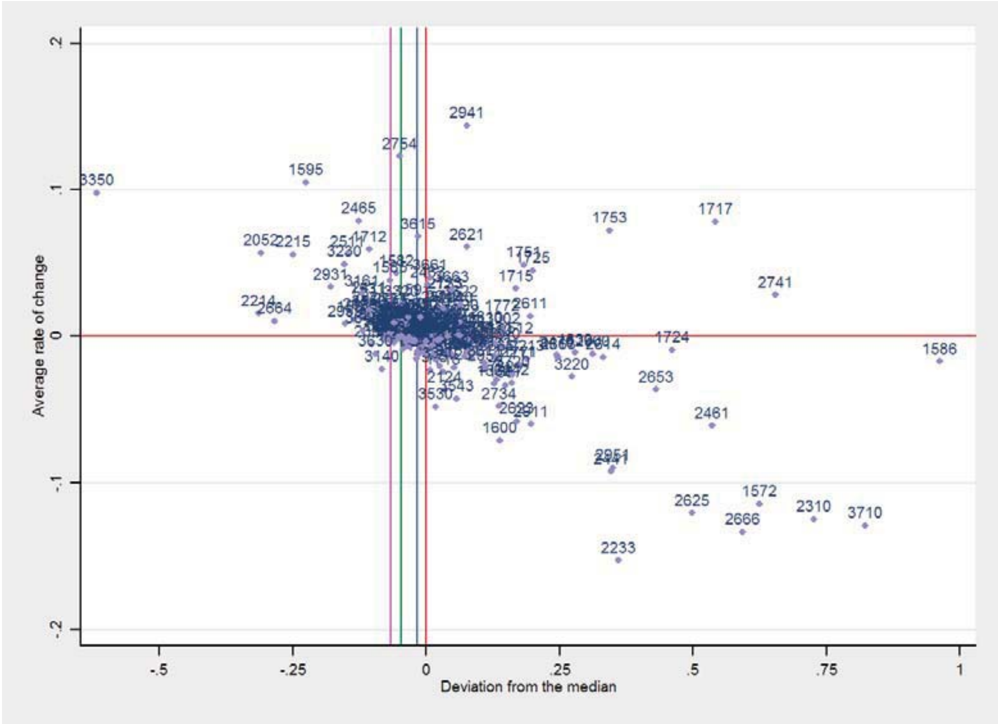


Figure 4.10: **Evolution of 4-digit industries**

The evolution of the industries may be examined by referring to the four quadrants of the plane, starting from the quadrant I (upper right)and going counter-clockwise. Industries taking place in quadrant I represent the ones which had high level of agglomeration in 2003 and experienced an increase in their agglomeration levels through the years. Table 4.21 presents the transition of 4-digit manufacturing industries between 2003 and 2008 with the purpose of displaying the direction of the change towards higher or lower agglomeration levels. Each column also shows the number of industries wirhin each range regarding the direction of change, positive or negative. There were 132 four-digit industries in 2003 which were highly agglomerated and in 2008 99 of them have remained in the in this category. 46 out of those 99 industries have experienced absolute increase in the index levels while 53 of them have declined. Quadrant I displays 40 out of these 46 industries as the origin refers to the median level which is higher than the 0.05 threshold for high agglomeration category. Industries taking place in this region, however, have different characteristics. For example industries 2941 and 2741 had a few large firms in 2003 (5 and 9 firms, respectively) and in 2008 there were even fewer firms in these industries (3 and 4 firms, respectively). Hence, increase in agglomeration in those industries may largely be attributable to increase in industrial concentration. Regarding industries 1717 and 1753, they had only 4 firms in each and this number increased to 14 and 10 firms, respectively. Even the number of firms increased, these industries are still industrially concentrated. On the other hand, industries such as 1715, 1725 and 1751 have great number of firms in 2003 and increased to 209, 259 and 261 firms in 2008, respectively. The increase in agglomeration in these industries may well be attributed to increase in raw geographic concentration rather than industrial concentration.

Quadrant II displays industries which have passed through a process of increasing geographic concentration, however reference lines distinguish these industries according to their agglomeration levels in 2003. Industries to the left of the purple line represent the ones with no agglomeration at all in 2003 but somehow increased their agglomeration through the period. Nevertheless, even though some industries have experienced increases in their agglomeration levels, they still have remained in the category of no agglomeration. First row of the table 4.21 shows the transition of the industries from no agglomeration category to others. Out of 30 industries (excluding 6 of them within no agglomeration and (-) sign since they are placed in quadrant III) which were not agglomerated at all in 2003, 8 of them were still in this category even their γ ameliorated. For example industries coded 3350, 2214 and 2664 have increased their agglomeration index by 0.49, 0.07 and 0.05 percentage points respectively but their agglomeration level did not change, remained at no agglomeration category. Besides, on the other hand, there exist some industries which had been categorized as not agglomerated in 2003 have changed category in terms of agglomeration levels, such that 5 industries have elevated to low and 4 industries to medium agglomeration group. For instance, industry 2052 shown in the second quadrant has experienced a transition from no agglomeration to medium level agglomeration category. Likewise, 13 industries have undergone a transition from no agglomeration to high agglomeration by attaining significant improvement in the index value such that above 0.25 percentage points, e.g. 1595, 2465, 1712, 2511, 2215 etc.. Except industry 1712, these industries had small number of firms (ranging between 5 to 14) in 2003 and even the number of firms doubled in these industries they became more geographically concentrated. So we may infer that those new firms have been located in a few regions leading to geographic concentration.

Industries falling in between the purple and green line are those with low level of geographic concentration in 2003 and experienced increase in its γ, hence enhanced in terms of agglomeration. Second row and positive columns of the table 4.21 show the number of firms in transition from low to other levels of agglomeration. Industries coded 2754 and 1582 may be given as examples to this group. In the same manner industries taking place between the green line and blue line represent those transited from medium level of agglomeration to the same category or a higher one. There are 14 industries in this group as the sum of positive columns of the third row shows. And finally industries lying within the blue and red line display the ones in high agglomeration category with a γ lying in range 0.05 and 0.067 and which still remained in this category in 2008. There are 6 industries taking place within this range but visually most notable one, industry 3615 may be given as an example with a 0.34 percentage points increase in the index value.[10]

Industries taking place in quadrant III are the ones which have γ's below the median and experienced a decline in index values. Negative columns of table 4.21 except high category reveals the number of industries in transition. 6 industries that were in the low category in 2003 have switched to no agglomeration in 2008. Amongst 8 industries that were categorized in the medium group in 2003, 4 of them switched to low and the other 4 to no agglomeration level.

Table 4.21: **Transition of 4-digit industries across different degrees of agglomeration between 2003 and 2008**

	γ in 2008								
γ in 2003	No agg.		Low		Medium		High		Total
change	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	
No agg.	8	6	5	0	4	0	13	0	36
Low	0	6	4	3	7	0	6	0	26
Medium	0	4	0	4	4	11	10	0	33
High	0	16	0	5	0	12	46	53	132
Total	8	32	9	12	15	23	75	53	227

And lastly, quadrant IV shows the transition of highly agglomerated industries which have experienced decreases in the level of γ's. First of all total number of industries in this group were 132 in 2003 where 16 of them have switched to no agglomeration, 5 of them to low and 12 of them to medium agglomeration category in 2008. 53 out of 132 highly agglomerated industries remained in the same category even they faced declines in their index levels. Furthermore, 99 of those industries have taken part within high agglomeration category in both years. Taking into account that there exists 128 industries in this category as of

[10] In fact, all highly agglomerated industries take place to the right of the blue line. In order to examine highly agglomerated industries which have increased their index value, one should refer to the region above the horizontal red line and to the right of blue line. There are 46 industries that were previously highly agglomerated and still so in 2008. As we are dealing with quadrants, 6 of them take place in quadrant II between the blue and red line and 40 of them in quadrant I.

2008, almost 77 % of these have also appeared within high category also in 2003.

4.3.2 Industrial Scope of Agglomeration

Taking into account the question "Do industry *groups* concentrate due merely to the fact that its *sub-industries* concentrate or there is a *common* effect on the industries of a higher industrial hierarchy group (e.g. agglomeration of 4-digit industries within 2-digit groups)?", co-agglomeration within industry groups are also examined.

The first statement implies that natural advantages and spillovers are *industry-specific* while the second implies them to be *group-specific*. To measure the degree to which the industries in the group are co-agglomerated; Ellison and Glaeser (1997) propose the use of a measure γ^c defined by

$$\gamma^c = \frac{\dfrac{G}{\left(1-\sum_m x_m^2\right)} - H - \left(1-\sum_j^r \hat{\gamma}_j w_j^2 (1-H_j)\right)}{\left(1-\sum_j^r w_j^2\right)} \tag{4.1}$$

where the industry group consists of r industries, H_j: plant Herfindahl of the jth industry, w_j: employment share of the jth industry in the group, $\hat{\gamma}_j$: agglomeration index of the jth industry, $H = \sum_j w_j^2 H_j$: group's plant Herfindahl index and G: raw concentration of employment in the group as a whole.

An estimate of $\gamma^c = 0$ may be interpreted as any spillovers/natural advantages found within the industry group are completely industry-specific. In other words there is no agglomeration in the industry *group* , hence agglomeration is observed simply owing to the concentration of its industries per se.

To measure the degree to which the industries in the group are co-agglomerated; equation (4.1) proposed by Ellison and Glaeser (1997) is explored for the case of Turkey manufacturing at the two-digit industry level for the 22 industry groups that contain more than one sub-industry. γ^c reflects the degree of correlation between the locations of establishments that belong to the same group and the scale of it is the same as that of γ. $\gamma^c = 0$ may be interpreted as indicating that there is no correlation across sub-industries, hence there is no agglomeration in the industry *group* and spillovers are industry-specific rather than group-specific.

Ellison and Glaeser (1997) also finds it useful to rescale the index γ^c with the weighted *EG*'s of the sub-industries to measure the strength of co-agglomerative forces relative to agglomerative forces.

$$\lambda = \frac{\gamma^c}{\sum w_j \hat{\gamma}_j} \tag{4.2}$$

A value of $\lambda = 0$ would indicate that sub-industries exhibit no co agglomeration at all, and a value $\lambda = 1$ indicate that natural advantages and/or spillovers that exist are group-specific rather than industry specific.

Table 4.22 reports the values of γ_c and λ obtained from four-digit sub-industries of each two-digit industry for year 2003. Regarding the index γ_c, 12 out of 22 two-digit industries exhibit co-agglomerative behaviour within the traditional ranges such that $\gamma_c \geq 0.02$. In 7 out of these 12 industries γ_c is found to be greater than 0.05, indicating that in these industries spillovers are group-specific rather than being industry-specific. The same characteristic is also valid for the year 2008 as shown in Table 4.23. 13 out of 22 two-digit industries exhibit co-agglomerative behaviour within the traditional ranges such that $\gamma_c \geq 0.02$. In 8 out of these 13 industries γ_c is found to be greater than 0.05

The five industries, namely, recycling (37), office machinery and computers (30), coke, refined petroleum products and nuclear fuel (23), motor vehicles, trailers and semi-trailers (34) and wearing apparel; dressing and dyeing of fur (18) appear within the most co-agglomerated in industries in both years. The spillovers are found to be group specific in these industries and this fact has been stable through the observed period. Ellison and Glaeser (1997) find substantial co-agglomeration of the three-digit sub-industries within the two-digit tobacco, textile, and lumber industries. In line with our findings, they find co-agglomeration in apparel and other textiles industry in US. Bertinelli and Decrop (2005) examines the co-agglomeration of the four-digit sub-industries within the two-digit according to the Nace 1.1 classification, likewise this study, for Belgian manufacturing industry. Hence their findings are directly comparable. The most co-agglomerated two-digit industries are found to be textile (17), clothes and fur industry (18), publishing, printing and reproduction of recorded media (22), production of medical, precision, optical and clock instruments (33) and production of office machines and computer materials (30). Excluding textile industry, their findings completely agree with our findings in terms of co-agglomeration patterns within two-digit industries.

Table 4.22: **Co-agglomeration within 2-digit industries (2003)**

NACE 2-digit	*Subind.*	H	G	γ^c	λ
37 Recycling	2	0.076	0.599	0.557	0.806
30 Office machinery and computers	2	0.095	0.268	0.372	-0.429
23 Coke, refined petroleum products and nuclear fuel	12	0.106	0.205	0.234	1.495
34 Motor vehicles, trailers and semi-trailers	3	0.012	0.101	0.108	0.922
20 Wood and of products of wood and cork, except furniture	6	0.002	0.073	0.066	0.593
18 Wearing apparel; dressing and dyeing of fur	6	0.001	0.054	0.064	0.898
15 Food products and beverages	30	0.001	0.049	0.052	0.393
19 Tanning and dressing of leather; man. of luggage, handbags, ...	3	0.003	0.047	0.050	0.873
31 Electrical machinery and apparatus n.e.c.	7	0.009	0.047	0.049	1.676
22 Publishing, printing and reproduction of recorded media	12	0.004	0.031	0.036	0.907
26 Other non-metallic mineral products	24	0.002	0.034	0.030	0.348
33 Medical, precision and optical instruments, watches and clocks	5	0.010	0.030	0.020	0.181
24 Chemicals and chemical products	18	0.008	0.039	0.015	0.124
27 Basic metals	16	0.028	0.064	0.010	0.081
17 Textiles	19	0.002	0.016	0.009	0.088
25 Rubber and plastic products	7	0.004	0.013	0.009	0.409
29 Machinery and equipment n.e.c.	22	0.002	0.009	0.005	0.078
36 Furniture; manufacturing n.e.c.	12	0.004	0.014	0.004	0.055
21 Pulp, paper and paper products	7	0.007	0.014	-0.001	-0.030
28 Fabricated metal products, except machinery and equipment	16	0.001	0.004	-0.003	-0.051
32 Radio, TV and communication equipment and apparatus	3	0.100	0.073	-0.004	-0.089
35 Other transport equipment	8	0.029	0.035	-0.090	-0.458

Table 4.23: **Co-agglomeration within 2-digit industries (2008)**

NACE 2-digit	H	G	γ^c	λ
23 Coke, refined petroleum products and nuclear fuel	0.123	0.216	0.272	1.942
37 Recycling	0.119	0.263	0.236	1.284
33 Medical, precision and optical instruments, watches and clocks	0.036	0.124	0.122	1.186
34 Motor vehicles, trailers and semi-trailers	0.014	0.101	0.110	1.082
18 Wearing apparel; dressing and dyeing of fur	0.001	0.062	0.071	0.876
30 Office machinery and computers	0.115	0.236	0.065	0.339
22 Publishing, printing and reproduction of recorded media	0.003	0.053	0.059	1.046
19 Tanning and dressing of leather; man. of luggage, handbags, ...	0.006	0.052	0.054	0.811
32 Radio, TV and communication equipment and apparatus	0.081	0.138	0.050	0.322
20 Wood and of products of wood and cork, except furniture	0.003	0.046	0.048	0.690
15 Food products and beverages	0.001	0.040	0.042	0.330
26 Other non-metallic mineral products	0.002	0.040	0.039	0.404
17 Textiles	0.001	0.029	0.022	0.140
21 Pulp, paper and paper products	0.006	0.021	0.015	0.404
24 Chemicals and chemical products	0.004	0.022	0.015	0.248
27 Basic metals	0.014	0.047	0.009	0.083
31 Electrical machinery and apparatus n.e.c.	0.007	0.018	0.004	0.092
28 Fabricated metal products, except machinery and equipment	0.001	0.007	0.002	0.042
29 Machinery and equipment n.e.c.	0.003	0.005	0.000	0.003
35 Other transport equipment	0.008	0.105	-0.000	-0.001
36 Furniture; manufacturing n.e.c.	0.002	0.016	-0.001	-0.009
25 Rubber and plastic products	0.002	0.004	-0.002	-0.068

One may argue that, co-agglomerative forces operating at different technology levels might be different and NACE classification system may fail to capture this fact. For instance, it may be expected that high tech industries are more likely to agglomerate in order to benefit from knowledge spillovers while low-tech and medium tech industries are more likely to gather together to take the advantage of input sharing and labour pooling. In order to see whether a potentially inappropriate definition of industrial activities masks this, we group industries according to technology levels in accordance with OECD classification and compute co-agglomeration index based on this. Table 4.24 displays that low and medium-low tech industries do not exhibit co-agglomeration behaviour, whereas medium-high tech industries show a considerable degree of co-agglomeration with a λ of 0.0769 and high tech industries show a moderate degree of co-agglomeration with a λ of 0.327.

Table 4.24: **Co-agglomeration within technology groups (2003)**

Technology level	Num.ind.	H	G	γ^c	λ
Low	97	0.0004	0.0013	-0.0024	-0.0260
Medium-low	67	0.0013	0.0045	0.0010	0.0131
Medium-high	53	0.0017	0.0095	0.0058	0.0769
High	11	0.0159	0.0485	0.0046	0.0327

4.3.3 Coagglomeration Between Industry Pairs

In addition to co-agglomeration within groups, co-agglomeration between industry pairs is also examined following Ellison, Glaeser, and Kerr (2010) which quantify industry pair co-agglomeration (as opposed to larger groups) in a way to reveal which particular industry pairs tend to locate together.

When *EG* (1997) co-agglomeration index is applied to industry pairs i and j instead of larger groups, it takes such a simple form :

$$\gamma_{ij}^c = \frac{\sum_{m=1}^{M}(s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^{M} x_m^2} \tag{4.3}$$

Ellison, Glaeser, and Kerr (2010) quantify industry pair co-agglomeration (as opposed to larger groups) in a way to reveal which particular industry pairs tend to locate together.

For the case of Turkey by exploring eq.(4.3), the sample contains 26335 industry pair observations for years 2003 and 2008: all distinct co-agglomeration pairs from 230 4-digit industries after discarding own-industry co-agglomerations. 4.25 shows the distribution of γ_{ij}^c across all industries. Around 14 per cent of industry pairs have medium level of agglomeration for both years. In 2003, 21.5 per cent and in 2008 19.2 per cent of the industry-pair co-agglomerations are calculated to be greater than 0.05, indicating higher levels of co-agglomeration.

Table 4.25: **Distribution of γ_{ij}^c across all industries**

	2003		**2008**	
	N	%	N	%
$\gamma_{ij}^c < 0$	13,354	50.71	13,280	50.43
$0 < \gamma_{ij}^c \leq 0.02$	3,677	13.96	4,075	15.47
$0.02 < \gamma_{ij}^c \leq 0.05$	3,656	13.88	3,920	14.89
$\gamma_{ij}^c > 0.05$	5,648	21.45	5,060	19.21
All	26,335	100.00	26,335	100.00

Ellison, Glaeser, and Kerr (2010) calculated the mean *EG* pairwise co-agglomeration as approximately zero. They relate this largely to the definition of the index such that their "benchmark measure of an area's *size* is its share of manufacturing employment, so each industry's deviations from the benchmark will be approximately uncorrelated with the average of the deviations of all other industries" (Ellison, Glaeser, and Kerr, 2010, p. 1199).

They regard the standard deviation of the co-agglomeration index as more interesting due to the fact that it reflects the extent to which industry pairs are positively and negatively co-agglomerated. So it is calculated to be 0.013 at the state level. They state that this measure is comparable with the mean within industry agglomeration level of 0.051 in Ellison and Glaeser (1997). As shown in Table 4.26 and 4.27 for the case of Turkey, mean of within industry agglomeration has been calculated as 0.106 and the standard deviation of the pairwise co-agglomeration index is calculated as 0.087 for the year 2003. These measures are calculated as 0.105 and 0.077 for the year 2008, respectively.

Table 4.26: **Descriptive statistics - All industries 2003**

	μ	σ	**N**	**median**	**min**	**max**	**skewness**	**kurtosis**
$\gamma_{ij}^c < 0$	-0.058	0.053	13354	-0.042	-0.320	-0.000	-1.511	5.378
$0 < \gamma_{ij}^c \leq 0.02$	0.009	0.006	3677	0.009	0.000	0.020	0.134	1.845
$0.02 < \gamma_{ij}^c \leq 0.05$	0.034	0.009	3656	0.033	0.020	0.050	0.183	1.847
$\gamma_{ij}^{>}0.05$	0.115	0.079	5648	0.090	0.050	1.003	3.227	20.267
All	0.001	0.087	26335	-0.001	-0.320	1.003	1.002	9.822

Table 4.27: **Descriptive statistics - All industries 2008**

	μ	σ	**N**	**median**	**min**	**max**	**skewness**	**kurtosis**
$\gamma_{ij}^c < 0$	-0.052	.0472	13,280	-0.038	-0.296	-0.000	-1.427	5.094
$0 < \gamma_{ij}^c \leq 0.02$	0.009	0.006	4,075	0.009	0.000	0.020	0.107	1.809
$0.02 < \gamma_{ij}^c \leq 0.05$	0.034	0.009	3,920	0.033	0.02	0.050	0.140	1.838
$\gamma_{ij}^{>}0.05$	0.109	0.070	5,060	0.086	0.050	0.906	2.988	18.299
All	0.001	0.077	26,335	-0.0005	-0.296	0.906	0.912	8.920

Figures 4.11 and 4.12 show the distribution of γ^c for all industries for years 2003 and 2008 respectively.

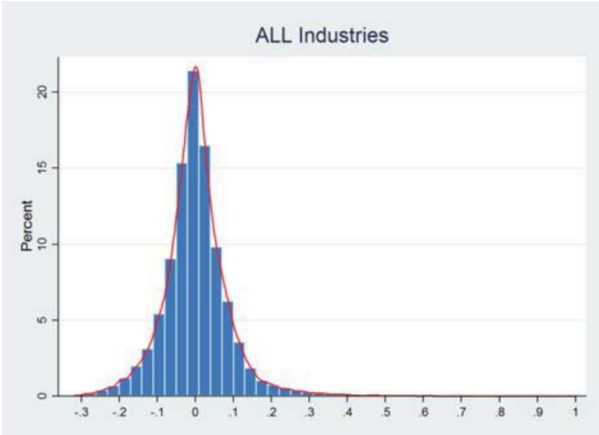


Figure 4.11: **Histogram of γ^c for all industries 2003**

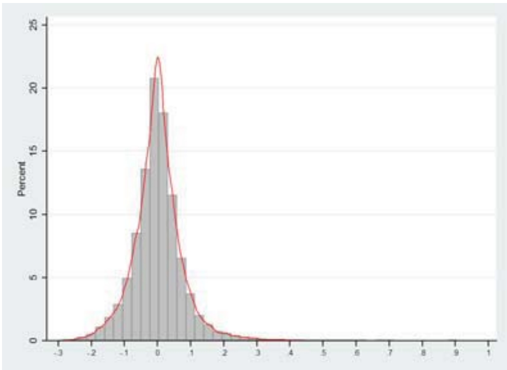


Figure 4.12: **Histogram of γ^c for all industries 2008**

Industry pairs with highest pairwise co-agglomerations are shown in Table 4.28 and 4.29. In 2003, broadly speaking, chemicals, other non-metallic mineral products and basic metal industries are paired with paper products, food and beverage, radio and TV equipment, other non-metallic mineral products. For the year 2008 textile , other non-metallic mineral products and chemical industries show co-agglomeration with textile, basic metals and furniture industries.

4.3.4 International Comparisons

Ellison and Glaeser (1997, p. 890) point out that "the index is designed to facilitate comparisons across industries, across countries, or over time". However, one should be cautious when dealing with comparisons. Regarding comparisons across industries, sectors entailing few observations should be considered carefully, as the index tends to be upward biased. Hence performing cross-country comparisons is even more critical. Apart from this, comparing same sectors in different countries raises several issues. Making comparisons across different industrial classifications

Table 4.28: **Highest Pairwise Co-agglomerations 2003**

Rank	Industry 1	Industry 2	γ_{12}^{c}
1	Precious metals production(2741)	(2111)Publishing of books	1.003
2	Basic pharmaceutical products(2441)	(2111)Publishing of books	0.918
3	Recycling of metal waste and scrap(3710)	(1572)Prepared pet foods	0.904
4	Coke oven products(2310)	(1572)Prepared pet foods	0.873
5	Coke oven products(2310)	(3710)Recycling of metal waste & scrap	0.864
6	Precious metals production(2741)	(2441)Basic pharmaceutical products	0.842
7	Ceramic insulators & insulating fittings(2623)	(2666)Other articles of concrete, plaster & cement	0.726
8	Ceramic insulators & insulating fittings(2623)	(3220)Instruments & appliances for measuring, checking,...	0.683
9	Other articles of concrete, plaster & cement(2666)	(3220)Instruments & appliances for measuring, checking,...	0.626
10	Ceramic insulators & insulating fittings(2623)	(2653) Plaster	0.616
11	Explosives(2461)	(2623) Ceramic insulators & insulating fittings	0.612
12	Plaster(2653)	(2666)Other articles of concrete, plaster & cement	0.606
13	Explosives(2461)	(2666)Other articles of concrete, plaster & cement & cement	0.602
14	Plastics in primary forms(2416)	(3710)Recycling of metal waste & scrap	0.573
15	Ceramic insulators & insulating fittings(2623)	(3530)Aircraft and spacecraft	0.572

Table 4.29: **Highest Pairwise Co-agglomerations 2008**

Rank	Industry 1	Industry 2	γ_{12}^{c}
1	Portable hand held power tools(2941)	(1586)Processing of tea & coffee	0.906
2	Throwing & preparation of silk...(1715)	(1717)Prep. & spinning of other textile fibres	0.594
3	Throwing & preparation of silk...(1715)	(1753)Non-wovens & articles made from it,...	0.529
4	Prep. & spinning of other textile fibres(1717)	(1751)Carpets and rugs	0.695
5	Prep. & spinning of other textile fibres(1717)	(1753)Non-wovens & articles made from it,...	0.884
6	Silk-type weaving(1724)	(1725)Other textile weaving	0.510
7	Carpets and rugs(1751)	(1753)Non-wovens & articles made from it,...	0.626
8	Prepared unrecorded media(2465)	(3621)Striking of coins	0.552
9	Prepared unrecorded media(2465)	(3661)Imitation jewellery	0.521
10	Glass fibres(2614)	(2732)Cold rolling of narrow strip	0.616
11	Glass fibres(2614)	(2754)Casting of other non-ferrous metals	0.591
12	Ceramic household & ornamental art.(2621)	(2741)Precious metals production	0.678
13	Cold rolling of narrow strip(2732)	(2754)Casting of other non-ferrous metals	0.675
14	TV & radio receivers, sound or video...(3230)	(2741)Precious metals production	0.491
15	Imitation jewellery(3661)	(3621)Striking of coins	0.532

may make the analysis inaccurate. Even this issue may easily be dealt by using or transforming to same classification scheme, the issue related to the spatial units is not an easy one to solve. When exploring concentration indices, one is limited to use existing spatial units within the country for which it's not easy to find comparable counterparts across countries. This is of great importance since the size of the index is very sensitive to locational fineness of the data, the index tends to be higher at more coarsened spatial units. A possible explanation lies behind the assumptions of the location model that treats natural advantages and spillovers as being uncorrelated across space. But if spillovers reach beyond borders or natural advantages are correlated across spatial units, measuring index at that spatial unit becomes incompatible with the true location model and thus nonsense (Alecke, Alsleben, Scharr, and Untiedt, 2006). Briefly, in such a case, the index fails to capture the entire range of spillovers and natural advantages as they are operating at a higher spatial unit, hence it is calculated lower. This arises due to the a-spatial nature of the index.

Bearing in mind all these potential shortcomings in making cross-country

Table 4.30: **Comparison with other countries**

Author(s)	Country	Mean *EG*
Ellison and Glaeser (1997)	USA	0.051
Rosenthal and Strange (2001)	USA	0.048
Devereux, Griffith, and Simpson (2004)	UK	0.033
Bertinelli and Decrop (2005)	Belgium	0.040*
Barrios, Bertinelli, Strobl, and Teixeira (2005)	Ireland-Portugal	0.042*-0.095*
Alecke, Alsleben, Scharr, and Untiedt (2006)	Germany	0.036
Lafourcade and Mion (2007)	Italy	0.033
Barrios, Bertinelli, Strobl, and Teixeira (2009)	Belgium-Ireland-Portugal	0.027*-0.038*-0.133*
Lu and Tao (2009)	China	0.014
Leahy, Palangkaraya, and Yong (2010)	Austria	0.044
This study	Turkey	0.112

∗weighted means

comparisons, still a cross country comparison may be performed cautiously on broader terms. To have an idea about the extent of agglomeration in different countries Table 4.30 shows mean levels of *EG* indices calculated by some notable studies as all of these studies well go beyond this in terms of research topic. A striking point is that the extent of overall agglomeration is similar for developed countries ranging between 0.03 and 0.05 which falls within the ranges of medium level agglomeration according to Ellison and Glaeser's (1997) classification. Two countries, namely Portugal and China, lie outside this range with values 0.133 (and 0.095 in Barrios, Bertinelli, Strobl, and Teixeira (2005)) and 0.014 respectively. Portugal having higher mean values of *EG* relative to other countries is associated with quite low levels of industrial concentration, as both Barrios, Bertinelli, Strobl, and Teixeira (2005) and Barrios, Bertinelli, Strobl, and Teixeira (2009) agree on this fact. On the other hand China's industrial agglomeration is observed to be lower when compared with those in developed countries. Lu and Tao (2009) relate this to some institutional factors such as local protectionism which may preclude the process of industrial agglomeration in China.

In this picture, Turkey arises as having high levels of agglomeration compared to developed countries with an average *EG* index of 0.112. This fact may hinge on to the differences in terms of transportation costs, labour market conditions, and more broadly any other factors influencing the location of plants across countries considered. Notwithstanding that may well be a possible case for Turkey as a developing country, it should completely be discussed in a deeper research framework.

Again broadly speaking, there is a stylized fact arising from the research on agglomeration that traditional and low-tech industries tend to show higher degrees of agglomeration relative to others. A notable inference that can be retrieved from the studies mentioned above is the presence of *textile* (or textile-related industries) ranking high in many of the countries, for instance in US (Ellison and Glaeser (1997), Rosenthal and Strange (2001)), UK (Devereux, Griffith, and Simpson (2004)), Belgium (Bertinelli and Decrop (2005)), Spain (Alonso-Villar, Chamorro-Rivas, and González-Cerdeira (2004)) and Italy (Lafourcade and Mion (2007)) for the sample of small plants. In this sector high proportion of the labour is unskilled and Ellison and Glaeser (1999) find that access to unskilled labour to be the most important factor in explaining geographic concentration for the textiles and apparel industries in US. Along with *textile*, *extraction* and *mining* industries are found to be among the most agglomerated ones. Alecke, Alsleben, Scharr, and Untiedt (2006) state that *extractive* industries dominate the top group within German manufacturing industries. In Alonso-Villar, Chamorro-Rivas, and González-Cerdeira (2004) also, *extractive* industries are found to be highly agglomerated for the Spanish case. Here, natural advantages arise as a plausible candidate for explaining agglomeration in these type of industries. Along with these sectors, in general there is a clear evidence on the high agglomeration of low-tech industries. These stylized facts are also valid for Turkey, 80 per cent of the 15 most agglomerated industries fall within the low and medium-low technology sectors. Consistent with the previous findings textile and traditional sectors dominate the group.

Taking into account the literature highlighting knowledge externalities, high-tech industries may be anticipated to appear among the most agglomerated ones. However, as a common aspect across country studies, high-tech industries do nor rank high. Agglomeration to a certain degree in high-tech industries is supported in a few studies including Ellison and Glaeser (1997), Alonso-Villar, Chamorro-Rivas, and González-Cerdeira (2004), Alecke, Alsleben, Scharr, and Untiedt (2006) and Bertinelli and Decrop (2005). There might be some reasons for these sectors to be less agglomerated. One reason is that "they are newer and agglomeration is a dynamic process and geographic concentration in these sectors might still be at an early stage" (Devereux, Griffith, and Simpson, 2004, p. 545). Another reason might be that even technological spillovers are important, geographic contiguity may be less important in today's world in capturing knowledge externalities due to the developments in communication technologies (Devereux, Griffith, and Simpson, 2004). Regarding Turkey, high-tech industries also exhibit lower levels of agglomeration such that almost half of the industries listed within the least agglomerated industries are high and medium-high tech industries.

5 DETERMINANTS OF AGGLOMERATION IN TURKISH MANUFACTURING INDUSTRY

The theoretical literature on agglomeration economies has been vast in the sense that scholars from various disciplines (regional & urban economics, NEG) have developed formal theoretical models explaining why economic activity ends up being concentrated in specific areas, which is already examined in the first chapter. In line with the theoretical developments in the area, questions regarding quantification and measurement of agglomeration apparently have come to light and has shown significant improvement over the last two decades, as mentioned in the second chapter. However, indices measuring geographic concentration lack providing information on the sources of agglomeration since generally they are not derived from theoretically founded models. Ellison and Glaeser (1997)'s index, even though have been derived from a location choice model, does not differentiate different sources of agglomeration either due to spillovers or natural advantages. Thus, examining the sources of agglomeration remains still as an interesting subject to be studied for researchers. Nevertheless, empirical evidence on the determinants/sources of agglomeration put forth by theoretical models still deemed to be relatively limited. A possible explanation behind that might be the difficulty in proxying different sources of agglomeration provided by the theory.

This section firstly reviews the existing empirical literature on determinants of agglomeration in order to gain an accurate understanding of theories that pave the way for empirical testing [1]. As a second step, a brief presentation of data, proxies for sources of agglomeration and the model will be presented in accordance with the existing literature. And lastly, it attempts to examine the determinants of agglomeration in Turkish manufacturing industries.

[1]The empirical testing of NEG models is out of the scope of this study. For the empirical literature on the predictions of NEG and urban agglomeration economies, the reader may refer to Rosenthal and Strange (2004) Head and Mayer (2004), Brakman, Garretsen, Gorter, van der Horst, and Schramm (2005), Bosker, Brakman, Garretsen, and Schramm (2010), Combes (2011), Puga (2010) and Redding (2010).

5.1 Related Literature

Empirical studies dealing with the determinants of agglomeration, basically follow a common modelling strategy where an index of geographic concentration as a dependent variable is regressed on a set of independent variables to stand as proxies for the sources of agglomeration. Generally these proxies are employed in order to account for *Marshallian externalities*, *transportation costs* and *internal increasing returns* á la NEG, *natural advantages* and other industry specific factors that are likely to affect agglomeration. The literature will be reviewed in terms of data used, proxies employed, modelling strategy and main results obtained in these studies. Figure 5.1 list detailed information related to data, models employed and variables used in some of the notable empirical studies in this subject.

Kim, Barkley, and Henry (2000) investigate the industry characteristics linked to establishment concentration in manufacturing industries in US non-metropolitan areas. They give a special importance to non-metropolitan areas because these areas inherit some disadvantages such as sparse local markets, thin labour pools, geographic isolation and lacking economic diversity, which are likely to be overcome by the externalities created by agglomerations. They relate industry agglomeration propensities to *industry structure, work force characteristics* and *buy-and-sell linkages* and provide a review of literature under each of these headings in order to form a basis for the empirical examination. They use k dispersion parameter [2] to measure non-metropolitan agglomeration propensities and subsequently as a dependent variable.The effect of industry structure on agglomeration is discussed elaborately in the study and takes an important place in the literature. So, it will be mentioned more in detail here. To account for *industry structure* they use average plant size, multi plant firms, product specialization and vertical integration as dependent variables. The effect of average plant size on agglomeration is expected to be ambiguous due to two factors. If relatively large production facilities contributes to spin-offs of new firms, attract input suppliers and specialized services to the area and encourage the develop-

[2] They prefer using k dispersion parameter due to shortcomings of the other concentration measures such as locational Gini, spatial HHI, EG. Basically in this methodology, the observed distribution of establishments is compared with the values predicted by a Poisson distribution (as a benchmark random distribution). If the observed and predicted values are significantly different, the null hypothesis of random distribution is rejected and observations are said to exhibit contiguous behaviour. Then, they are fitted to the negative binomial distribution which is developed to account for distributions in which there is over-dispersion. If the observed spatial distribution of firms are approximated by the negative binomial distribution, the exponent k of the negative binomial distribution is used as a measure of concentration in the industry

ment of labor pools with industry-specific skills, it may have a positive effect on agglomeration. on the other hand, they may contribute to higher local input prices which would result in spatial dispersion. Multi plant firms are expected to be less spatially concentrated as this structure provides an alternative to spatial clustering for reducing costs of technical information, coordinating activities etc. Product diversity in the industry may stimulate agglomeration to enhance inter-firm networking, reduce information exchange and transaction costs among firms. Alternatively industry product diversity may lead to spatial dispersion if establishments in the industry specialize in different products which require different production process, dissimilar input and product markets. So the affect of product diversity on agglomeration is also ambiguous. Lastly, vertical integration is expected to have negative effect on agglomeration as they are less dependent on local input and product markets and less reliant on inter-firm networking for information and technology exchange.

To account for *workforce characteristics* three variables (low, medium, high) related to skills requirements and a proxy variable for labour intensity is used as shown in Figure 5.1. A positive relation is anticipated in three labour skill categories if industry clusters are associated with labour pools with specific skills. Industry labour intensity will be related with agglomeration if labour pooling is most beneficial to labour intensive industries. *Buy-and-sell linkages* are captured by four input or output linkage variables. In three of the variables (ratio of input expenditures for raw materials, ratio of expenditure for external R&D, technical and professional services and ratio of input expenditure from the same three-digit industry) a positive relationship with agglomeration is expected if industry establishments select similar locations in order to gain from proximity to specialized input and product markets. Lastly a proxy variable for industries with highly localized input (e.g. logging, sawmills, millwork, grain mill products, meat products) or product (e.g. newspaper, commercial printing, concrete and gypsum, metalworking) markets is selected. As dependence on local markets discourage clustering, a negative relationship is expected.

The relationship between concentration and industry characteristics is estimated by using OLS. It is worth noting that, smaller k values reflect more concentration, thus negative coefficients indicate an increasing effect on the concentration levels. The findings of the study indicate that agglomeration is positively associated with average establishment size, raw material inputs and labour intensity of the production process. In addition, employment shares of high skill and low skill occupations are found to be positively affecting clustering of firms. On the other hand, multi plant firm structure, a relatively large share of medium skilled workers and reliance on local input and product markets have negative impacts on agglomeration of establishments.

Rosenthal and Strange (2001), examine determinants of agglomeration in US manufacturing industries and focus on the three micro-foundations of agglomeration that have been outstanding in the literature, namely knowledge spillovers, labour pooling and input sharing. They also control for transportation costs and natural advantages. Dependent variable they use to measure concentration is EG index. Proxies employed for related theories are generally straightforward as may be tracked in Figure 5.1, however the one accounting for transportation costs is a kind of ingenious. To account for per mile cost of shipping the product, *inventories per $ of shipment*, defined as the value of end-of-year inventories divided by the value of shipments is employed. "Industries that produce highly perishable products face high product shipping costs per unit distance and, therefore, will seek to locate close to their markets, ceteris paribus. With multiple markets, such industries will tend to display less agglomeration. Conversely, industries that produce non-perishable products face lower product shipping costs and should display more agglomeration" (Rosenthal and Strange, 2001, p. 202). They conduct the analysis separately at different geographic levels (zip code, country and state) by using OLS and 2-digit industry level fixed effects estimation. They show that different agglomerative forces are at work at different geographical scales. Results indicate that labour market pooling hypothesis is found to be most robust positively affecting agglomeration at all levels of geography. On the other hand knowledge spillovers have positive influence only at the zip code level as they attenuate rapidly. Their contribution takes place at the local level. Reliance on manufactured inputs and natural resources contributes to agglomeration at the state level but not evident at other geographical levels.

In the same vein with Rosenthal and Strange (2001), **Alecke, Alsleben, Scharr, and Untiedt (2006)** examine the determinants of agglomeration in German manufacturing industries by relating the EG index to a variety of industry characteristics to find out which Marshallian externalities are actually at work. Also controlling for transportation costs, increasing internal returns, natural advantages and potential congestion effects, they also conduct analysis at different aggregation levels of geography. They find strong evidence that transportation costs significantly reduce agglomeration while increasing internal returns increase it which of both findings are coherent with the predictions of new economic geography. Input sharing hypothesis is supported at the county level while for labour pooling there is evidence in a more aggregated geography level. However knowledge spillovers are not supported by German data. They provide two explanations for that. One is that spillovers could work at an extremely localized level that their approach can not capture. And the other is knowledge spillovers are spatially bounded due to substantial spatial transaction costs in Germany, which justifies their approach.

Likewise, **Barrios, Bertinelli, Strobl, and Teixeira (2009)** provide en empirical analysis of the determinants of agglomeration for three small European countries, namely Belgium, Ireland and Portugal where again EG index is regressed on a set of variables representing various theoretical approaches put forward in the literature. None of the theories were supported for the case of Ireland where all coefficients were found to be insignificant. For Belgium, only backward linkages argument of NEG is supported and found to have a positive impact on agglomeration as expected. Regarding Portugal, backward linkages, knowledge spillovers and labour pooling hypotheses are all supported empirically with an expected positive role on agglomeration. However, they note that their results are not completely robust due to data quality issues and add that cross-country regressions may not be the right approach to capture determinants of agglomeration.

Besides empirical studies on developed countries, **Lu and Tao (2009)** investigate the determinants of geographic concentration of Chinese manufacturing industries with a special focus given to local protectionism. They argue that with the economic reform in 1978, market forces for industrial agglomeration should have ameliorated some of the poor location choices of economic activities caused by the central planning. However, they also state that the same reform has led to the rise of local protectionism in some regions which decelerates market driven agglomeration process. In their study they provide strong evidence supporting the obstructor role of local protectionism on agglomeration and this result is robust to the use instrumental variable estimation and omitted variables problems. Meanwhile, traditional determinants of agglomeration are largely found to be consistent with the previous findings in the literature. Knowledge spillovers and input sharing proxies are found to be significant at all geographic levels lending them a positive role for agglomeration. Labour market pooling is supported only at city level with a positive effect while scale economies and natural resource endowments do have significant positive effects at each geographic level.

Author	Year	Country	Data	Industry class.	Spatial unit	Theory	Model	Dep. v.	Control variables	Type of spillovers	Expected sign
Kim, Barkley and Henry	2000	US	1992 - Enhanced County Business Patterns (Establishment data); 1992 - Census of Manufacturers (Industry characteristics data); 1990 - Census of Population (Industry occupational data)	SIC / 3-digit (140) (In the analysis 113 SIC ind. used due to estimation results)	Multi-county Component Economic Areas (CEAs) - nonmetropolitan CEAs (38)	External economies: static localization economies vs MAR dynamic externalities, NEG	OLS - two estimations. In estimation 2, labor skill percentages are weighted by industry labor intensity	*k*(unweighted)*, k^{a}(geo.size weighted) & k^{t}(emp. weighted)	**Average establishment size** (total emp./num. of establishments)	Industry structure	ambiguous (+/-)
									Multiplant firms (num. of est./number of companies)		-
									Product spec. (Sum sq. of commodity shares of toti output)		ambiguous (+/-)
									Vertical integration (Industry VA/Ind. tot. output)		-
									High skill (Percent emp. professional and technical)	Labor pooling effects	+
									Med. skill (Percent emp. pericision prod., craft, repair)		+
									Low skill (Percent emp. operators, fabricators, laborers)		+
									Labor intensity (Industy payroll/Industry VA)		ambiguous (+/-)
									Percent input purchases from agriculture, forestry, fishing & mining	Buy and sell linkages	+
									Percent input purchases from External R&D services		+
									Percent input purchases from the same 3-digit ind.		+
									Number of establishments in the industry in nonmet. counties		-
Rosenthal and Strange	2001	US	2000-Q4 establishment-level data from Dun and Bradstreet (D&B) MarketPlace	SIC / 4-digit (459) 3-digit (140) 2-digit (20)	Zipcode, County, State	Marshallian externalities, NEG and NTT	OLS & 2-digit fixed effects & 3-digit fixed effects	EG	***Nonmanufactured inputs per $ of shipment***	Input sharing	+
									Manufactured inputs per $ of shipment (cost of inputs purchased from manuf. sec./total shipments)		+
									Net productivity (value of shipments - value of inputs.)/number of workers	Labor market pooling	+
									Management workers / (Management + Production) workers		+
									Percentage of workers with PhD, Master's and Bachelor's degree		+
									Innovations per $ of shipment (firms<500)	Knowledge spillovers	+
									Innovations per $ of shipment (firms>500)		+
									Inventories per $ of shipment (value of end-of-year inv./(value of shipments)	Transportation costs	+
									Energy per $ shipment	Natural advantages	+
									Natural resources per $ shipment		+
									Water per $ shipment		+
Alecke et al.	2006	Germany	1998 plant-level data	Nace Rev. 1.1 / 3-digit (116 ind.) 2-digit (27 ind.)	County (440), Labor market region (225), Planning region (97)	Marshallian externalities, NEG and NTT	OLS	EG	***Share of services in inputs*** (technical and industrial services/total shipments)	Input sharing	+
									Share of manufactured goods in inputs (manufactured inputs/total shipments)		+
									Share of workers with specialized occupation (emp. in spec. occ./overall emp.)	Labor market pooling	+
									Share of workers with university degree (workers uni. degree/overall emp.)		+
									Dummy for "high-tech industry" ***Dummy for "medium-tech industry"***	Knowledge spillovers	+
									Trade costs (tons/€) weight (imports+exports)/value (imports+exports)	Transportation costs	+
									Av. establishment size	Internal increasing returns	+
									Dummy for "extractive industries"	Natural advantages	+
									Size (total emp. of industry)	Potential congestion effects	
Barrios et al.	2009	Belgium	1998 - Social Security database	ISIC Rev. 1 / 4-digit (63)	*Arrondissement* (43-NUTS3), *Communes* (589-NUTS5)	NEG & Marshallian externalities	OLS (2-digit ISIC dummies) & Oaxaca decomposition	EG	***Total purchases of goods & services*** (total inputs-mil.€)	Forward & backward linkages (NEG)	+
		Ireland	1998 - Forfas Emp. Survey		*County* (27-NUTS4), *Township* (504-btw.NUTS4&5)				***Sectorial share of skilled works*** (% of tot.pop. with higher educ. diploma)	Labor pooling	+
		Portugal	1998 - Ministry of Employment		*Distritos* (18-rougher than NUTS3), *Concelhos-municipos* (275-NUTS4)				***Total intra-mural R&D expenditure*** (mil.€)	Knowledge spillovers	+
									Purchases of energy products (mil.€)	Natural advantages	+
									Gross investment in tangible goods (mil.€)		+
									Gross investment in machinery (mil.€)		+
									Av. plant size (persons employed/number of plants)	Proxy for increasing returns	+
Lu and Tao	2009	China	Annual Survey of Industrial Firms (ASIF) 1998-2005 -Firm level data, not plant.	Chinese Standart of Industrial Classification, 4-digit (540) 3-digit (171) 2-digit (29)	County, city, province	Marshallian externalities, resource endowments and scale economies	Pooled OLS & industry fixed effects & IV estimation (3 geog. levels and 3 estimations)	EG	**Share of SOEs in employment**	local protectionism	_
									Wage premium $wage\ prem_i = \frac{\sum_r \frac{wage_{ir}}{wage_r} * Emp_{ir}}{\sum_r Emp_{ir}}$	Labor pooling	+
									Purchased-inputs intensity (Purchased-inputs/total output)	input sharing	+
									New products to output ratio (Total new products of an industry/total output of the industry)	Knowledge spillovers	+
									Average firm size (Total output of an industry/number of firms in the industry)	Scale economies	+
									Agricultural products usage ratio (Share of inputs from agricultural sectors)	Resource endowments	+
									Mining products usage ratio (Share of inputs from mining sectors)		+
									Share of SOEs in 1985	IV for Share of SOEs 1998-2005	

Figure 5.1: **Control variables in selected literature**

5.2 Explaining Agglomeration

As mentioned in section 3.1.5, the *EG* index can not distinguish between natural advantages and spillovers that may drive agglomeration. Moreover, spillovers are captured in a broad sense. In accordance with the theoretical sources of agglomeration discussed in section 2.2, the degree to which agglomerative externalities explain interindustry differences in geographic concentration is analysed in this section. Accordingly, as a modelling strategy *EG* index is regressed on a set of variables that proxy Marshallian externalities and other controls as sources of agglomeration. Firstly, the data is introduced in order to get an insight about its structure. Then, dependent variables that are constructed consistent with the empirical literature is presented. Finally, regression results are displayed.

5.2.1 Data and Methodology

In order to conduct analysis of determinants of agglomeration, three different datasets are exploited. The first one is the "Annual Industry and Service Statistics" (AISS) which is also employed in the previous chapter. However previous chapter made use of local units dataset which allowed us to consider locational information of plants belonging to establishments in the main data. Now, since the analysis requires to gather information on various establishment level variables (as mentioned in section 4.2) and obtain industrial aggregates by using them, the main dataset of AISS is used. Other two datasets explored include "Foreign Trade Statistics" and "Research and Development Activities Survey of Industry and Service Enterprises" compiled by Turkish Statistical Institute which of both are establishment level micro data. These datasets are used in order to derive variables to proxy transportation costs and knowledge spillovers, respectively, which are explained in detail in the next subsection. First, these three datasets are merged by using a common key variable and then aggregated at the 4-digit industry level in order to get control variables that would reflect industrial characteristics. The dependent variable of the analysis comes from the previous chapter, *EG* indices of agglomeration calculated again for 4-digit industries. The restriction of the analysis to NACE Rev. 1.1 classification in the previous chapter also determines the industrial classification to be used here, and hence compulsorily the period to be covered which is 2003-2008.

The factors that help to explain the causes of agglomeration is analysed by using an empirical specification that allows one to relate the *EG* index of industry agglomeration to industry characteristics and agglomeration forces put-forward by the theory. Having data over time for the same cross section units allows us to use panel data models. Baltagi (2001, p. 5-9) lists several benefits as well as limits of using panel data. However, the major power of using panel data arises as it allows one to control for unobserved cross section heterogeneity. Concerning this study, unobserved industry characteristics most probably differs across industries which validates conducting analysis in a panel data framework. A simple panel data model can be built in the following form within the context of this study:

$$\gamma_{it} = \alpha + \beta X_{it} + u_{it} \tag{5.1}$$

where i and t index industry and time, respectively, γ depicts the EG industry agglomeration index, X is a vector of industry characteristics that explain agglomeration and u is the idiosyncratic error term.

The panel data model introduced in equation (5.1) can be estimated by different estimation techniques [3]. This study follows one way error correction component models with fixed effect and random effect considerations. Baltagi (2001, p.12) emphasize that that fixed effect models are appropriate in cases where the study focuses on a given set of firms, countries or regions. This study instead focuses on a group of firms that belong to the same industry. Hence industries rather than firms constitute the cross section unit. Revisiting equation (5.1) the *fixed effect* model can be constructed as follows:

$$\gamma_{it} = \alpha + \beta X_{it} + \mu_i + v_{it} \tag{5.2}$$

As obvious as seen one way error component is decomposed as follows, $u_{it} = +v_{it}$ where μ_i is the unobserved heterogeneity. It is assumed to be a fixed parameter and is by construction correlated with explanatory variables. The remainder disturbances are assumed to be stochastic with v_{it} identically and independently distributed $IID(0, \sigma_v^2)$. Moreover X_{it} is assumed to be independent of the v_{it} for all i and t. Eliminating unobserved heterogeneity component basically forms the background of the fixed effects model estimation. Wooldridge (2010, p.267) asserts that there are several transformations to accomplish this purpose. Commonly, *fixed effects* transformation, or which can be labelled as *within* transformation is used. Fixed effects transformation yields a two step procedure where in the first stage equation (5.2) is averaged over $t = 1, 2, ..., T$ to obtain the cross section equation. Then in the second stage this cross section equation is subtracted from

[3]For an exhaustive presentation of panel data models, plese see Baltagi (2001), Part 5 in Cameron and Trivedi (2005) and Chapter 10 and 11 in Wooldridge (2010).

equation (5.2) Wooldridge (2010). Estimating the resultant equation by pooled ordinary least squares (OLS) seems appropriate as the unobserved effect of the model is removed, which was allowed to be correlated with explanatory variables by construction.

In *random effect* models unobserved heterogeneity is assumed to be random, which requires $\mu_i \sim IID(\sigma_\mu^2)$, $v_i \sim IID(\sigma_v^2)$ and the μ_i are independent of the v_{it}. Baltagi (2001) emphasized that if a predetermined number of observations are drawn from a population random effects model would be an appropriate specification. He gives the most common example as household surveys. If expected individual effects are uncorrelated with the control variables, then unobserved heterogeneity can be modelled as randomly distributed across cross section units. Random effect model contains the unobserved heterogeneity which is assumed to be uncorrelated with the explanatory variables, unlike fixed effects model. However, random effect model accounts for the implied serial correlation in the composite error component by using Generalized Least Squares (GLS) Wooldridge (2010, p.257). Variance-covariance matrix of the composite error term implies homoskedastic variance and serial correlation is also allowed for the disturbances for the same cross section units over time (Baltagi, 2001).

Given two error components model above, central question arises as the choice of the accurate model during the estimation procedure. The critical assumption made in error component models above is that error component should not be correlated with the explanatory variables. Disturbances contain individual heterogeneity component (μ_i) which are unobserved and may be correlated with the X_{it}. In this case, since $E(u_{it} \mid X_{it}) \neq 0$ GLS estimator obtained by using random effect model becomes biased and inconsistent. However, the within transformation removes μ_i and leaves the within estimator unbiased and consistent (Baltagi, 2001). Hausman (1978) proposes comparing two estimators, namely $\widehat{\beta}_{GLS}$ and $\widehat{\beta}_{within}$, where both of them are consistent under the null hypothesis of errors and explanatory variables are uncorrelated, $H_0 : E(u_{it} \mid X_{it}) = 0$. But, if H_0 is not true they will have different probability limits. As a matter of fact, $\widehat{\beta}_{within}$ is consistent irrespective of H_0 whether true or false , while $\widehat{\beta}_{GLS}$ is best linear unbiased estimator (BLUE), consistent and asymptotically efficient if H_0 is true, but inconsistent otherwise. Basically it tests whether the difference between the two estimates is statistically significantly different from zero or not. Yet, it is worth mentioning that Hausman test does not compare the appropriateness of the fixed or random effect models, rather it gives information regarding the consistency of the random effect model which is the more efficient model under the true null hypothesis. Keeping in mind this, considering the nature of the data and sampling procedure would help one better to decide on the right model to employ.

5.2.2 Control Variables

Table 5.1 lists the control variables used to proxy agglomeration theories. Marshall (1890) identifies three specific channels of externalities that may contribute to industrial agglomeration process. These channels are defined as input sharing, labour market pooling and knowledge spillovers.

Input sharing refers to linkages between intermediate and final good suppliers. Some inputs of production that can not be traded and procured easily by final good producers, will be produced and supplied by other providers within the same localities. These inputs will be produced by these providers at a lower cost due to specialization in production. This in turn will benefit both final good producers and also input providers. A good proxy for input sharing would be the use of input-output tables which gives information about the buy and sell relationships between sectors. Most of the studies, i.e. Kim, Barkley, and Henry (2000), Rosenthal and Strange (2001), Alecke, Alsleben, Scharr, and Untiedt (2006), (Lu and Tao, 2009), use share of manufactured inputs as measured by the ratio of cost of inputs purchased from manufacturing sector to total shipments. Unfortunately, input-output table for 4-digit industries is not available for Turkish manufacturing case regarding the period in question. The most recent input-output table for Turkish manufacturing industries belongs to year 2002 at 2-digit industry classification. For this reason, input sharing is proxied by the share of raw material purchases to total expenditures which is available in AISS dataset. Similarly, Barrios, Bertinelli, Strobl, and Teixeira (2009) use the same proxy to account for input sharing in their analysis of determinants of agglomeration in three European countries due to data unavailability.

An industry requiring workers with industry specific skills would better locating in a location where its supply is high because it increases the probability of finding appropriate workers. This is known as *labour market pooling* hypothesis and it is one of the Marshallian micro-foundations which is most difficult to proxy, as also expressed by Rosenthal and Strange (2001). The problem arises from the difficulty of identifying industry characteristics with regards to the specialization of the industry's labor force. Different studies hava made use of diverse set of proxies to account for labour pooling due to their data availability. For instance Alecke, Alsleben, Scharr, and Untiedt (2006), Rosenthal and Strange (2001) and Barrios, Bertinelli, Strobl, and Teixeira (2009) have used share of workers with higher degree education (B.A, Masters and PhD) and in addition the share of management workers in total employment (former two studies), while Kim, Barkley, and Henry (2000) explored the advantages of occupational data and grouped occupations according to their skill requirements as low, medium and high. Clearly occupational attributes identify industry characteristics better than educational attributes regarding workers' specialization, but still one should expect highly educated workers most likely to perform specialized tasks. So, if pooling is possible a positive relationship is expected between the agglomeration measure and labour pooling proxy.

Table 5.1: **Description of the variables used in the regressions**

Variable	Description	Type of spillovers
sh_purch_raw	$\frac{\text{Purchase of raw materials}}{\text{Total expenditures}}$	*input sharing*
labint	$\frac{\text{Industry payroll}}{\text{Industry value added}}$	*labor market pooling*
sh_RD_exp	$\frac{\text{Total R\&D expenditures}}{\text{Total expenditure}}$	*knowledge spillovers*
sh_RD_emp	$\frac{\text{Total R\&D employment}}{\text{Total employment}}$	*knowledge spillovers*
tra_cost1	$\frac{1}{UV} = \frac{\text{weight(IM+X)}}{\text{value(IM+X)}}$	*trasportation costs*
tra_cost2	$\frac{\text{End-of-year inventories}}{\text{Production value}}$	*trasportation costs*
avfirms	$\frac{\text{Persons employed}}{\text{Number of firms}}$	*internal IRTS*
sh_purch_energy	$\frac{\text{Purchases of energy products}}{\text{Total expenditures}}$	*natural advantages*

However, neither educational nor occupational attributes are available in AISS data. Alternatively, another proxy proposed by Kim, Barkley, and Henry (2000) is utilised instead. They construct labour intensity variable as measured by the ratio of industry payroll to industry value added, to control for the relative importance of labour as an input in the industry's production process. Yet, its relationship with the agglomeration measure is not so straightforward, it depends on the relative dominance of two opposing forces. If labour intensive industries mostly benefit from scale economies associated with labour pooling, then it will be positively related. On the other side, industries with relatively large demands for labour may be less willing to agglomerate due to higher wages associated with agglomerated locations, hence a negative relationship may be anticipated.

Knowledge spillovers or *informational spillovers* is the third and most widely discussed Marshallian agglomeration force by geographers, regional scientists and urban economists. Due to the public-good nature of knowledge, once it is created by one firm its usage by other firms does not reduce its content. Given that different firms own different information, the advantages brought by locating closer increase as the number of firms involved increase. So a positive relationship is expected between knowledge spillovers and agglomeration. Knowledge spillovers may be accounted for by employing various variables constructed by using data on patents, innovations, new products, R&D expenditures, etc. Since patent or innovation data is not available for the period in question, we use two variables to proxy knowledge spillovers, namely the share of R&D expenditures in total expenditures and the share of R&D employment in total employment of an industry.

As we have seen in second chapter *transportation costs* and *internal increasing returns* are keystones in NEG which predicts that industries with increasing returns to scale in production technology and lower transportation costs tend to be more concentrated. Although testing the empirical validity of NEG models is completely a different task as this study does not attempt to undertake, still it is worth testing whether these predictions find any support in Turkish data. Since transportation costs are not directly observable in the dataset, the ingenious proxies employed in a few empirical studies are used to account for. Grounding on the finding of Hummels (1999) that explicit costs such as tariffs and freight costs are the most important components in trade costs for the majority of traded goods, Alecke, Alsleben, Scharr, and Untiedt (2006) measure the average trade cost of an industry by the inverse of its unit value [4]. Reciprocal of the unit value is computed as $\frac{1}{UV} = \frac{\text{weight(IM+X)}}{\text{value(IM+X)}}$ and measures the amount of weight per Turkish lira (TL). In other words it measures how many tons of weight is carried by 1 TL. So an increase in this measure means that the amount of weight carried per TL has increased, implying a decline in transportation costs as 1 TL is able carry more weight than before. So, due to the nature of the variable constructed, a positive sign is expected while it points to a negative relationship between transportation costs and agglomeration consistent with the literature.

Also, another proxy is employed to account for *transportation costs* relying on Rosenthal and Strange (2001). They point out that there is close link between perishability of a good and the cost of transporting it. Industries producing highly perishable goods face higher transportation costs per unit distance because they have to transport their goods as soon as possible to their markets. So in order to avoid high transportation costs, they seek to locate closer to their markets. Therefore such industries tend to display less agglomeration. And an opposite result would arise for the industries producing non-perishable goods. Accordingly, they show a compelling support for using 'inventories' to proxy for

[4] Unit value is commonly used as a proxy for trade prices in empirical research in international economics and measures the monetary value of per unit weight traded. Unit value is computed as the ratio of value of total imported and exported goods to their weight, algebraically shown as $UV = \frac{\text{value(IM+X)}}{\text{weight(IM+X)}}$.

perishability. They display that industries producing highly perishable goods have very low levels of inventory to output ratios while the ones producing non-perishable goods have high levels. So they use inventory to output ratio as a proxy for transportation costs. Following Rosenthal and Strange (2001), we also employ end of year inventories to production value. An increase in inventory to output variable for an industry would signal lower transportation costs faced by that industry, hence more tendency for agglomeration. Thus, a positive sign is also expected in this variable.

NEG literature emphasizes the role of increasing returns to scale internal to the firm together with transportation costs in determining the location of economic activity. The empirical studies mentioned above make use of plant size or average plant size to account for internal returns to scale which is very central in NEG. However, there is point that needs to be clarified here. First of all, NEG literature builds on the existence of increasing returns to scale that are *internal* to the firm not to the industry. Using average plant size of an industry as a proxy for increasing returns to scale at the firm level implicitly requires one to assume that production technology for each firm in a specific industry is the same. Even if it may not be an unrealistic assumption, it is crucial to be stressed. However, such a discussion does not exist in these studies. Possibly it might have deemed so straightforward that is not worth discussing. Keeping this point in mind and assuming the similar production technology for firms under the same industry, then one can argue that agglomeration tends to be more significant in industries exhibiting greater scale economies and expect a positive relationship. Alternatively, Kim, Barkley, and Henry (2000) also mentions about the potential effects of industry average plant size on the propensity to agglomerate without referring to NEG and they state that these effects are less clear. Industries exhibiting relatively larger scale economies may agglomerate if large plant size contributes to the spin-offs of new firms, if input suppliers and specialized service providers are attracted to the region to serve them, or if large firms encourage the development of labour pools with industry specific skills. On the other hand, large plants may contribute to higher local input prices which leads firms to avoid locations with large plants resulting in spatial dispersion. Thus, the sign of the average firm size variable depends on which force prevails. Bearing in mind these discussions, industry average firms size is included as a proxy for increasing returns.

Finally, *natural advantages* in terms of resource endowments are traditionally considered to be important in determining agglomeration. Generally the impacts of resource endowments are controlled by using proxies such as energy, natural resource, water, mining product, agricultural product and water usage ratios derived from input-output tables. As mentioned before, unfortunately input-output tables are not available to obtain the flows among sectors. So, purchases of energy products are used as a generic variable instead, to account for natural advantages, as proposed by Barrios, Bertinelli, Strobl, and Teixeira (2009).

5.2.3 Regression Results

A natural first step is to pool the data and estimate the model using ordinary least squares. Regressing γ_{it} on explanatory variables included in vector X_{it} yields consistent estimates of β if the composite error u_{it} in the pooled model is uncorrelated with the explanatory variables. As Cameron and Trivedi (2009) states the error u_{it} is likely to be correlated over time for a given industry so they propose using cluster-robust standard errors that cluster on the cross-section units. Therefore, each specification is run both with robust (unclustered) and cluster robust standard errors.

Table 5.2 presents pooled OLS estimates. (a) of each cloumn presents estimation results run with robust standard errors while column (b) shows the ones estimated with cluster-robust standard errors which are shown below coefficient estimates in parenthesis. In model (1.a), *sh_RD_emp*, *avfsize* and *sh_purch_energy* are marginally significant at 10 % significance level and have positive signs as expected. *labint* variable has a negative sign and highly significant. This results conforms with Kim, Barkley, and Henry (2000)'s prediction about labour market pooling hypothesis. They assert that industries with relatively large demands for labour may be less willing to agglomerate due to higher wages associated with agglomerated locations, hence a negative relationship may be anticipated. Whenever cluster-robust standard errors are used (model 1.b) all of the marginally significant variables lose significance and labour pooling variable becomes marginally significant. This is due to the fact that cluster-robust standard errors are larger. As discussed in the previous section, it may be suspected that the variable which proxy natural advantages is a poor proxy since many of the industries purchase energy in order to conduct production. So most probably is fails to capture natural advantages unless this information is derived from input-output tables. Therefore, this variable is dropped and the model is estimated again. In model (2.a), variables proxying transportation costs and average firm size have positive and significant coefficients as expected and *labint* variable has a negative sign as in the previous model which is highly significant. The same model estimated with cluster-robust (2.b) standard errors displays that only *trade_cost1* and *labint* variables are marginally significant keeping their signs. Next two estimations drop insignificant coefficients obtained in the previous model respectively and estimate the model again as shown in columns (3.a) through (4.b). What arises from this picture is that transportation costs and labour pooling variables are robust to all specifications (except model (1) for transportation costs). However, input sharing hypothesis is not supported in none of the specifications.

Alternative variables to proxy transportation costs and knowledge spillovers are used in order to find out whether results are robust to this specifications as shown in Appendix I. Table AI.3 employs share of R&D expenditures instead of R&D employment. The results obtained are the same as in Table 5.2 where transportation costs and labour market pooling variables are marginally significant in model (4.b) with the same signs. Regarding transportation costs, another proxy, *tra_cost2* which is measured by the ratio of end-of-year inventories to production value, is used in Table AI.4 and Table AI.5 with two different knowledge spillovers variables employed in each. In both tables transportation costs and labour pooling variables are significant in models (1.a) through (4.b) while in cluster-robust estimations show lower significance for both of them.

Input sharing hypothesis is not supported in none of the specifications displaying insignificant coefficients in all specifications. Knowledge spillovers hypothesis is also very weakly supported in models (1.a), (3.a) and (4.a) in Table 5.2.

In the next step, panel data analysis is performed which takes into account unobserved industry characteristics. Both fixed and random effect models are employed that are differentiated due to their assumptions made about unobserved heterogeneity as discussed in the previous chapter. First, fixed effect estimation results are shown in Table 5.3. Each variable is added step by step through models (1) to (6), then variables with insignificant coefficients are dropped respectively. Input sharing variable *sh_purch_raw* has a positive sign as expected and is marginally significant in models (2) to (4). But with the inclusion of variable *labint* it loses significance and *labint* becomes highly significant then. Transportation costs variable is highly significant in all models with a positive sign. This shows that this variable is quite robust to all specifications. Surprisingly, the variable that proxy knowledge spillovers has a negative sign, contrarily to theory, at a 5% significance level. It is not so straightforward to interpret this result as it suggests that knowledge spillovers generate a dispersion effect rather than concentration. Average firm size which is included in order to capture internal increasing returns, has no significant effect on the agglomeration of industries.

Table 5.2: **Pooled OLS model**

	(1.a)	(1.b)	(2.a)	(2.b)	(3.a)	(3.b)	(4.a)	(4.b)
sh_purch_raw	0.005	0.005	-0.050	-0.050	-0.049	-0.049		
	(0.058)	(0.077)	(0.062)	(0.098)	(0.062)	(0.098)		
sh_RD_emp	0.371*	0.371	0.290	0.290	0.342*	0.342	0.370**	0.370
	(0.203)	(0.333)	(0.195)	(0.340)	(0.191)	(0.326)	(0.188)	(0.325)
tra_cost1	0.011	0.011	0.013**	0.013*	0.013**	0.013*	0.013**	0.013*
	(0.007)	(0.009)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.007)
avfsize	0.000*	0.000	0.000**	0.000				
	(0.000)	(0.000)	(0.000)	(0.000)				
labint	-0.057***	-0.057*	-0.056***	-0.056*	-0.055***	-0.055*	-0.051***	-0.051*
	(0.020)	(0.033)	(0.019)	(0.031)	(0.019)	(0.031)	(0.016)	(0.026)
sh_purch_energy	0.784*	0.784						
	(0.407)	(0.710)						
Constant	0.091**	0.091	0.144***	0.144*	0.147***	0.147*	0.112***	0.112***
	(0.045)	(0.057)	(0.049)	(0.078)	(0.049)	(0.078)	(0.010)	(0.016)
R^2	0.040	0.040	0.030	0.030	0.028	0.028	0.027	0.027
$Adj.R^2$	.0359721	.0359721	.0261205	.0261205	.0249799	.0249799	.0245872	.0245872
N	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000
F	4.410	1.535	5.142	2.046	5.197	2.004	6.877	2.649

Standard errors in parentheses

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

(a)-robust s.e. (unclustered), (b)-cluster-robust s.e.

Labour pooling variable has a negative effect on agglomeration which is highly significant in all the models where appeared. It can be interpreted in the same way as in Table 5.2. It may be the case that centrifugal forces are more at work than centripetal forces in the form of higher wages in agglomerated regions that push firms apart to locate there due to cost disadvantage.

Table 5.4 displays the fixed effect estimations with robust standard errors. In these estimations, *sh_RD_exp* that proxy knowledge spillovers and *labint* that proxy labour pooling continue to have a negative and statistically significant effects on industrial agglomeration, while the result for transportation costs, *tra_cost1*, becomes much weaker than before but still has a positive effect on agglomeration.

Alternative variables for transportation costs and knowledge spillovers are also employed in fixed effect models which are not reported here since they yield similar results, but available upon request. In the parsimonious model (9) estimated with *tra_cost1* and *sh_RD_emp* with classical standard errors, transportation costs and labour pooling variables have the same signs and magnitudes as in Table 5.3 while the coefficient of the knowledge spillover variable displays an increase with a negative sign again. In the model estimated with robust standard errors, knowledge spillovers loses significance. In other two models where *tra_cost2* is employed with each of the knowledge spillover variables, transportation costs and knowledge spillovers variables lose significance and only labour pooling variable remains highly significant.

Table 5.5 presents random effects estimations. There is suggestive evidence for the importance of transportation costs. In the random effects estimations, transportation costs continue to have positive and highly statistically significant (1 % significance level) impacts on industrial agglomeration. But the evidence for Marshallian micro-foundations is either weak or contrary to the standard expectation. There is weak evidence for input sharing as shown in models (2) to (4) it is marginally significant and becomes insignificant with the inclusion of *labint*. On the other hand, empirical results obtained for labour market pooling and knowledge spillovers do not conform with the classical expected effects of these channels. These latter two channels of externalities are expected to have positive impact on industrial agglomeration while we observe a negative sign for the coefficients of related variables. In addition there is no evidence for the impact of internal increasing returns and natural advantages on the agglomeration of industries.

Table 5.3: **Fixed effects model**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sh_purch_raw	0.062	0.081**	0.074*	0.074*	0.046	0.046	0.046	0.046	
	(0.040)	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.040)	(0.040)	
tra_cost1		0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
sh_RD_exp			-0.311**	-0.311**	-0.320**	-0.317**	-0.317**	-0.320**	-0.338**
			(0.158)	(0.158)	(0.158)	(0.158)	(0.158)	(0.158)	(0.157)
avfsize				-0.000	-0.000	-0.000			
				(0.000)	(0.000)	(0.000)			
labint					-0.041***	-0.041***	-0.041***	-0.041***	-0.044***
					(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
sh_purch_energy						-0.069	-0.070		
						(0.278)	(0.278)		
Constant	0.067**	0.044*	0.051*	0.052*	0.086***	0.087***	0.087***	0.086***	0.118***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.029)	(0.030)	(0.029)	(0.028)	(0.006)
R^2	-0.198	-0.181	-0.178	-0.179	-0.168	-0.169	-0.168	-0.167	-0.168
N	1332	1314	1314	1314	1314	1314	1314	1314	1314
F	2.448	10.285	8.162	6.130	7.163	5.975	7.176	8.962	11.499

Standard errors in parentheses

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

Table 5.4: **Fixed effects model with robust S.E.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sh_purch_raw	0.062	0.081	0.074	0.074	0.046	0.046	0.046	0.046	
	(0.064)	(0.055)	(0.052)	(0.052)	(0.046)	(0.046)	(0.046)	(0.046)	
tra_cost1		0.009	0.009	0.009	0.009*	0.009*	0.009*	0.009*	0.009*
		(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
sh_RD_exp			-0.311***	-0.311***	-0.320***	-0.317***	-0.317***	-0.320***	-0.338***
			(0.114)	(0.114)	(0.088)	(0.093)	(0.093)	(0.088)	(0.089)
avfsize				-0.000	-0.000	-0.000			
				(0.000)	(0.000)	(0.000)			
labint					-0.041**	-0.041**	-0.041**	-0.041**	-0.044***
					(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
sh_purch_energy						-0.069	-0.070		
						(0.390)	(0.392)		
Constant	0.067	0.044	0.051	0.052	0.086***	0.087***	0.087***	0.086***	0.118***
	(0.043)	(0.036)	(0.034)	(0.035)	(0.031)	(0.032)	(0.031)	(0.030)	(0.007)
R^2	0.001	0.017	0.020	0.019	0.028	0.027	0.028	0.029	0.028
N	1332	1314	1314	1314	1314	1314	1314	1314	1314
F	0.944	3.278	4.879	3.682	4.341	3.845	4.623	5.431	7.439

Standard errors in parentheses

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

And finally random effects estimations with robust standard errors are carried out and displayed in Table 5.6. The coefficients of the variables and their signs do not alter, but their significance levels change due to the robust standard errors. For instance *tra_cost1* became marginally significant, while the significance level of *sh_RD_exp* increased and *labint* did not change. Alternating the knowledge spillover variable and use *sh_RD_emp* instead of *sh_RD_exp* results remain the same to a very large extent, except only the magnitude of the alternative variable changes. Similarly, second proxy is used also for transportation costs. And the models (1) through (9) are estimated again with this second proxy for each knowledge spillover variable. When *tra_cost2* and *sh_RD_exp* take place together in these models, it's observed that transportation costs, knowledge spillovers and labour pooling variables are highly significant with the same signs mentioned above under default standard errors. Moreover, the magnitude of the coefficient of *tra_cost2* is observed to be far larger than *tra_cost1*, more than ten times. However, transportation costs lose significance in models that are estimated with robust standard errors. We may infer that default standard errors are underestimated for *tra_cost2* variable such that with the inclusion of robust standard errors it loses all significance. And finally, the results for these models with default standard errors, where *tra_cost2* and *sh_RD_emp* take place together, are very similar to the previous estimation results in terms of significance, sign and magnitude . However, when they are estimated with robust standard errors, the only variable which remains significant is the *labint*, labour pooling variable.

Table 5.7 shows a comparison between random effects and fixed effects estimations. As mentioned in the previous section, both of the estimators are consistent under the null hypothesis of errors and explanatory variables are uncorrelated. If individual effects are random but not fixed, both of the estimators will be consistent but random effects estimator will be more efficient since it's standard errors will be smaller compared to fixed effects estimator which only considers within variation and the remaining variation will be reflected in standard errors by high values. But, if the null hypothesis is not true they will have different probability limits and random effects estimator will be inconsistent. However, it is important to underline again, that Hausman test does not compare the appropriateness of the fixed or random effect models, rather it gives information regarding the consistency of the random effect model which is the more efficient model under the true null hypothesis.

Based on this observation we fail to reject the null hypothesis that difference in coefficients is not systematic. In other words we accept the null hypothesis that individual effects are random, hence random effects estimator is more efficient.

Table 5.5: **Random effects model**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sh_purch_raw	0.037	0.067*	0.061*	0.061*	0.035	0.039	0.039	0.035	
	(0.037)	(0.036)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	
tra_cost1		0.010***	0.010***	0.010***	0.010***	0.010***	0.010***	0.010***	0.010***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
sh_RD_exp			-0.250	-0.251*	-0.257*	-0.265*	-0.263*	-0.255*	-0.269*
			(0.152)	(0.153)	(0.152)	(0.152)	(0.152)	(0.152)	(0.151)
avfsize				0.000	0.000	0.000			
				(0.000)	(0.000)	(0.000)			
labint					-0.041***	-0.042***	-0.042***	-0.041***	-0.044***
					(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
sh_purch_energy						0.244	0.249		
						(0.241)	(0.241)		
Constant	0.084***	0.056**	0.061**	0.060**	0.093***	0.086***	0.087***	0.095***	0.120***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.029)	(0.030)	(0.030)	(0.029)	(0.012)
$R^2_{overall}$	0.003	0.012	0.09	0.010	0.017	0.023	0.022	0.015	0.018
N	1332	1314	1314	1314	1314	1314	1314	1314	1314
$Wald\_chi2$	0.97	24.07	26.80	26.85	38.78	39.72	39.59	38.61	37.70

Standard errors in parentheses

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

Table 5.6: **Random effects model with robust S.E.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sh_purch_raw	0.037	0.067	0.061	0.061	0.035	0.039	0.039	0.035	
	(0.054)	(0.047)	(0.045)	(0.045)	(0.039)	(0.039)	(0.039)	(0.039)	
tra_cost1		0.010*	0.010*	0.010*	0.010*	0.010*	0.010*	0.010*	0.010*
		(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
sh_RD_exp			-0.250*	-0.251*	-0.257**	-0.265*	-0.263*	-0.255*	-0.269**
			(0.137)	(0.135)	(0.130)	(0.137)	(0.139)	(0.132)	(0.134)
avfsize				0.000	0.000	0.000			
				(0.000)	(0.000)	(0.000)			
labint					-0.041**	-0.042**	-0.042**	-0.041**	-0.044***
					(0.017)	(0.018)	(0.018)	(0.017)	(0.016)
sh_purch_energy						0.244	0.249		
						(0.435)	(0.434)		
Constant	0.084**	0.056*	0.061**	0.060**	0.093***	0.086***	0.087***	0.095***	0.120***
	(0.038)	(0.032)	(0.031)	(0.030)	(0.029)	(0.028)	(0.029)	(0.029)	(0.013)
$R^2_{overall}$	0.003	0.012	0.009	0.009	0.017	0.023	0.022	0.015	0.018
N	1332	1314	1314	1314	1314	1314	1314	1314	1314
$Wald_c hi2$	0.47	6.23	8.50	9.26	13.76	12.98	12.44	13.25	13.12

Standard errors in parentheses

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

Table 5.7: **Hausman Test: FE or RE**

	RE	FE	Difference	(S.E)
sh_purch_raw	0.039	0.046	0.007	0.0134
	(0.037)	(0.040)		
sh_RD_exp	-0.265*	-0.317**	-0.051	0.0433
	(0.152)	(0.158)		
tra_cost1	0.010***	0.009***	-0.001	0.0005
	(0.002)	(0.002)		
avfsize	0.000	-0.000	0.0000	0.0001
	(0.000)	(0.000)		
labint2	-0.042***	-0.041***	0.001	0.0025
	(0.012)	(0.012)		
sh_purch_energy	0.244	-0.069	-0.314	0.1382
	(0.241)	(0.278)		

$chi2(6) = 9.51$
$Prob > chi2 = 0.1468$
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Estimation results for the traditional determinants of industrial agglomeration are mostly inconsistent with the findings in the existing literature, especially Marshallian forces. In both the pooled and panel estimations, there is almost no evidence for input sharing. This is largely due to the proxy employed, which is defined as the share of raw material purchased within all purchases. An ideal proxy to would be derived from input-output tables which display interactions among sectors in the form of buy and sell linkages. In Turkey, the lately available input-output table belongs to 2002 at two-digit industrial classification level, which is too broad for this analysis. It might be argued that, input-output structure among industries does not change rapidly, so one can use the available data for all years. However, even if data would have been available at the four-digit level for the year 2002, it would be treated as a time invariant variable which would restrain one to employ fixed effects models which accounts for individual heterogeneity.

The second Marshallian mechanism labour market pooling is found to have a statistically significant negative effect on industrial agglomeration. This is just the opposite of the standard expectation where a positive effect is anticipated. Kim, Barkley, and Henry (2000) have mentioned this as a potential outcome that may arise in the case of increasing wages due to increased demand for labour in agglomerated areas, hence which means increasing costs for firms that disincentive them to agglomerate. To put differently, centrifugal forces may well outweigh centripetal forces resulting in dispersion rather than explanation. This explanation might be valid for the case of Turkey, but due to its ad hoc nature it sounds a little loose. In order to have a solid explanation about this result, it might be supported with the wage data related to specific skill occupations (which labour market pooling hypothesis centrally mentions) in most agglomerated regions. On the other hand, the possibility of poor proxy may well be valid also for labour market pooling variable. An ideal proxy would be obtained by gathering information about skill levels of workers from an occupational data, or more implicitly by considering their education levels via relating it to skills. However, the dataset employed in this study does not convey such an information.

Knowledge spillovers hypothesis, the third Marshallian mechanism, empirically yields the most unlikely result with a negative effect on agglomeration. The coefficients of both share of R&D expenditure and share of R&D employment are negative and statistically significant at different levels of significance in both pooled and panel models. Regarding empirical literature, its effect on agglomeration is either found to be insignificant or positive. So it is not straightforward to provide a plausible explanation for this result. The easiest way would be to accuse data of not being capable of explaining the knowledge spillover phenomenon, which is probable in any case. But beyond that, it is important to recall tacit nature of knowledge discussed before, which attenuate with distance hence likely to contribute to agglomeration at smaller geographic scales. This fact is empirically supported by many studies suggesting that spillover effects are much localized (Audretsch and Feldman (1996), Rosenthal and Strange (2004), Audretsch and Feldman (2004)). Therefore it might be due to the fact that geographic units are so broadly defined (NUTS-2) that enables index to capture knowledge spillovers which are possibly at work at finer units of geography. Since data is available only at NUTS-2 level, it can not be tested empirically but provided as a possible explanation.

One of the main postulates of NEG, the one regarding the positive effect of declining transportation costs on agglomeration, is the most empirically supported hypothesis by the data. Transportation costs proxied by two alternative variables, as described in Table 5.1, both have statistically significant negative effects on industrial agglomeration and they are robust to inclusion of other control variables and alternative specifications. Note that, the sign of the coefficient of the related variable is positive but it is interpreted inversely due to the construction of the variable.

Finally, Increasing returns to scale and natural advantages hypotheses are not supported. Using average plant size of an industry as a proxy for increasing returns to scale at the firm level may not be able to capture this fact, because it implicitly requires one to assume that production technology for each firm in a specific industry is the same, which may not be the really case. Regarding natural advantages, it is proxied by share of purchases of energy products which is likely to be a poor proxy. Again it would be ideal to use input-output tables to see whether regional variations in resource endowments do matter in determining the patterns of industrial agglomeration.

In summary, we find strong evidence supporting the role of transportation costs in facilitating industrial agglomeration in Turkey; however, Marshallian externalities are not supported. Nevertheless, it is important to recognize that, finding no evidence for Marshallian externalities does not necessarily mean that these mechanisms do not work in Turkey, but it might be argued that the available data fails to detect these mechanisms.

6 CONCLUSION

Location of economic activity has been examined within numerous disciplines. Originating from the early discussions of von Thünen the field, which may be called broadly as spatial economics, has attracted attention from many scholars related to different disciplines. 1950's witnessed the rise of regional science which is followed by urban economics in 1960's. However, it had to await for almost thirty years for location to be considered as an important integral part of economics. With the rise of new economic geography (NEG) in early 1990's, the attention of scientists has shifted on the role of space and geography in understanding the distribution of economic activity.

Beyond theoretical discussions, empirical studies display evidence that most of the economic activities tend to cluster in certain locations. In this respect, quantifying agglomeration has gained a significant interest from researchers and it has been a center of another line of research.

This dissertation examines the geographic distribution of Turkish manufacturing industries for the post-2000 period. The analysis is mainly based on the index developed by Ellison and Glaeser (1997), which has been designed to allow for comparisons across industries and countries. The intention has been to provide a rigorous descriptive analysis of the distribution of manufacturing activity in Turkey.

Initial findings of the dissertation find supporting evidence that manufacturing industries in Turkey are highly agglomerated. Regarding years 2003 and 2008, 57.9 % and 55.9 % of the industries are highly agglomerated displaying a value $\gamma \geqslant 0.05$, 14.5 % and 17.5 % of them are moderately concentrated with $0.02 < \gamma \leqslant 0.05$, respectively. Ignoring some of the industries which display high level of agglomeration due to high industrial concentration, average level of agglomeration, measured by the mean of γ calculated at the four-digit industry level by two-digit industries, is highest in textile (17), other non-metallic mineral products (26) and food products and beverages (15) industries in 2003. These industries include a great number of four-digit sub-industries and on average display a high degree of agglomeration with a γ of 0.17, 0.15 and 0.13, respectively.

Likewise, textile (17), basic metals (27) and food products and beverages (15) industries are observed to be highly agglomerated with average values of 0.26, 0.14 and 0.13, respectively in 2008. It is obvious that agglomeration level in textile industries on average has increased through the time period. In terms of more specific four-digit industries, processing of tea and coffee (1586), silk-type weaving (1724), manufacture of plaster (2653), manufacture of glass fibres (2614), manufacture of weapons and ammunition (2960) and dressing and dyeing of fur (1830) industries were among the most localized industries in 2003. Most agglomerated industries in 2008 include processing of tea and coffee (1586), manufacture of carpets and rugs (1751), other textile weaving (1725), silk-type weaving (1724) and manufacture of ceramic household (2621). Five out of 15 most localized industries in 2008 has also been ranked within the top 15 list in 2003 (1717, 1586, 2741, 1753 and 1724). It is also worth noting that 11 out of 20 most localized industries in 2008 have also been listed within the top 20 regarding year 2003. In other words, it may said that 55 % of the industries have remained in the top list during the period. Moreover, five textile (17) related industries in this group have all enhanced their rankings relative to 2003. Also on average textile (17) industries have increased their index values.

A striking feature arises from this picture. Most localized industries in both years encompass low-technology industries. Within this technology group textile and traditional industries are observed to be dominant. Based on the OECD technology classification scheme, on average low and medium-low technology sectors show a higher degree of agglomeration in Turkey which display average agglomeration index of 0.114 and 0.118 in 2003 and 0.130 and 0.094 in 2008, respectively. Specifically, above the 0.05 threshold, the share of low-technology sectors are much higher than other technology levels for both years. Still medium-high and high technology sectors on average display agglomeration above the 0.05 threshold, but they are far below the other two lower technology groups.

The degree of overall agglomeration is observed to be stable throughout the period. Nonetheless, a value of around 0.1 still represents a high level of localization on average. The stability in agglomeration levels in most of the Turkish manufacturing industries is a common pattern observed in other countries. However some industries experienced remarkable changes in their agglomeration levels.

When agglomeration patterns in Turkey are compared to developed countries; Turkey arises as having high levels of agglomeration with an average EG index of 0.112. This fact may hinge on to the differences in terms of transportation costs, labour market conditions, and more broadly any other factors influencing the location of plants across countries considered. Still, it may well be a possible case for Turkey as a developing country whose dynamics behind agglomeration motives are likely to be different than developed countries.

The stylized fact arising from the research on agglomeration that traditional and low-tech industries tend to show higher degrees of agglomeration relative to others, also hold for the case of Turkey. 80 and 85 per cent of the 20 most agglomerated industries fall within the low and medium-low technology sectors in 2003 and 2008, respectively. Consistent with the previous findings textile and traditional sectors dominate the group. However, there is limited evidence on the agglomeration of high-tech industries, which has been subject to policy considerations in developed countries.

Uncovering agglomeration patterns in Turkish manufacturing industries, the study attempted to analyse the factors behind industrial agglomeration. As long as the data set employed allows, the sources of agglomeration are analysed with reference to Marshallian externalities and postulates of NEG. There is suggestive evidence that transportation costs play an important role in determining industrial agglomeration. However, Marshallian forces of agglomeration do not find any support in the Turkish data. While input sharing hypothesis does not find any evidence in Turkish data, labour market pooling and knowledge spillovers hypotheses have opposite impacts on industrial concentration contrary to a priori expectations. While it is possible to provide a plausible explanation for the negative effect of labour market pooling on agglomeration, it is not so straightforward for knowledge spillovers. The negative relation between labour market pooling hypothesis and industrial agglomeration may be explained by centrifugal forces that overweigh centripetal forces in the form of increasing wages in the agglomerated locations which deter firms to agglomerate further in those locations. However, regarding knowledge spillovers, the negative effect is not easy to explain. The quality of the variables that proxy knowledge spillovers and the relevance of the geographical units to capture this might be provided as loose explanations.

Over all the conclusions, the limitations of the study cover a number of issues. First of all, geographic scale of the available data is quite broad and only available at NUTS-2 level which does not allow one to consider geographic scope of agglomeration at different scales. This is important in the sense that different agglomeration mechanisms work at different scales of spatial units. Secondly, the inferences made about determinants of agglomeration should be taken cautiously. Better proxies may be employed for Marshallian externalities. Regarding input sharing, input-output tables may be explored if a recent version becomes available covering the period in question in our analysis. In addition to that, household labour force surveys may be explored to have better proxies for labour market pooling which would yield information about occupation and education level. And thirdly, the analysis is conducted without taking into consideration the firm size. However, both theoretical and empirical studies show that agglomeration patterns differ between different firm sizes. The latter two limitations of the study also point the way for further research.

Central contributions of the dissertation may be captured in a few points. First of all, geographic concentration of economic activity has first been explored by following the framework proposed by Ellison and Glaeser (1997) for the case of Turkey. Second contribution derives from the first one, such that the framework employed requires one to explore firm level micro data to conduct the analysis. In this respect, micro data is firstly explored in this line of research. And finally, its novelty lies behind the period it considers. Previous studies examining the geographic concentration of economic activities in Turkey have largely employed highly aggregated data at provincial or regional level and covered a certain period of time, namely 1980-2000, due to the inconsistency of regional data from then on. Therefore, we have very limited knowledge about geographic concentration of industries for the post-2000 period. So, examining the post-2000 period, which has not been examined within this line of research before, will be shedding a new light on the agglomeration phenomenon in Turkey.

BIBLIOGRAPHY

Akgüngör, S. (2003): "Exploring Regional Specialization in Turkey's Manufacturing Industry," in *Regional Studies Association International Conference*.

——— (2006): "Geographic concentrations in Turkey's manufacturing industry: identifying regional highpoint clusters," *European Planning Studies*, 14(2), 169–197.

Albert, J. M., M. R. Casanova, and V. Orts (2012): "Spatial location patterns of Spanish manufacturing firms*," *Papers in Regional Science*, 91(1), 107–136.

Alecke, B., C. Alsleben, F. Scharr, and G. Untiedt (2006): "Are there really high-tech clusters? The geographic concentration of German manufacturing industries and its determinants," *The Annals of Regional Science*, 40(1), 19–42.

Alonso-Villar, O., J.-M. Chamorro-Rivas, and X. González-Cerdeira (2004): "Agglomeration economies in manufacturing industries: the case of Spain," *Applied Economics*, 36(18), 2103–2116.

Arbia, G. (2001a): "Modelling the geography of economic activities on a continuous space," *Papers in Regional Science*, 80(4), 411–424.

——— (2001b): "The role of spatial effects in the empirical analysis of regional concentration," *Journal of Geographical systems*, 3(3), 271–281.

Arbia, G., M. Copetti, and P. Diggle (2009): "Modelling Individual Behaviour of Firms in the Study of Spatial Concentration," in *Growth and Innovation of Competitive Regions*, ed. by U. Fratesi, and L. Senn, Advances in Spatial Science, pp. 297–327. Springer Berlin Heidelberg.

Arbia, G., G. Espa, D. Giuliani, and A. Mazzitelli (2009): "Clusters of firms in space and time," *Department of Economics, University of Trento, Italia Department of Economics Working Papers*.

Arbia, G., G. Espa, and D. Quah (2008): "A class of spatial econometric methods in the empirical analysis of clusters of firms in the space," *Empirical Economics*, 34(1), 81–103.

Audretsch, D. B., and M. P. Feldman (1996): "R&D spillovers and the geography of innovation and production," *The American Economic Review*, 86, 630–640.

——— (2004): "Knowledge spillovers and the geography of innovation," in *Handbook of Regional and Urban Economics*, ed. by V. J.Henderson, and J. F. Thisse, vol. 4, chap. 61, pp. 2713–2739. Elsevier, Amsterdam: North-Holland.

Balassa, B. (1966): "Tariff reductions and trade in manufacturers among the industrial countries," *The American Economic Review*, 56(3), 466–473.

Baltagi, B. (2001): *Econometric analysis of panel data*. John Wiley & Sons, England, 2 edn.

Barkley, D. L., Y. Kim, and M. S. Henry (2001): "Do manufacturing plants cluster across rural areas? Evidence from a probabilistic modeling approach," Discussion paper, Clemson University, Regional Economic Development Research Laboratory (REDRL).

Barlet, M., A. Briant, and L. Crusson (2013): "Location Patterns of Service Industries in France: A Distance-Based Approach," *Regional Science and Urban Economics*, 43, 338–351.

Barrios, S., L. Bertinelli, E. Strobl, and A.-C. Teixeira (2005): "The dynamics of agglomeration: evidence from Ireland and Portugal," *Journal of Urban Economics*, 57(1), 170 – 188.

Barrios, S., L. Bertinelli, E. Strobl, and A. C. Teixeira (2009): "Spatial Distribution of Manufacturing Activity and its Determinants: A Comparison of Three Small European Countries," *Regional Studies*, 43(5), 721–738.

Beaudry, C., and A. Schiffauerova (2009): "Who's right, Marshall or Jacobs? The localization versus urbanization debate," *Research Policy*, 38(2), 318–337.

Behrens, K., and T. Bougna (2015): "An anatomy of the geographical concentration of Canadian manufacturing industries," *Regional Science and Urban Economics*, 51, 47–69.

Behrens, K., and J.-F. Thisse (2007): "Regional economics: A new economic geography perspective," *Regional Science and Urban Economics*, 37(4), 457–465.

Bertinelli, L., and J. Decrop (2005): "Geographical agglomeration: Ellison and Glaeser's index applied to the case of Belgian manufacturing industry," *Regional Studies*, 39(5), 567–583.

Beyene, J., and R. Moineddin (2005): "Methods for confidence interval estimation of a ratio parameter with application to location quotients," *BMC Medical Research Methodology*, 5(1), 32.

Billings, S. B., and E. B. Johnson (2012): "The location quotient as an estimator of industrial concentration," *Regional Science and Urban Economics*, 42(4), 642–647.

——— (2014): "Measuring Agglomeration: Which estimator should we use?," .

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2010): "Adding geography to the new economic geography: bridging the gap between theory and empirics," *Journal of Economic Geography*, 10(6), 793–823.

Brakman, S., H. Garretsen, J. Gorter, A. van der Horst, and M. Schramm (2005): *New economic geography, empirics, and regional policy*, no. 58 in ISBN 9789058332813. CPB Netherlands Bureau for Economic Policy Analysis.

Briant, A., P.-P. Combes, and M. Lafourcade (2010): "Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?," *Journal of Urban Economics*, 67(3), 287 – 302.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics: Methods and Applications*. Cambridge University press, New York, 1 edn.

Cameron, A. C., and P. K. Trivedi (2009): *Microeconometrics Using Stata*. Stata Press, College Station, Texas.

Cassey, A. J., and B. O. Smith (2014): "Simulating confidence for the Ellison-Glaeser index," *Journal of Urban Economics*, 81(0), 85 – 103.

Chamberlin, E. (1933): *The Theory of Monopolistic Competition,*. Cambridge, MA: Harvard University Press.

Christaller, W. (1933): *Central Places in Southern Germany*. London, Prentice Hall.

Combes, P.-P. (2011): "The empirics of economic geography: how to draw policy implications?," *Review of World Economics*, 147(3), 567–592.

Combes, P.-P., and H. G. Overman (2004): "Chapter 64: The spatial distribution of economic activities in the European Union," in *Cities and Geography*, ed. by J. V. Henderson, and J.-F. Thisse, vol. 4 of *Handbook of Regional and Urban Economics*, pp. 2845 – 2909. Elsevier.

Committee, N. P. (2008): "Paul Krugman: Trade and Geography - Economies of Scale, Differentiated Products and Transport Costs," Nobel Prize in Economics documents 2008-2, Nobel Prize Committee.

Cutrini, E. (2006): "The Balassa index meets the dissimilarity Theil index: A decomposition methodology for location studies," *Universita Politecnica Delle Marche, Working Paper*, (274).

De Benedictis, L., and M. Tamberi (2004): "Overall specialization empirics: techniques and applications," *Open Economies Review*, 15(4), 323–346.

Devereux, M. P., R. Griffith, and H. Simpson (2004): "The geographic distribution of production activity in the UK," *Regional Science and Urban Economics*, 34(5), 533–564.

Dixit, A. K., and J. E. Stiglitz (1977): "Monopolistic competition and optimum product diversity," *The American Economic Review*, 67(3), 297–308.

Doğruel, F., A. S. Doğruel, and B. C. Karahasan (2011): "Market Potential and Regional Disparities in Turkey," ERSA conference papers ersa11p1601, European Regional Science Association.

Doğruel, S., and F. Doğruel (2008): "Türkiye Sanayiine Sektörel Bakış (Turkish Industry: A Sectoral Overview)," Discussion paper.

Dumais, G., G. Ellison, and E. L. Glaeser (2002): "Geographic Concentration As A Dynamic Process," *The Review of Economics and Statistics*, 84(2), 193–204.

Duranton, G., and H. Overman (2005): "Testing for localization using micro-geographic data," *The Review of Economic Studies*, 72(4), 1077–1106.

Duranton, G., and H. G. Overman (2008): "Exploring the detailed location patterns of UK manufacturing industries using microgeographic data*," *Journal of Regional Science*, 48(1), 213–243.

Ellison, G., and E. Glaeser (1999): "The geographic concentration of industry: Does natural advantage explain agglomeration?," *American Economic Review*, 89(2), 311–316.

Ellison, G., E. Glaeser, and W. Kerr (2010): "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economic Review*, 100(3), 1195–1213.

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

Feng, J., and M. Ji (2011): "Integrating location quotient, local Moran's I and geographic linkage for spatial patterning of industries in Shanghai, China," in *Geoinformatics, 2011 19th International Conference on*, pp. 1–6. IEEE.

Feser, E. J. (2000): "On the Ellison–Glaeser geographic concentration index," .

Fujita, M. (2010): "The evolution of spatial economics: from Thünen to the new economic geography," *Japanese Economic Review*, 61(1), 1–32.

Fujita, M., and P. Krugman (2004): "The new economic geography: Past, present and the future," *Papers in regional science*, 83(1), 139–164.

Fujita, M., P. R. Krugman, and A. Venables (2001): *The Spatial Economy: Cities, regions, and international trade*. MIT press.

Fujita, M., and J.-F. Thisse (1996): "Economics of agglomeration," *Journal of the Japanese and international economies*, 10(4), 339–378.

Fujita, M., and J. F. Thisse (2002): *Economics of agglomeration: Cities, industrial location, and regional growth*. Cambridge University Press, Cambridge.

Fujita, M., and J. F. Thisse (2009): "New economic geography: An appraisal on the occasion of Paul Krugman's 2008 Nobel Prize in Economic Sciences," *Regional Science and Urban Economics*, 39(2), 109–119.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992): "Growth in Cities," *Journal of Political Economy*, 100(6), 1126–52.

Gokan, T. (2010): *Spatial Statistics and Industrial Location in CLMV: An Interim Report*chap. Chapter 2: On the usage of the measurements of geographical concentration and specialization with areal data. Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO).

Guillain, R., and J. Le Gallo (2007): "Agglomeration and dispersion of economic activities in Paris and its surroundings: An exploratory spatial data analysis," *LEG-Document de travail-Economie*, 1.

Guimarães, P., O. Figueiredo, and D. Woodward (2007): "MEASURING THE LOCALIZATION OF ECONOMIC ACTIVITY: A PARAMETRIC APPROACH*," *Journal of Regional Science*, 47(4), 753–774.

——— (2009): "Dartboard tests for the location quotient," *Regional Science and Urban Economics*, 39(3), 360–364.

——— (2011): "Accounting for Neighboring Effects in Measures of Spatial Concentration," *Journal of Regional Science*, 51(4), 678–693.

Haaland, J. I., H. J. Kind, and K.-H. Ulltveit-Moe (1999): "What Determines the Economic Geography of Europe?," CEPR Discussion Papers 2072, C.E.P.R. Discussion Papers.

Hallet, M. (2000): "Regional specialisation and concentration in the EU," Economic Papers 141, European Commission, Directorate-General for Economic and Financial Affair.

Hausman, J. (1978): "Specification Test in Econometrics," *Econometrica*, 46, 1251–1271.

Head, K., and T. Mayer (2004): "The empirics of agglomeration and trade," in *Handbook of Regional and Urban Economics - Volume IV: Cities and Geography*, ed. by J. V. Henderson, and J. Thisse, vol. 4, pp. 2609–2669. Elsevier.

Holmes, T. J., and J. J. Stevens (2004a): "Geographic concentration and establishment size: analysis in an alternative economic geography model," *Journal of Economic Geography*, 4(3), 227–250.

——— (2004b): "Spatial distribution of economic activities in North America," *Handbook of regional and urban economics*, 4, 2797–2843.

Hoover, E. M. (1937): "Spatial price discrimination," *The Review of Economic Studies*, 4(3), 182–191.

Hoover, Edgar M., J. (1936): “The Measurement of Industrial Localization,” *The Review of Economics and Statistics*, 18(4), pp. 162–171.

Hotelling, H. (1929): “Stability in competition,” *Economic Journal*, 39, 41–57.

Hummels, D. (1999): “Toward a Geography of Trade Costs,” GTAP Working Papers 17, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Isard, W. (1949): “The general theory of location and space-economy,” *The Quarterly Journal of Economics*, 63, 476–506.

Kaldor, N. (1935): “Market imperfection and excess capacity,” *Economica*, 2(5), 33–50.

Karlsson, C. (2008): “Introduction,” in *Handbook of research on cluster theory*, ed. by C. Karlsson, vol. 1, chap. 1, pp. 1–19. Edward Elgar Publishing.

Kim, S. (1995): “Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987,” *The Quarterly Journal of Economics*, 110(4), 881–908.

——— (1999): “Regions, resources, and economic geography: Sources of US regional comparative advantage, 1880–1987,” *Regional Science and Urban Economics*, 29(1), 1–32.

Kim, Y., D. L. Barkley, and M. S. Henry (2000): “Industry characteristics linked to establishment concentrations in nonmetropolitan areas,” *Journal of Regional Science*, 40(2), 234–259.

Klier, T., and D. P. McMillen (2008): “Evolving Agglomeration in the Us Auto Supplier Industry*,” *Journal of Regional Science*, 48(1), 245–267.

Koh, H.-J., and N. Riedel (2014): “Assessing the localization pattern of German manufacturing and service industries: a distance-based approach,” *Regional Studies*, 48(5), 823–843.

Koopmans, T. C., and M. Beckmann (1957): “Assignment problems and the location of economic activities,” *Econometrica*, 25(1), 53–76.

Krugman, P. (1980): “Scale economies, product differentiation, and the pattern of trade,” *The American Economic Review*, 70(5), 950–959.

——— (1991a): “Increasing returns and economic geography,” *The Journal of Political Economy*, 99(3), 483–499.

——— (1993): "First nature, second nature, and metropolitan location," *Journal of Regional Science*, 33(2), 129–144.

——— (1998): "What's new about the new economic geography?," *Oxford review of economic policy*, 14(2), 7–17.

——— (1999): "The role of geography in development," *International regional science review*, 22(2), 142–161.

——— (2009): "The increasing returns revolution in trade and geography," *The American Economic Review*, 99(3), 561–571.

Krugman, P., and A. J. Venables (1995): "Globalization and the Inequality of Nations," *The Quarterly Journal of Economics*, 110(4), 857–880.

Krugman, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of international Economics*, 9(4), 469–479.

——— (1981): "Intraindustry specialization and the gains from trade," *The Journal of Political Economy*, 89(5), 959–973.

——— (1991b): *Geography and trade*. MIT press.

Lafourcade, M., and G. Mion (2003): "Concentration, Spatial Clustering and the Size of Plants: Disentangling the sources of co-location externalities," *Available at SSRN: http://ssrn.com/abstract=981397*.

——— (2007): "Concentration, agglomeration and the size of plants," *Regional Science and Urban Economics*, 37(1), 46–68.

Launhardt, W. (1885): *Mathematische Begründung Der Volkswirtschaftslehre*. Leipzig: B.G. Teubner. Translated from German by B.G. Teubner (1993) Mathematical Principles of Economics, Aldershot: Edward Elgar.

Leahy, A., A. Palangkaraya, and J. Yong (2010): "Geographical Agglomeration in Australian Manufacturing," *Regional Studies*, 44(3), 299–314.

Lösch, A. (1940): *The Economics of Location*. English translation by Wiiliam H. Woglom, New Haven, Yale University Press, 1954.

Lu, J., and Z. Tao (2009): "Trends and determinants of China's industrial agglomeration," *Journal of Urban Economics*, 65(2), 167–180.

Lu, Z., A. Flegg, and X. Deng (2011): "Regional specialization: a measure method and the trends in China," MPRA Paper 33867, University Library of Munich, Germany.

Marcon, E., and F. Puech (2003): "Evaluating the geographic concentration of industries using distance-based methods," *Journal of Economic Geography*, 3(4), 409–428.

——— (2012): "A typology of distance-based measures of spatial concentration," <halshs-00679993v3>.

Marshall, A. (1890): *Principles of eEonomics*. Macmillan London, 8th edition published in 1920.

Martin, R., and P. Sunley (2003): "Deconstructing clusters: chaotic concept or policy panacea?," *Journal of Economic Geography*, 3(1), 5–35.

Maurel, F., and B. Sédillot (1999): "A measure of the geographic concentration in French manufacturing industries," *Regional Science and Urban Economics*, 29(5), 575–604.

Midelfart-Kvarnik, K.-H., H. G. Overman, S. J. Redding, and A. J. Venables (2004): "5. The Location of European Industry," in *European Integration and the Functioning of Product Markets*, ed. by K. S. Adriaan H. Dierx, Fabienne Ilzkovitz, pp. 113–168. Edward Elgar Publishing.

Moineddin, R., J. Beyene, and E. Boyle (2003): "On the location quotient confidence interval," *Geographical Analysis*, 35(3), 249–256.

Nakajima, K., Y. Saito, and I. Uesugi (2012): "Measuring economic localization: Evidence from Japanese firm-level data," *Journal of the Japanese and International Economies*.

O'Donoghue, D., and B. Gleave (2004): "A note on methods for measuring industrial agglomeration," *Regional Studies*, 38(4), 419–427.

Ohlin, B. (1933): *Interregional and International Trade,*. Cambridge, MA: Harvard University Press.

Openshaw, S. (1977): "A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling," *Transactions of the Institute of British Geographers*, 2(4), pp. 459–472.

Openshaw, S. (1984): *The Modifiable Areal Unit Problem*, CATMOG - Concepts and Techniques in Modern Geography No:38. Geo Books, Norwich, Regency House, 34 Duke Street, Norwich NR3 3AP, England, ISBN: 0860941345, ISSN: 0306-6142.

Ottaviano, G., and J.-F. Thisse (2004): "Agglomeration and economic geography," *Handbook of regional and urban economics*, 4, 2563–2608.

Palander, T. (1935): *Beiträge Zur Standortstheorie*. Uppsala: Almqvist & Wiksells Boktryckeri AB.

Pede, V. O., R. J. Florax, and H. L. de Groot (2011): "Technological Leadership and Sectoral Employment Growth:A Spatial Econometric Analysis for U.S. Counties," Working Papers 11-1, Purdue University, College of Agriculture, Department of Agricultural Economics.

Puga, D. (2010): "The Magnitude and Causes of Agglomeration Economies," *Journal of Regional Science*, 50(1), 203–219.

Redding, S. J. (2010): "The Empirics of New Economic Geography," *Journal of Regional Science*, 50(1), 297–311.

Ripley, B. (1977): "Modelling spatial patterns," *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 172–212.

Robinson, J. (1933): *Economics of Imperfect Competition*. The London: Macmillan.

Rodrik, D. (2007): *Industrial Development for the 21st Century*chap. 1.1. Industrial Development: Stylized Facts and Policy Directions, pp. 7–29. United Nations, New York.

Romer, P. M. (1986): "Increasing returns and long-run growth," *The journal of political economy*, 94(5), 1002–1037.

——— (1987): "Growth based on increasing returns due to specialization," *The American Economic Review*, 77(2), 56–62.

——— (1990): "Are Nonconvexities Important for Understanding Growth?," *American Economic Review*, 80(2), 97–103.

Rosenthal, S., and W. Strange (2004): "Evidence on the nature and sources of agglomeration economies," in *Handbook of Regional and Urban Economics*, ed. by V. J.Henderson, and J. F. Thisse, vol. 4, chap. 49, pp. 2119–2171. Elsevier, Amsterdam: North-Holland.

Rosenthal, S. S., and W. C. Strange (2001): "The determinants of agglomeration," *Journal of Urban Economics*, 50(2), 191–229.

Ruiz-Valenzuela, J., R. Moreno-Serrano, and E. Vaya-Valcarce (2006): “Concentration of the Economic Activity: Comparing Methodologies and Geographic Units,” ERSA conference papers ersa06p197, European Regional Science Association.

Samuelson, P. A. (1954): “The transfer problem and transport costs, II: Analysis of effects of trade impediments,” *The Economic Journal*, 64, 264–289.

——— (1983): “Thünen at two hundred,” *Journal of Economic Literature*, 21, 1468–1488.

Scholl, T., and T. Brenner (2012): “Detecting Spatial Clustering Using a Firm-Level Cluster Index,” Working Papers on Innovation and Space 2012-02, Philipps University Marburg, Department of Geography.

Scitovsky, T. (1954): “Two Concepts of External Economies,” *Journal of Political Economy*, 62(2), 143–151.

Sohn, J. (2004a): “Do birds of a feather flock together?: Economic linkage and geographic proximity,” *The Annals of Regional Science*, 38(1), 47–73.

——— (2004b): “Information technology in the 1990s: More footloose or more location-bound?*,” *Papers in Regional Science*, 83(2), 467–485.

Starrett, D. (1978): “Market allocations of location choice in a model with free mobility,” *Journal of Economic Theory*, 17(1), 21–37.

Tian, Z. (2013): “Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method,” *Journal of Regional Analysis & Policy*, 43(2), 186.

Venables, A. J. (1996): “Equilibrium Locations of Vertically Linked Industries,” *International Economic Review*, 37(2), 341–359.

Vogiatzoglou, K. (2006): “Increasing Agglomeration or Dispersion? Industrial Specialization and Geographic Concentration in NAFTA,” *Journal of Economic Integration*, 21(2), pp. 379–396.

von Thünen, J. (1826): *Der Isolierte Staat in Beziehung auf Landscaft und Nationalökonomie.* English Translation by C.M. Warterberg, von Thünen’s Isolated State, Oxford: Pergamon Press, 1912, Hamburg.

Weber, A. (1909): *Theory of the Location of Industries.* Chicago: The University of Chicago Press, 1929. Translition by C. J. Friedrich.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*. MIT press, Cambridge, Massachusetts, 2 edn.

APPENDIX

Appendix I

Table AI.1: **Cross-correlation table, 2003**

Variables	Number of firms	γ	G_i	H	Gini
Number of firms	1.0000				
γ	-0.0692	1.0000			
G_i	-0.2464	0.7732	1.0000		
H	-0.3061	0.2142	0.7472	1.0000	
Gini	-0.5610	0.3109	0.6244	0.6118	1.0000

Table AI.2: **Cross-correlation table, 2008**

Variables	Number of firms	γ	G_i	H	Gini
Number of firms	1.0000				
γ	-0.0525	1.0000			
G_i	-0.2456	0.8299	1.0000		
H	-0.3520	0.1271	0.6223	1.0000	
Gini	-0.5504	0.3107	0.6165	0.6087	1.0000

Table AI.3: **Pooled OLS model-2**

	(1.a)	(1.b)	(2.a)	(2.b)	(3.a)	(3.b)	(4.a)	(4.b)
sh_purch_raw	-0.003	-0.003	-0.053	-0.053	-0.053	-0.053		
	(0.058)	(0.078)	(0.061)	(0.098)	(0.061)	(0.098)		
sh_RD_exp	0.247	0.247	0.225	0.225	0.258	0.258	0.287	0.287
	(0.303)	(0.447)	(0.269)	(0.411)	(0.282)	(0.422)	(0.286)	(0.430)
tra_cost1	0.011	0.011	0.013**	0.013*	0.012**	0.012*	0.013**	0.013*
	(0.007)	(0.009)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.007)
avfsize	0.000**	0.000	0.000**	0.000				
	(0.000)	(0.000)	(0.000)	(0.000)				
labint	-0.056***	-0.056*	-0.056***	-0.056*	-0.054***	-0.054*	-0.050***	-0.050*
	(0.020)	(0.033)	(0.019)	(0.031)	(0.019)	(0.031)	(0.016)	(0.026)
sh_purch_energy	0.739*	0.739						
	(0.405)	(0.707)						
Constant	0.098**	0.098*	0.147***	0.147*	0.152***	0.152*	0.114***	0.114***
	(0.044)	(0.057)	(0.048)	(0.078)	(0.049)	(0.078)	(0.010)	(0.016)
R^2	0.038	0.038	0.029	0.029	0.026	0.026	0.025	0.025
$Adj.R^2$	.0339681	.0339681	.0251522	.0251522	.0234478	.0234478	.0228657	.0228657
N	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000	1314.000
F	4.213	1.510	4.922	2.010	4.526	1.796	6.007	2.389

Standard errors in parentheses

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

(a)-robust s.e. (unclustered), (b)-cluster-robust s.e.

Table AI.4: **Pooled OLS model-3**

	(1.a)	(1.b)	(2.a)	(2.b)	(3.a)	(3.b)	(4.a)	(4.b)
sh_purch_raw	0.014	0.014	-0.087	-0.087	-0.086	-0.086		
	(0.060)	(0.086)	(0.061)	(0.093)	(0.060)	(0.093)		
sh_RD_emp	0.263	0.263	0.109	0.109	0.130	0.130	0.176	0.176
	(0.211)	(0.339)	(0.196)	(0.352)	(0.195)	(0.352)	(0.194)	(0.355)
tra_cost2	0.198***	0.198*	0.176***	0.176*	0.178***	0.178*	0.180***	0.180*
	(0.052)	(0.102)	(0.053)	(0.100)	(0.053)	(0.100)	(0.052)	(0.099)
avfsize	0.000	0.000	0.000	0.000				
	(0.000)	(0.000)	(0.000)	(0.000)				
labint	-0.069***	-0.069**	-0.067***	-0.067**	-0.066***	-0.066**	-0.060***	-0.060**
	(0.020)	(0.032)	(0.018)	(0.028)	(0.018)	(0.028)	(0.016)	(0.025)
sh_purch_energy	1.387***	1.387**						
	(0.366)	(0.684)						
Constant	0.053	0.053	0.157***	0.157**	0.158***	0.158**	0.097***	0.097***
	(0.050)	(0.075)	(0.050)	(0.079)	(0.050)	(0.079)	(0.013)	(0.024)
R^2	0.057	0.057	0.022	0.022	0.021	0.021	0.018	0.018
$Adj.R^2$	.0523848	.0523848	.0178296	.0178296	.0182848	.0182848	.0158897	.0158897
N	1331.000	1331.000	1331.000	1331.000	1331.000	1331.000	1331.000	1331.000
F	6.761	2.325	6.913	3.264	8.716	3.999	11.252	5.088

Standard errors in parentheses

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

(a)-robust s.e. (unclustered), (b)-cluster-robust s.e.

Table AI.5: **Pooled OLS model-4**

	(1.a)	(1.b)	(2.a)	(2.b)	(3.a)	(3.b)	(4.a)	(4.b)
sh_purch_raw	0.011	0.011	-0.085	-0.085	-0.085	-0.085		
	(0.060)	(0.088)	(0.060)	(0.093)	(0.060)	(0.093)		
sh_RD_exp	0.247	0.247	0.225	0.225	0.236	0.236	0.279	0.279
	(0.365)	(0.433)	(0.323)	(0.367)	(0.324)	(0.368)	(0.331)	(0.380)
tra_cost2	0.197***	0.197*	0.172***	0.172*	0.175***	0.175*	0.177***	0.177*
	(0.052)	(0.100)	(0.053)	(0.099)	(0.053)	(0.098)	(0.052)	(0.097)
avfsize	0.000	0.000	0.000	0.000				
	(0.000)	(0.000)	(0.000)	(0.000)				
labint	-0.069***	-0.069**	-0.067***	-0.067**	-0.067***	-0.067**	-0.060***	-0.060**
	(0.020)	(0.031)	(0.018)	(0.028)	(0.018)	(0.028)	(0.016)	(0.025)
sh_purch_energy	1.359***	1.359**						
	(0.364)	(0.681)						
Constant	0.056	0.056	0.156***	0.156*	0.158***	0.158**	0.098***	0.098***
	(0.050)	(0.076)	(0.050)	(0.079)	(0.050)	(0.080)	(0.013)	(0.024)
R^2	0.056	0.056	0.022	0.022	0.022	0.022	0.019	0.019
R^2	.0521522	.0521522	.0185609	.0185609	.0189898	.0189898	.0166998	.0166998
N	1332.000	1332.000	1332.000	1332.000	1332.000	1332.000	1332.000	1332.000
F	6.784	2.350	6.885	3.226	8.704	3.931	11.169	5.045

Standard errors in parentheses

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

(a)-robust s.e. (unclustered), (b)-cluster-robust s.e.

Appendix II

Table AII.1: **Statistical Classification of Economic Activities in The European Community NACE Rev. 1.1**

Group	Division	Class	Definition
DA			**Manufacture of food products, beverages and tobacco**
15			**Manufacture of food products and beverages**
	15.1		Production, processing and preserving of meat and meat products
		15.11	Production and preserving of meat
		15.12	Production and preserving of poultrymeat
		15.13	Production of meat and poultrymeat products
	15.2	15.20	Processing and preserving of fish and fish products
	15.3		Processing and preserving of fruit and vegetables
		15.31	Processing and preserving of potatoes
		15.32	Manufacture of fruit and vegetable juice
		15.33	Processing and preserving of fruit and vegetables n.e.c.
	15.4		Manufacture of vegetable and animal oils and fats
		15.41	Manufacture of crude oils and fats
		15.42	Manufacture of refined oils and fats
		15.43	Manufacture of margarine and similar edible fats
	15.5		Manufacture of dairy products
		15.51	Operation of dairies and cheese making
		15.52	Manufacture of ice cream
	15.6		Manufacture of grain mill products, starches and starch products
		15.61	Manufacture of grain mill products
		15.62	Manufacture of starches and starch products
	15.7		Manufacture of prepared animal feeds
		15.71	Manufacture of prepared feeds for farm animals
		15.72	Manufacture of prepared pet foods
	15.8		Manufacture of other food products
		15.81	Manufacture of bread; manufacture of fresh pastry goods and cakes
		15.82	Manufacture of rusks and biscuits; manufacture of preserved pastry goods and cakes
		15.83	Manufacture of sugar
		18.84	Manufacture of cocoa; chocolate and sugar confectionery
		15.85	Manufacture of macaroni, noodles, couscous and similar farinaceous products
		15.86	Processing of tea and coffee
		15.87	Manufacture of condiments and seasonings
		15.88	Manufacture of homogenized food preparations and dietetic food
		15.89	Manufacture of other food products n.e.c.
	15.9		Manufacture of beverages
		15.91	Manufacture of distilled potable alcoholic beverages
			Continued

Table AII.1 – continued

Group	Division	Class	Definition
		15.92	Production of ethyl alcohol from fermented materials
		15.93	Manufacture of wines
		15.94	Manufacture of cider and other fruit wines
		15.95	Manufacture of other non-distilled fermented beverages
		15.96	Manufacture of beer
		15.97	Manufacture of beer
		15.98	Production of mineral waters and soft drinks
16			**Manufacture of tobacco products**
	16.0	16.00	Manufacture of tobacco products
DB			**Manufacture of textiles**
17			**Manufacture of textiles**
	17.1		Preparation and spinning of textile fibres
		17.11	Preparation and spinning of cotton-type fibres
		17.12	Preparation and spinning of woollen-type fibres
		17.13	Preparation and spinning of worsted-type fibres
		17.14	Preparation and spinning of flax-type fibres
		17.15	Throwing and preparation of silk, including from noils, and throwing and texturing of synthetic or artificial filament yarns
		17.16	Manufacture of sewing threads
		17.17	Preparation and spinning of other textile fibres
	17.2		Textile weaving
		17.21	Cotton-type weaving
		17.22	Woollen-type weaving
		17.23	Worsted-type weaving
		17.24	Silk-type weaving
		17.25	Other textile weaving
	17.3	17.30	Finishing of textiles
	17.4	17.40	Manufacture of made-up textile articles, except apparel
	17.5		Manufacture of other textiles
		17.51	Manufacture of carpets and rugs
		17.52	Manufacture of cordage, rope, twine and netting
		17.53	Manufacture of non-wovens and articles made from non-wovens, except apparel
		17.54	Manufacture of other textiles n.e.c.
	17.6	17.60	Manufacture of knitted and crocheted fabrics
	17.7		Manufacture of knitted and crocheted articles
		17.71	Manufacture of knitted and crocheted hosiery
		17.72	Manufacture of knitted and crocheted pullovers, cardigans and similar articles
18			**Manufacture of wearing apparel; dressing and dyeing of fur**
	18.1	18.10	Manufacture of leather clothes
	18.2		Manufacture of other wearing apparel and accessories
		18.21	Manufacture of workwear

Continued

Table AII.1 – continued

Group	Division	Class	Definition
		18.22	Manufacture of other outerwear
		18.23	Manufacture of underwear
		18.24	Manufacture of other wearing apparel and accessories n.e.c.
	18.3	18.30	Dressing and dyeing of fur; manufacture of articles of fur
DC			**Manufacture of leather and leather products**
19			**Tanning and dressing of leather; manufacture of luggage**, **handbags, saddlery, harness and footwear**
	19.1	19.10	Tanning and dressing of leather
	19.2	19.20	Manufacture of luggage, handbags and the like, saddlery and harness
	19.3	19.30	Manufacture of footwear
DD			**Manufacture of wood and wood products**
20			**Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and** plaiting materials
	20.1	20.10	Sawmilling and planing of wood; impregnation of wood
	20.2	20.20	Manufacture of veneer sheets; manufacture of plywood, laminboard, particle board, fibre board and other panels and boards
	20.3	20.30	Manufacture of builders' carpentry and joinery
	20.4	20.40	Manufacture of wooden containers
	20.5		Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials
		20.51	Manufacture of other products of wood
		20.52	Manufacture of articles of cork, straw and plaiting materials
DE			**Manufacture of pulp, paper and paper products; publishing and printing**
21			**Manufacture of paper and paper products**
	21.1		Manufacture of pulp, paper and paperboard
		21.11	Manufacture of pulp
		21.12	Manufacture of paper and paperboard
	21.2		Manufacture of articles of paper and paperboard
		21.21	Manufacture of corrugated paper and paperboard and of containers of paper and paperboard
		21.22	Manufacture of household and sanitary goods and of toilet requisites
		21.23	Manufacture of paper stationery
		21.24	Manufacture of paper stationery
		21.25	Manufacture of other articles of paper and paperboard n.e.c.
22			**Publishing,printing and reproduction of recorded media**
	22.1		Publishing
		22.11	Publishing of books
		22.12	Publishing of newspapers
		22.13	Publishing of journals and periodicals
			Continued

Table AII.1 – continued

Group	Division	Class	Definition
		22.14	Publishing of sound recordings
		22.15	Other publishing
	22.2		Printing and service activities related to printing
		22.21	Printing of newspapers
		22.22	Printing n.e.c.
		22.23	Bookbinding
		22.24	Pre-press activities
		22.25	Ancillary activities related to printing
	22.3		Reproduction of recorded media
		22.31	Reproduction of sound recording
		22.32	Reproduction of video recording
		22.33	Reproduction of computer media
DF			**Manufacture of coke, refined petroleum products and nuclear fuel**
23			**Manufacture of coke, refined petroleum products and nuclear fuel**
	23.1	23.10	Manufacture of coke oven products
	23.2	23.20	Manufacture of refined petroleum products
	23.3	23.30	Processing of nuclear fuel
DG			**Manufacture of chemicals, chemical products and man-made fibres**
24			**Manufacture of chemicals and chemical products**
	24.1		Manufacture of basic chemicals
		24.11	Manufacture of industrial gases
		24.12	Manufacture of dyes and pigments
		24.13	Manufacture of other inorganic basic chemicals
		24.14	Manufacture of other organic basic chemicals
		24.15	Manufacture of fertilizers and nitrogen compounds
		24.16	Manufacture of plastics in primary forms
		24.17	Manufacture of synthetic rubber in primary forms
	24.2	24.20	Manufacture of pesticides and other agro-chemical products
	24.3	24.30	Manufacture of paints, varnishes and similar coatings, printing ink and mastics
	24.4		Manufacture of pharmaceuticals, medicinal chemicals and botanical products
		24.41	Manufacture of basic pharmaceutical products
		24.42	Manufacture of pharmaceutical preparations
	24.5		Manufacture of soap and detergents, cleaning and polishing preparations, perfumes and toilet preparations
		24.51	Manufacture of soap and detergents, cleaning and polishing preparations
		24.52	Manufacture of perfumes and toilet preparations
	24.6		Manufacture of other chemical products
		24.61	Manufacture of explosives

Continued

Table AII.1 – continued

Group	Division	Class	Definition
		24.62	Manufacture of explosives
		24.63	Manufacture of explosives
		24.64	Manufacture of photographic chemical material
		24.65	Manufacture of prepared unrecorded media
		24.66	Manufacture of other chemical products n.e.c.
	24.7	24.70	Manufacture of man-made fibres
DH			**Manufacture of rubber and plastics products**
25			**Manufacture of rubber and plastic products**
	25.1		Manufacture of rubber products
		25.11	Manufacture of rubber tyres and tubes
		25.12	Retreading and rebuilding of rubber tyres
		25.13	Manufacture of other rubber products
	25.2		Manufacture of plastic products
		25.21	Manufacture of plastic plates, sheets, tubes and profiles
		25.22	Manufacture of plastic packing goods
		25.23	Manufacture of builders' ware of plastic
		25.24	Manufacture of other plastic products
DI			**Manufacture of other non-metallic mineral products**
26			**Manufacture of other non-metallic mineral products**
	26.1		Manufacture of glass and glass products
		26.11	Manufacture of flat glass
		26.12	Shaping and processing of flat glass
		26.13	Manufacture of hollow glass
		26.14	Manufacture of glass fibres
		26.15	Manufacture and processing of other glass,
	26.2		Manufacture of non-refractory ceramic goods other than for construction purposes; manufacture of refractory ceramic prod.
		26.21	Manufacture of ceramic household and ornamental articles
		26.22	Manufacture of ceramic sanitary fixtures
		26.23	Manufacture of ceramic insulators and insulating fittings
		26.24	Manufacture of other technical ceramic products
		26.25	Manufacture of other ceramic products
		26.26	Manufacture of refractory ceramic products
	26.3	26.30	Manufacture of ceramic tiles and flags
	26.4	26.40	Manufacture of bricks, tiles and construction products, in baked clay
	26.5		Manufacture of cement, lime and plaster
		26.51	Manufacture of cement
		26.52	Manufacture of lime
		26.53	Manufacture of plaster
	26.6		Manufacture of articles of concrete, plaster and cement
		26.61	Manufacture of concrete products for construction purposes
		26.62	Manufacture of plaster products for construction purposes
		26.63	Manufacture of ready-mixed concrete

Continued

Table AII.1 – continued

Group	Division	Class	Definition
		26.64	Manufacture of mortars
		26.65	Manufacture of fibre cement
		26.66	Manufacture of other articles of concrete, plaster and cement
	26.7	26.70	Cutting, shaping and finishing of ornamental and building stone
	26.8		Manufacture of other non-metallic mineral products
		26.81	Production of abrasive products
		26.82	Manufacture of other non-metallic mineral products n.e.c.
DJ			**Manufacture of basic metals and fabricated metal products**
27			**Manufacture of basic metals**
	27.1	27.10	Manufacture of basic iron and steel and of ferro-alloys
	27.2		Manufacture of tubes
		27.21	Manufacture of cast iron tubes
		27.22	Manufacture of steel tubes
	27.3		Other first processing of iron and steel
		27.31	Cold drawing
		27.32	Cold rolling of narrow strip
		27.33	Cold forming or folding
		27.34	Wire drawing
	27.4		Manufacture of basic precious and non-ferrous metals
		27.41	Precious metals production
		27.42	Aluminium production
		27.43	Lead, zinc and tin production
		27.44	Copper production
		27.45	Other non-ferrous metal production
	27.5		Casting of metals
		27.51	Casting of iron
		27.52	Casting of steel
		27.53	Casting of light metals
		27.54	Casting of other non-ferrous metals
28			**Manufacture of fabricated metal products, except machinery and equipment**
	28.1		Manufacture of structural metal products
		28.11	Manufacture of metal structures and parts of structures
		28.12	Manufacture of builders' carpentry and joinery of metal
	28.2		Manufacture of tanks, reservoirs and containers of metal; manufacture of central heating radiators and boilers
		28.21	Manufacture of tanks, reservoirs and containers of metal
		28.22	Manufacture of central heating radiators and boilers
	28.3	28.30	Manufacture of steam generators, except central heating hot water boilers
	28.4	28.40	Forging, pressing, stamping and roll forming of metal; powder metallurgy
	28.5		Treatment and coating of metals; general mechanical engineering

Continued

Table AII.1 – continued

Group	Division	Class	Definition
		28.51	Treatment and coating of metals
		28.52	General mechanical engineering
	28.6		Manufacture of cutlery, tools and general hardware
		28.61	Manufacture of cutlery
		28.62	Manufacture of tools
		28.63	Manufacture of locks and hinges
	28.7		Manufacture of other fabricated metal products
		28.71	Manufacture of steel drums and similar containers
		28.72	Manufacture of light metal packaging
		28.73	Manufacture of wire products
		28.74	Manufacture of fasteners, screw machine products, chain and springs
		28.75	Manufacture of other fabricated metal products n.e.c.
DK			**Manufacture of machinery and equipment**
29			**Manufacture of machinery and equipment n.e.c.**
	29.1		Manufacture of machinery for the production and use of mechanical power, except aircraft, vehicle and cycle engines
		29.11	Manufacture of engines and turbines, except aircraft, vehicle and cycle engines
		29.12	Manufacture of pumps and compressors
		29.13	Manufacture of taps and valves
		29.14	Manufacture of bearings, gears, gearing and driving elements
	29.2		Manufacture of other general purpose machinery
		29.21	Manufacture of furnaces and furnace burners
		29.22	Manufacture of lifting and handling equipment
		29.23	Manufacture of non-domestic cooling and ventilation equipment
		29.24	Manufacture of other general purpose machinery n.e.c.
	29.3		Manufacture of agricultural and forestry machinery
		29.31	Manufacture of agricultural tractors
		29.32	Manufacture of other agricultural and forestry machinery
	29.4		Manufacture of machine tools
		29.41	Manufacture of portable hand held power tools
		29.42	Manufacture of other metalworking machine tools
		29.43	Manufacture of other machine tools n.e.c.
	29.5		Manufacture of other special purpose machinery
		29.51	Manufacture of machinery for metallurgy
		29.52	Manufacture of machinery for mining, quarrying and construction
		29.53	Manufacture of machinery for food, beverage and tobacco processing
		29.54	Manufacture of machinery for textile, apparel and leather production
		29.55	Manufacture of machinery for paper and paperboard production
		29.56	Manufacture of other special purpose machinery n.e.c.

Continued

Table AII.1 – continued

Group	Division	Class	Definition
	29.6	29.60	Manufacture of weapons and ammunition
	29.7		Manufacture of domestic appliances n.e.c.
		29.71	Manufacture of electric domestic appliances
		29.72	Manufacture of non-electric domestic appliances
DL			**Manufacture of electrical and optical equipment**
30			**Manufacture of office machinery and computers**
	30.0		Manufacture of office machinery and computers
		30.01	Manufacture of office machinery
		30.02	Manufacture of computers and other information processing equipment
31			**Manufacture of electrical machinery and apparatus n.e.c.**
	31.1	31.10	Manufacture of electric motors, generators and transformers
	31.2	31.20	Manufacture of electricity distribution and control apparatus
	31.3	31.30	Manufacture of insulated wire and cable
	31.4	31.40	Manufacture of accumulators, primary cells and primary batteries
	31.5	31.50	Manufacture of lighting equipment and electric lamps
	31.6		Manufacture of electrical equipment n.e.c.
		31.61	Manufacture of electrical equipment for engines and vehicles n.e.c.
		31.62	Manufacture of other electrical equipment n.e.c.
32			**Manufacture of radio, television and communication equipment and apparatus**
	32.1	32.10	Manufacture of electronic valves and tubes and other electronic components
	32.2	32.20	Manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy
	32.3	32.30	Manufacture of television and radio receivers, sound or video recording or reproducing apparatus and associated goods
33			**Manufacture of medical, precision and optical instruments, watches and clocks**
	33.1	33.10	Manufacture of medical and surgical equipment and orthopaedic appliances
	33.2	33.20	Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process control equipment
	33.3	33.30	Manufacture of industrial process control equipment
	33.4	33.40	Manufacture of optical instruments and photographic equipment
	33.5	33.50	Manufacture of watches and clocks
DM			**Manufacture of transport equipment**
34			**Manufacture of motor vehicles, trailers and semi-trailers**
	34.1	34.10	Manufacture of motor vehicles
	34.2	34.20	Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers
	34.3	34.30	Manufacture of parts and accessories for motor vehicles and their engines

Continued

Table AII.1 – continued

Group	Division	Class	Definition
35			**Manufacture of other transport equipment**
	35.1		Building and repairing of ships and boats
		35.11	Building and repairing of ships
		35.12	Building and repairing of pleasure and sporting boats
	35.2	35.20	Manufacture of railway and tramway locomotives and rollingstock
	35.3	35.30	Manufacture of aircraft and spacecraft
	35.4		Manufacture of motorcycles and bicycles
		35.41	Manufacture of motorcycles
		35.42	Manufacture of bicycles
		35.43	Manufacture of invalid carriages
	35.5	35.50	Manufacture of other transport equipment n.e.c.
DN			**Manufacturing n.e.c.**
36			**Manufacture of furniture; manufacturing n.e.c.**
	36.1		Manufacture of furniture
		36.11	Manufacture of chairs and seats
		36.12	Manufacture of other office and shop furniture
		36.13	Manufacture of other kitchen furniture
		36.14	Manufacture of other furniture
		36.15	Manufacture of mattresses
	36.2		Manufacture of jewellery and related articles
		36.21	Striking of coins
		36.22	Manufacture of jewellery and related articles n.e.c.
	36.3	36.30	Manufacture of musical instruments
	36.4	36.40	Manufacture of sports goods
	36.5	36.50	Manufacture of games and toys
	36.6		Miscellaneous manufacturing n.e.c.
		36.61	Manufacture of imitation jewellery
		36.62	Manufacture of brooms and brushes
		36.63	Other manufacturing n.e.c.
37			**Recycling**
	37.1	37.10	Recycling of metal waste and scrap
	37.2	37.20	Recycling of non-metal waste and scrap

Table AII.2: **Nomenclature of Territorial Units for Statistics (NUTS) of Turkey**

NUTS-1	NUTS-2	NUTS-3
İstanbul Region (TR1)	İstanbul Subregion (TR10)	İstanbul Province (TR100)
West Marmara Region (TR2)	Tekirdağ Subregion (TR21)	Tekirdağ Province (TR211)
		Edirne Province (TR212)
		Kırklareli Province (TR213)
	Balıkesir Subregion (TR22)	Balıkesir Province (TR221)
		Çanakkale Province (TR222)
Aegean Region (TR3)	İzmir Subregion (TR31)	İzmir Province (TR310)
	Aydın Subregion (TR32)	Aydın Province (TR321)
		Denizli Province (TR322)
		Muğla Province (TR323)
	Manisa Subregion (TR33)	Manisa Province (TR331)
		Afyonkarahisar Province (TR332)
		Kütahya Province (TR333)
		Uşak Province (TR334)
East Marmara Region (TR4)	Bursa Subregion (TR41)	Bursa Province (TR411)
		Eskişehir Province (TR412)
		Bilecik Province (TR413)
	Kocaeli Subregion (TR42)	Kocaeli Province (TR421)
		Sakarya Province (TR422)
		Düzce Province (TR423)
		Bolu Province (TR424)
		Yalova Province (TR425)
West Anatolia Region (TR5)	Ankara Subregion (TR51)	Ankara Province (TR511)
	Konya Subregion (TR52)	Konya Province (TR521)
		Karaman Province (TR522)
Mediterranean Region (TR6)	Antalya Subregion (TR61)	Antalya Province (TR611)
		Isparta Province (TR612)
		Burdur Province (TR613)
	Adana Subregion (TR62)	Adana Province (TR621)
		Mersin Province (TR622)
	Hatay Subregion (TR63)	Hatay Province (TR631)
		Kahramanmaraş Province (TR632)
		Osmaniye Province (TR633)
Central Anatolia Region (TR7)	Kırıkkale Subregion (TR71)	Kırıkkale Province (TR711)
		Aksaray Province (TR712)
		Niğde Province (TR713)
		Nevşehir Province (TR714)
		Kırşehir Province (TR715)
	Kayseri Subregion (TR72)	Kayseri Province (TR721)
		Sivas Province (TR722)
		Yozgat Province (TR723)
West Black Sea Region (TR8)	Zonguldak Subregion (TR81)	Zonguldak Province (TR811)
		Continued

Table AII.2 – continued

NUTS-1	NUTS-2	NUTS-3
		Karabük Province (TR812)
		Bartın Province (TR813)
	Kastamonu Subregion (TR82)	Kastamonu Province (TR821)
		Çankırı Province (TR822)
		Sinop Province (TR823)
	Samsun Subregion (TR83)	Samsun Province (TR831)
		Tokat Province (TR832)
		Çorum Province (TR833)
		Amasya Province (TR834)
East Black Sea Region (TR9)	Trabzon Subregion (TR90)	Trabzon Province (TR901)
		Ordu Province (TR902)
		Giresun Province (TR903)
		Rize Province (TR904)
		Artvin Province (TR905)
		Gümüşhane Province (TR906)
Northeast Anatolia Region (TRA)	Erzurum Subregion (TRA1)	Erzurum Province (TRA11)
		Erzincan Province (TRA12)
		Bayburt Province (TRA13)
	Ağrı Subregion (TRA2)	Ağrı Province (TRA21)
		Kars Province (TRA22)
		Iğdır Province (TRA23)
		Ardahan Province (TRA24)
Central East Anatolia Region (TRB)	Malatya Subregion (TRB1)	Malatya Province (TRB11)
		Elazığ Province (TRB12)
		Bingöl Province (TRB13)
		Dersim Province (TRB14)
	Van Subregion (TRB2)	Van Province (TRB21)
		Muş Province (TRB22)
		Bitlis Province (TRB23)
		Hakkari Province (TRB24)
Southeast Anatolia Region (TRC)	Gaziantep Subregion (TRC1)	Gaziantep Province (TRC11)
		Adıyaman Province (TRC12)
		Kilis Province (TRC13)
	Şanlıurfa Subregion (TRC2)	Şanlıurfa Province (TRC21)
		Diyarbakır Province (TRC22)
	Mardin Subregion (TRC3)	Mardin Province (TRC31)
		Batman Province (TRC32)
		Şırnak Province (TRC33)
		Siirt Province (TRC34)

OYA KENT
Okan University, Tuzla Campus
34959, Tuzla/Istanbul, TURKEY
Email: oya.kent@okan.edu.tr

Date and place of birth

25/06/1982, Bursa, Turkey

Education

09/2010 – 09/2015: **Yıldız Technical University, İstanbul, Turkey**
PhD in Economics

10/2008 – 04/2010: **Università di Bologna, Bologna, Italy**
PhD in economics (- not completed)

09/2006 – 01/2007: **Universite Des Sciencse et Technologies De Lille 1, Lille, France**
Research study in the scope of Socrates-Erasmus Exchange programme

10/2004 – 08/2007: **Marmara University, Istanbul, Turkey**
M.A. in Economics (Eng.)

09/1999 – 06/2004: **Marmara University, Istanbul, Turkey**
B.A. in Economics (Eng.)

Academic positions and employment

02/2014 – 07/2014 : Universitat de Barcelona Barcelona-Spain
Research Institute of Applied Economics (IREA)
Regional Quantitative Analysis Research Group (AQR)
Full Time - Research Fellow

10/2010 – present : Okan University Istanbul-Turkey
Faculty of Economics and Administrative Sciences
Department of Banking and Finance – Okan University
Research Center for Financial Risks
Full Time – Research Assistant

09/2007 – 09/2008 : Bahçeşehir University Istanbul-Turkey
Faculty of Economics and Administrative Sciences
Department of Economics
Full Time – Research Assistant

Research interests

Economics of geography, regional economics, industrial economics, industrial policies and regional integration, panel data analysis, spatial data analysis and spatial econometrics

Language

English-Fluent, Turkish-Native, Italian-Advanced, French-Intermediate, Spanish-Intermediate

Computer skills

STATA, E-views, MATLAB, Latex, MS Office Applications

Awards, grants and scholarships

- Middle East Economic Association (MEEA) Travel Grant for ASSA 2015 Conference (2014)
- International Postdoctoral Research Fellowship Programme (for PhD Students)
  (Code: BIDEB 2214-A), by The Scientific and Technology Research Council of Turkey (TUBITAK), 2014
- PhD Scholarship, via Universita di Bologna, Bologna, ITALY, 2008
- Erasmus Grant , via European Commission for exchange studies at Université de Lille 1, Lille, FRANCE, 2006

Publications

Articles in Refereed Journals:

- "Ana Akıma Bir Alternatif: Nöroiktisat", (An Alternative to the Mainstream: Neuroeconomics), *Istanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi* (Istanbul Commerce University Journal Of Institute of Social Sciences), Vol: 10(19), 157-176, 2011.
- "Measuring commercial banks' performance in Turkey: A proposed model" with Suat Teker and Dilek Teker, *Journal of Applied Finance & Banking*, Vol: 1(3), 97-112, 2011
- "Long-run relation between interest rates and inflation: Evidence from Turkey" with Dilek Teker and Elçin Aykaç Alp, *Journal of Applied Finance & Banking*, Vol: 2(6), 41-54, 2012
- "Determinants of sovereign rating: Factor based ordered probit models for panel data analysis modeling framework" with Aynur Pala and Dilek Teker, International *Journal of Economics and Financial Issues,* Vol: 3(1), 122-132, 2012.

Working papers:

- "Religious regulation and economic prosperity: A cross-country examination" , Yıldız Technical University, Department of Economics, Istanbul, Advanced Econometrics Research Paper, 2012.
- "An examination of persistence of profits in Turkish Banking System" with Murat Donduran and Adem Yekeler, Yıldız Technical University, Department of Economics.

Papers presented in Refereed International Conferences

- "Agglomeration and Geographic Concentration in Turkish Manufacturing Industries"1st Annual International Conference on Social Sciences (AICSS) Istanbul, May 21-23, 2015
- "Agglomeration and Geographic Concentration in Turkish Manufacturing Industrie", 35th Annual Meeting of the MEEA, Allied Social Science Associations, Boston, MA, January 3 -6 2015
- "Unraveling the Institutional Mechanisms of Income Inequality in the MENA" with Fırat Bilgel and B.Can Karahasan, 35th Annual Meeting of the MEEA, Allied Social Science Associations, Boston, MA, January 3 -6 2015
- "Long-run relation between interest rates and inflation: Evidence from Turkey", Istanbul Finance Congress, Istanbul, 2012

Academic seminars, summer schools and trainings

- The International Iranian Economic Association (IIEA) "Advances in Economics: A PhD-Level Training Workshop", Bilgi University, İstanbul, Turkey August 26-29, 2015
- "MATLAB Seminars", lecturer Prof. Hüseyin Taştan, İstanbul Commerce University Faculty of Economics and Administrative Sciences, Sütlüce Campus, İstanbul, 2013.
- "EYS 2008- Economics Summer Seminars – Panel Data Econometrics (Theoretical), Panel Data Econometrics (Application)" lecturer Prof. Haluk Erlat and Assoc. Prof. Bülent Güloğlu, Pamukkale University & Economics Association of Turkey, Denizli, 2008

Memberships

American Economic Association (AEA), Middle East Economic Association (MEEA)

References

- Prof. Murat DONDURAN, Department of Economics, Yıldız Technical University
  E-mail: donduran@yildiz.edu.tr
- Prof. Fatma DOĞRUEL, Department of Economics, Marmara University
  E-mail: fatma.dogruel@marmara.edu.tr
- Prof. Erhan ASLANOĞLU, Department of Economics, Piri Reis University
  E-mail: easlanoglu@pirireis.edu.tr