openloc

working paper series

Agglomeration, related variety and vertical integration

Giulio Cainelli Donato Iacobucci

WP 4/2011

This research was funded by the Autonomous Province of Trento, as the sponsor of the OPENLOC research project under the call for proposals "Major Projects 2006". Partners of the project are: the E. Mach Foundation, the Manchester Institute of Innovation Research, the Trento Museum of Natural Sciences, the University of Bologna and the University of Trento

PROVINCIA AUTONOMA DI TRENTO

Agglomeration, related variety and vertical integration[1]

Giulio Cainelli[*]

Dipartimento di Scienze Economiche e Aziendali "Marco Fanno", Università degli Studi di Padova
Via del Santo 33, 35123, Padova, and CERIS-CNR, Milan, Italy
Phone: +39- 049 827.4227
Fax: +39- 049827.4211
giulio.cainelli@unipd.it

Donato Iacobucci

Dipartimento di Ingegneria Informatica, Gestionale e dell'Automazione, Università Politecnica delle Marche,
Via Brecce Bianche Monte Dago, Ancona, 60131, Italy
Phone: +39-071-2204482
Fax: +39-071-2204474
iacobucci@univpm.it

ABSTRACT

Several recent studies investigate the relation between geographic concentration of production and vertical integration, based on the hypothesis that spatial agglomeration of firms in the same industry facilitates input procurement thereby reducing the degree of vertical integration. The present paper contributes to this debate by also considering the effects of industry variety at the local level. Specifically, we consider two forms of variety: unrelated variety and vertically related variety. The latter index is constructed using information drawn from input-output tables and captures the opportunities for outsourcing within the local system. We consider inter-industry vertical integration by taking account of the ownership of activities with input-output linkages. Using a dataset of 24,663 Italian business groups in 2001, we estimate Tobit models to investigate the influence of vertically related variety and other agglomeration forces on the degree of vertical integration of groups. Our evidence confirms that vertical integration is influenced by industry specialization at the local level. We also find that the higher the vertically related variety, the lower the need for firms to integrate activities since they have more opportunities to acquire intermediate goods and services within the local system.

Keywords: vertical integration, agglomeration, related-variety, business groups

---

[1] We thank the Editor and three referees of this journal for their valuable and constructive comments and suggestions that allowed us to greatly improve the paper. We also thank Roberto Ganau and Yanlei Zhang for commenting a previous version of this paper.

[*] Corresponding author

1. INTRODUCTION

Several studies focus on the role of spatial agglomeration in affecting the organization and strategies of firms. They highlight, in particular, the relation between the geographic concentration of production and vertical integration (Cainelli and Iacobucci 2009; Holmes 1999; Helsley and Strange 2007; Brookfield 2008; Diez-Vial and Alvarez-Suescun 2010). The hypothesis common to these works is that spatial agglomeration of firms in the same industry facilitates input procurement and reduces the degree of vertical integration. All these empirical studies use information at plant or firm level, and although they adopt different methodological approaches and definitions, they find a negative relationship between productive specialization at the local level and vertical integration.

The present paper aims to contribute to this literature. Based on a dataset of 24,663 Italian business groups for the year 2001 and estimating Tobit models, we investigate the role of specialization and variety of local systems in influencing vertical integration decisions. Unlike previous works, we study inter-industry integration taking the business group as our unit of analysis and define vertical integration as the ownership and control of manufacturing and service activities along the same production and supply chain. Following a methodology developed by Fan et al. (2009), we measure vertical integration using information on the economic activities owned by the same business group, and information from input-output tables. This measure depends on two components: i) the presence of vertically related industries within the same group; ii) the intensity of the exchanges between these industries measured by the technical coefficients of the input-output table.

We consider the geographic concentration of firms belonging to the same industry (specialization), as well as other agglomeration forces such as unrelated and vertically related variety: we measure this latter by developing a new index which exploits information from input-output tables. Our geographical unit is province since most input-output exchanges related to key firm inputs take place within provinces.[2]

Starting from the premise that vertical integration choices are influenced by agglomeration, this paper makes three main contributions to the empirical literature. First, we show that in the case of inter-industry integration also, the higher the level of industry specialization at the local level the lower is the degree of vertical integration. Second, we find that the higher the vertically related variety of the local system the lower is the need for firms to integrate activities since they can more

[2] In Italy, province is an administrative entity between municipality and region. Italy is comprised of 103 provinces.

easily acquire intermediate goods and services from neighbor firms. In other words, we show that it is not variety '*per se*' that matters for vertical integration choices, but the presence of vertically related activities. Third, we show that the role of agglomeration forces is strongly conditioned by the spatial heterogeneity of the Italian economy. In particular, we find that in the so-called 'Third Italy' the link between agglomeration and vertical integration is stronger, confirming the role in this area played by industrial districts and local production systems, while in the so-called 'Industrial Triangle' (North-West area), characterized by the presence of larger firms and older industrial areas, the influence of specialization and local variety is not significant.

The paper is organized as follows. Section 2 provides a review of the literature on the relations between agglomeration, variety and vertical integration. Section 3 discusses the data and methodology. Section 4 presents the results of the empirical analysis. Section 5 discusses some robustness checks and Section 6 presents the main conclusions.

2. RELATED LITERATURE

There is an abundant and still growing literature on how the agglomeration of industries influences the performance and organizational choices of firms. Most of this work focuses on how the specialization and variety of local systems influence productivity, innovation and growth (Audretsch and Feldman 1996; Glaeser et al. 1992; Henderson, Kuncoro, and Turner 1995; Boschma and Iammarino 2009). There are a few studies that investigate the relations between industry structure at the local level – in terms of agglomeration and variety – and vertical integration choices (Brookfield 2008; Cainelli and Iacobucci 2009; Diez-Vial and Alvarez-Suescun 2010; Holmes 1999; Ono 2007). These studies differ in terms of the definition of vertical integration, agglomeration and variety, as well as the methodologies used to measure them and to test the empirical relations. However, all confirm the hypothesis of a negative relationship between the spatial agglomeration of industries and the degree of vertical integration of firms.

In this paper we develop this literature by considering inter-industry vertical integration at the business group level and developing an index of vertically related variety at the local level.

2.1 Vertical integration

There are several different ways of defining and measuring vertical integration. A first difference is in the span of activities included in the definition: they can be activities belonging to the same industry (intra-industry integration) or activities that encompass different industries (inter-industry integration). For example, Diez-Vial & Alvarez-Suescun (1999) study the degree of intra-industry integration by considering the presence of different meat processing stages in the same establishment. Studies that use the Adelman (1955) index (share of purchased input on sales) and consider overall input purchased by firms irrespective of their industry, refer to both intra- and inter-industry integration (Holmes 1999). Studies that use the input-output tables to detect the presence of vertical integration focus on inter-industry integration (Acemoglu et al. 2010; Cainelli and Iacobucci 2009).

Other differences are related to the organizational unit vertical integration is referred to: technical unit (plant) or economic unit (firm or business group). At the plant level, Diez-Vial & Alvarez-Suescun (2010) consider the presence of adjacent stages of production within the same plant and Ono (2007) looks at whether or not manufacturing establishments outsource specific business services. The majority of studies use firm level accounting data and industry level census data to calculate the Adelman index or some modified version of it (Holmes 1999; Li and Lu 2009). Other studies consider all the activities owned by the same firm and whether they show potential input-output linkages (Acemoglu et al. 2010; Cainelli and Iacobucci 2009).

In the present study we focus on inter-industry vertical integration, which we define as control over production activities in industries with input-output relations. Industries are delimited according to the NACE (2-digit level) classification of economic activities; input-output exchanges between industries are detected and measured using input-output tables. When a firm controls production activities belonging to different industries it is common to manage them in different plants, and most often, in different legal units; "vertical integration can be accomplished by placing two or more seemingly distinct firms under common control – that is, in a business group" (Fan et al. 2009).

We take the business group as our unit of analysis; this allows us to define inter-industry vertical integration as control, by the same owner, of manufacturing and service activities. For each group we construct a continuous index of vertical integration following the methodology developed by Fan et al. (2009). This index depends on two elements: i) the presence in the group of firms with potential input-output linkages (e.g. a textile firm and a clothing firm); ii) the intensity of the input-output relations between those firms, measured using the input-output coefficients of the industries to which the firms belong. In taking account of the ownership of vertically related activities we consider vertical integration as the result of strategic choices rather than technical constraints. Section 3.2 gives details on the calculation of this index of vertical integration.

2.2 The role of agglomeration and specialization

There is a small, but growing number of contributions investigating how agglomeration forces influence the vertical integration decisions of firms. The common hypothesis in these studies is that the spatial proximity of firms belonging to the same (or related) industries facilitates economic interactions among them (Boschma 2005); specifically, agglomeration should reduce opportunism, facilitate the co-ordination of production plans and investment, and reduce the transaction costs of market exchanges (Wood and Parr 2005; Enright 1995; Helsley and Strange 2007). For this reason, it is expected that agglomerated firms will be more disintegrated (less integrated) than firms operating outside these areas. This is true in the case of intra-industry integration, given the greater opportunities to exchange goods and services with firms in the same industry. It should apply also to the case of inter-industry integration since the agglomeration of firms belonging to a specific industry will attract related industries (suppliers) to locate in the same area (Porter 1998; Boschma and Wenting 2007).

Several empirical studies investigate the influence of agglomeration and specialization on the degree of vertical integration. Holmes (1999) studies the link between industry localization and vertical disintegration using plant level data on the U.S. manufacturing sector. He measures vertical disintegration using the Adelman index, defined as the ratio between the value of purchased inputs and sales (which he calls 'purchased-inputs intensity'). This index captures both intra- and inter-industry integration since it refers to all the inputs purchased by the firm. The concentration and variety of industries are measured using three different indicators: (i) employees working on the same activity (own-industry); (ii) employees working on different activities within the same two-digit code (related-industry); (iii) employees in different industries (other-manufacturing). The geographical unit in his study is the county. Holmes (1999) assumes that plants in the same county will be located at its geographic center; he then considers all the plants in the given county and in counties with centers within a 50 mile radius. He finds a positive (negative) relation between the level of own-industry (i.e. agglomeration) and vertical disintegration (integration); however, the presence of related-industries (i.e. related variety) and other-industries (i.e. unrelated variety) has little impact on the degree of vertical integration. As further confirmation of the role of agglomeration, he finds also that establishments of spatially concentrated industries show a lower degrees of vertical integration (Holmes 1999).

Brookfield (1999) provides similar results from an investigation of the relation between industry clustering and vertical integration, based on the Taichung machine tools district in Taiwan. To measure of vertical integration Brookfield uses materials and subcontracting costs as a percentage of total costs (a variation of the Adelman index). He compares the level of vertical integration among machine tools firms located in the Taichung district, with the level in firms located in other areas. The Taichung district is home to a large share (over 70%) of the machine tools firms in Taiwan, and it shows a high level of specialization (measured as the number of machine tools firms in total firms in the area). Brookfield (1999) finds that firms located in the Taichung area are less vertically integrated than firms located elsewhere in the island.

Ono (2007) estimates the effect of market thickness on the likelihood of outsourcing business services – that is, advertising, bookkeeping and accounting, legal services, and software and data-processing services - by U.S. manufacturing firms. He uses plant level data to detect whether or not services are outsourced (but not the share of outsourcing). His geographical unit is the primary metropolitan statistical area (PMSA). Market thickness is associated with the size of the local system, which is measured as the number of employees in the area. As a proxy for agglomeration and specialization Ono (2007) considers the share of industries that make intensive use of specific business services. He finds that plants located in larger cities are more likely to outsource services and specialization index shows little significance for explaining outsourcing decisions: only for legal services does the presence of industries that use this service intensively have a significant impact on the probability of outsourcing the service by other manufacturing firms located in the area.

Li and Lu (2008) replicate Holmes's (2009) methodology using data on Chinese manufacturing firms to measure the effects of the geographic concentration of industries on the vertical disintegration of firms. Following Holmes (1999), Li and Lu measure the degree of vertical disintegration using the Adelman index. The geographic concentration of industrial activity is measured at the level of the Chinese province (31), which is a much larger geographical unit than in other studies of spatial agglomeration. The population size of the average Chinese province is similar to the population size of many Western countries. Also, their analysis is not at firm level because the data are aggregated at levels of province and industry. The geographic concentration of activities is measured in terms of employment in neighbor firms (i.e. firms located in the same province). Li and Lu (1999) uses the same three indices of agglomeration and variety defined in Holmes (2009): the own-industry index, which considers employees in the same four-digit industry; the related-industry index, based on employees belonging to the same two-digit industry, but different four-digit industries; and the other-industry index, which considers employees in different two-digit industries. Li and Lu (2009) show that all three indices of geographic agglomeration play a significant role in determining the degree of disintegration; that is, the higher the level of geographic concentration of industrial activity, in both related and unrelated industries, the lower the needs for firms to integrate activities along the production chain.

Cainelli and Iacobucci (2009) consider backward vertical integration on the basis of control of production activities. They use data on Italian business groups to analyze the role of technology in influencing the choice of controlling activities along the production chain (inter-industry integration), taking account also of the role of agglomeration forces, captured by the location of the group in an industrial district. Consistent with Holmes (2009), they find that agglomeration has a negative impact on vertical integration, but the significance of the effect depends on the industry considered. They do not consider the presence of industry variety at the local level.

Diez-Vial & Alvarez-Suescun (2010) study the impact of agglomeration (physical proximity) on degree of vertical integration, using data on meat industry establishments in Spain. Vertical integration is based on the presence of different meat processing stages within the same establishment (intra-industry integration). Agglomeration is measured by the presence of firms belonging to the same industry, within different radiuses, ranging from 1 to 10 km. The authors find that firms located in agglomerated areas tend to internalize fewer stages in the production chain. This result holds for all distances considered, but the effect is stronger for firms located at a radius of between 1 and 2.5 km.

All these studies find that industry specialization at the local level (i.e. agglomeration)[3] plays a negative role in the level of vertical integration of firms. This results holds for intra-industry integration, given the higher possibility of exchanging goods and services with firms in the same industry. In the case of inter-industry integration, the empirical results are less clear cut (Cainelli and Iacobucci 2009; Ono 2007). However, the theory suggests that agglomeration of firms in the same industry should favor the presence in the same area of vertically related industries, thus enhancing the possibilities for outsourcing at the local level. As Marshall (1920) noted, "when an industry has thus chosen a locality for itself … 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" (Marshall 1920, IV.x.3). Porter is even more explicit on this point in stating that "clusters encompass an array of linked industries and other entities important to competition. They include, for example, suppliers of specialized inputs such as components, machinery and services, and providers of specialized infrastructures" (Porter 1998, p. 78).

For this reason, and in the case of inter-industry integration, we propose the following hypothesis:

H_1: The higher the level of industry specialization at local level, the lower the level of vertical integration of firms;

---

[3] Specialization, concentration and agglomeration at the local level refer to different concepts. However, given the high level of empirical correlation between them they are often used synonymously in the literature.

2.3 The role of related and unrelated variety

The empirical results on the relation between local level industry variety and vertical integration are more controversial. This could be due to theoretical and empirical reasons. At a theoretical level, the role of variety in fostering innovation through knowledge spillover across industries is well demonstrated; however, the mechanisms through which variety should favor market exchanges of intermediate goods – thereby reducing the need for vertical integration – are less clear. The weak empirical relation between variety and vertical integration could be attributable to the way variety is defined and measured. In the studies by Holmes (1999) and Li and Lu (2008) related and unrelated variety are defined on the basis of the system of classification of economic activities: related activities are considered as those with the same two-digit code, and unrelated activities as those belonging to different two-digit codes. This way of defining related and unrelated variety to study vertical integration, however, has several limitations. For example, the presence in the same area of textile and clothing firms (which belong to different two digit codes), and the presence of food and clothing firms, are both considered examples of unrelated variety, although the opportunities for local level outsourcing by clothing firms in the former case, is much higher than in the latter case. Similar problems occur in the case of related variety; industries in the same two-digit codes (e.g. men's and women's clothing), although relevant for knowledge spillovers, may not be similarly important in the case of input-output exchanges. For this reason, we develop an index of related variety that captures the opportunities for market exchanges at the local level; we call this index vertically related variety.[4]

A recent strand in the literature introduced the concept of *related variety* at local level, as a key concept to explain firm performance and regional growth (Boschma and Iammarino 2009; Boschma, Eriksson, and Lindgren 2009; Boschma, Minondo, and Navarro 2010; Frenken, Van Oort, and Verburg 2007). Previous work suggests that a diversified local structure may be preferable to specialization because variety facilitates the generation of new ideas, induces knowledge spillovers, and provides more valuable resources for radical innovation (Jacobs 1969; Duranton and Puga 2001; Glaeser et al. 1992). The debate on the role of specialization and variety is associated with two conceptualizations of knowledge spillovers at local level: specialization generates spillovers within the same industry as a result of so-called MAR (Marshall-Arrow-Romer) externalities; unrelated variety captures the phenomenon of cross-fertilization of ideas across different industries (so-called Jacobs externalities). The literature on related variety tries to address this debate by arguing that it is not variety per se that

---

[4] We thank an anonymous referee for suggesting this term.

matters, but the presence of industries that are related in terms of shared or complementary resources, knowledge, and institutions. Specialization is associated with similarity, which is beneficial for efficiency and incremental innovation but can hamper the possibilities for radical innovation and diversification because it can lead to “cognitive lock-in, because no much learning will take place when agents have exactly the same competences” (Boschma and Frenken 2009). At the same time, too much diversity (implied by the concept of unrelated variety) can be detrimental because some degree of cognitive proximity is required to ensure effective communication and interactive learning among firms from different sectors (Nooteboom et al. 2007; Nooteboom 2000). The debate on the role of related variety is concerned mainly with technological relatedness, its main aim being to study the conditions for local level learning, innovation, and diversification (Boschma and Frenken 2009; Frenken, Van Oort, and Verburg 2007).

An important empirical question is how to define and measure relatedness at the local level. As mentioned in the previous section, a common way to conceptualize related and unrelated variety is to use industry boundaries defined by industrial classification systems. The use of industrial classification codes to measure relatedness between sectors is justifiable from the point of view that classification systems group industries on the basis of common raw materials and production processes, but has some drawbacks because it “neglects the fact that spillovers might be best promoted when there is a degree of relatedness between relatively distinct sectors in terms of, for example, products, knowledge base, technology or skills” (Bishop and Gripaios 2010). Some recent studies have developed concepts and measures of variety at local level that account for presence of specific relations between industries (Frenken, Van Oort, and Verburg 2007).

Empirical studies have used two approaches to measure ‘relatedness’ between industries. One approach is to identify it indirectly (co-occurrence analysis) by observing the ‘revealed’ associations between industries at the firm or geographical level. For example, Hidalgo et al. (2007), following the methodological approach in Porter (2003), capture relatedness between sectors on the basis that “if two goods are related because they require similar institutions, infrastructure, physical factors, technology, or some combination thereof, they will tend to be produced in tandem…” (Hidalgo et al. 2007). Porter (2003) detects the presence of similar products by observing the production of goods within the same state; Hidalgo et al. (2007) measure proximity between industries by analyzing national export patterns.

Other studies consider the association within the same plant or firm, between products and employees’ skills. Teece et al. (1994) and Bryce & Winter (2009) consider the co-occurrence of products at firm level, while Neffke and Svensson Henning (2008) and Neffke et al. (2009) consider the co-occurrence of products at plant level. See also Boschma & Frenken (2009) for a discussion of these measures.

The second approach uses an a priori criterion of relatedness between industries. Frenken et al. (2007) define an index of technological relatedness between industries by considering the similarity between the input mixes of two industries. Boschma et al. (2010) propose two alternative measures of related variety: the first is based on Porter's (2003) cluster classification of industries; the second is based on export data and the idea that if a country reveals comparative advantages in products *i* and *j*, the two industries involved are likely to show commonalities in terms of factors of production and knowledge bases. Boschma and colleagues show that these two ways of defining and measuring related variety "better identify the relatedness across industries than the conventional measures based on standard industry or product classification" (Boschma, Minondo, and Navarro 2010).

To define our index of vertically related variety we follow this second approach. Our aim is to capture the role of related variety as a source of production efficiency through the exploitation of outsourcing opportunities at local level. We follow Fan and Lang (2000) and Frenken et al. (2007) to develop an index that uses a specific criterion for relatedness between industries that takes account of the opportunities for exchanges goods and services at the local level, identified by input-output data. We call this index vertically related variety. It depends on the presence of vertically related industries within the same geographical area, and on the intensity of input-output exchanges between these industries. Details on the construction of our index are provided in section 3.3.

The richer (in terms of input-output related industries) the local system in which a firm operates, the higher its outsourcing possibilities and the lower the need to control (integrate) activities along the production chain. Our resulting hypothesis is:

H_2: the higher the level of vertically related variety at local level, the lower the level of vertical integration of firms.

3. DATA AND METHODOLOGY

3.1 Business groups

For the empirical analysis we use an ISTAT (Italian National Statistical Institute) dataset on business groups.[5] Business groups are identified on the basis of control links between pairs of legal units. European level operational guidelines for the identification of control for statistical purposes

[5] The dataset was provided as part of a research agreement between ISTAT, CERIS-CNR and Università Politecnica delle Marche.

(Eurostat 2003) are that: i) a legal unit directly owns at least 50% plus 1, of the voting rights in another legal unit; ii) a legal unit owns an equal share of voting rights, with respect to other units, in another legal unit; iii) a legal unit, through its subsidiaries, owns at least 50% plus 1 of the voting rights in another legal unit; iv) a legal unit is fully responsible for the balance sheet statements of another legal unit and no other legal unit consolidates the accounts of that legal unit. The algorithm used to empirically identify business groups applies these operational guidelines to the shareholdings matrix to derive a control matrix.

The dataset on business groups was built by integrating three different statistical sources: (i) the archive of all shareholders of non-listed companies; (ii) the archive of all shareholders of listed companies; (iii) the archive of firms' consolidated accounts. These sources are sufficient to guarantee complete coverage of the shareholder structure of all limited companies. Elementary information on shareholdings from these three sources is integrated and chains of direct and indirect control are re-constructed. The data refer to 2001. For each legal unit belonging to a group, information is available on its activity (at the 5-digit level), location, number of employees, sales, share of ownership, etc. The industry and geographical location (province) of each group is determined by the economic activity and location of the largest company – measured by number of employees.[6]

To measure the degree of vertical integration we take account only of firms located in Italy. The original dataset contains 92,474 firms, 4,786 of which are located abroad. We do not have information on the activities and sales of the latter, and for this reason they are excluded from the analysis. This exclusion be considered a limitation of our study, given the increasing level of internationalization of production might. However, it is not so for four reasons. First, the number of companies controlled abroad accounts for only around 5% of the total. Second, consideration of domestic companies is more congruent with the input-output table for domestic exchanges that we use to measure vertical integration. Third, foreign owned companies are mostly controlled by large business groups and in the empirical estimate we control for the size of groups. Finally, international outsourcing is less widespread in Italy compared with other industrialized countries (Mariotti and Mutinelli 2005). When we exclude foreign located firms, the resulting dataset contains observations on 87,688 firms and 24,663 business groups.

Firms controlled by groups are mostly located in the same province (75% of group controlled firms). Even whn controlled firms are located in different provinces, the concentration of firms located in the same province as the head of the group is very high (78% of the total).

[6] For further details on the business groups dataset see Cainelli & Iacobucci (Cainelli and Iacobucci 2007).

3.2 Measuring vertical integration

One of the main problems in assessing the degree of vertical integration is the availability of information on control, by the same firm, of adjacent activities along the production chain (Acemoglu et al. 2010). To overcome this, we use data on the production activities controlled by a business group; in other words, we consider ownership and control of economic activities as the basis for identifying vertical integration. This is appropriate because the problems that arise from market transactions refer to contractual relations between independent firms.

The definition of a continuous measure of vertical integration in business groups is based on the methodology proposed by Fan and Lang (2000) and Fan et al. (2009). To compute this measure, we use the latest available *use* table at purchasing prices (ISTAT 2005).

Given that v_{ij} = value of input acquired by industry *i* from industry *j* on total output of industry *i*, v_{ji} = share of input acquired by industry *j* from industry *i* on total output of industry *j*, the input-output index for a pair of industries is defined as:

$$V_{ij} = \frac{v_{ij} + v_{ji}}{2}$$

The higher the index, the higher the input-output exchanges between the two industries. Excluding agriculture and fisheries, the input-output matrix available for the Italian economy includes 54 sectors.[7] Excluding exchanges within the same sector, we have 1,431 pairs for which we can calculate the value of V_{ij}.[8] For example, the clothing industry acquires 0.2780 euro from the textile sector for each euro of output; the textile sector acquires 0.0099 euro of input from the clothing industry for each euro of output. Thus, the input-output index for these two industries is 0.1439.

To identify vertical integration we use information on the activities controlled by the same business group. Some groups include more than one company belonging to the same industry.[9] To calculate our index of vertical integration we aggregate group sales by industry. This reduces the number of different codes in groups to 52,859. Given that the overall number of groups is 24,663, this means an average of about 2 codes per group.

If *i* is the primary industry in a group *f*, for each of the other industries $j \neq i$ controlled by the same group, we can construct a measure of the actual presence of vertical integration as the ratio of group *f* employees in industry *j* (the supply industry), and the employees in industry *i* (the acquiring industry).

---

[7] The Italian input-output table uses the classification of economic activities NaceRev.1.1 .

[8] This is the result of combining 2 industry codes out of 54: ${}_nC^k = \frac{n!}{k!(n-k)!}$.

[9] And also because we aggregate the industries to which a company belongs according to the aggregation level in the input-output table.

$$w_{ijf} = \frac{\text{emp}_{jf}}{\text{emp}_{if}}$$

Given that industry *i*, by definition, is the primary industry: $emp_i > emp_j$.

Moreover, it could be that $emp_j/emp_i > V_{ij}$. For this reason, to measure the effective presence of vertical integration for each group *f* and for each pair of industries *ij*, we redefine the index w_{ijf} in the following way:

$$w_{ijf} = \max\left(\frac{\text{emp}_{\text{jf}}}{\text{emp}_{\text{if}}}, V_{ij}\right)$$

For each group *f* (*f*=1, 2, …, 24,633), the effective presence of vertical integration is measured by the index:

$$\text{VI}_f = \sum_{j=1}^{n} w_{ijf}$$

where *i* denotes the primary industry of group *f*; $j \neq i$ indicates all the other industries in which group *f* has controlled firms. The value of this index depends on two factors: (i) the number of different industries controlled by the same group; (ii) the intensity of the input-output relations among those industries.

3.3 *Independent variables*

Following the theoretical discussion in section 2, we define different types of agglomeration forces: (i) geographical concentration of industries, to capture industry specialization at the local level; (ii) unrelated variety; (iii) vertically related variety.

Geographical concentration is measured using a Balassa specialization index calculated as follows:

$$\text{Spec}_{\text{i,k}} = \frac{\frac{l_{i,k}}{l_k}}{\frac{l_{i,Italy}}{l_{Italy}}}$$

where $l_{i,k}$ denotes employment of industry *i* in province *k*, l_k total employment of province *k*, $l_{i,Italy}$ total employment of industry *i* in Italy, and finally l_{Italy} total manufacturing employment in Italy.

Following Frenken et al. (2007), we measure unrelated variety within a province *k* as the entropy index at the two-digit level. The index is calculated for each province *k* as follows:

$$\text{Uvariety}_{\text{k}} = \sum_{g=1}^{G} S_g \log_2\left(\frac{1}{S_g}\right)$$

where S_g (g=1, 2, …, G) are the shares of employees at the two-digit level. The minimum possible value is zero, corresponding to the case where all employees are concentrated in the same industry. The maximum value is obtained when employees are distributed equally across all industries.[10] The higher the value of the index, the higher the variety of the industries located in the same geographical area (in our case the province).

For vertically related variety we adapt the index of relatedness between industries developed by Los (2000) and Frenken et al. (2007). Given $k=1\ldots m$: geographic areas (provinces in our case), $i=j=1\ldots n$: industries (at 2 digit level), s_i, s_j: employees in industries i and j , v_{ij} [0, 1]: measure of input-output exchanges between sectors, our index of vertically related variety is calculated for each province k as follows:

$$VR\,\text{variety}_k = \frac{\sum_{i}^{n}\sum_{j}^{n} s_{ik} s_{jk} v_{ij}}{\sum_{i}^{n}\sum_{j}^{n} s_{ik} s_{jk}}$$

The index varies from *1* (maximum value) to *1/n* (minimum value). It is clear from the construction (and easily verified through simulations) that the value of the index is influenced by both dimensions of: i) degree of concentration of activities in a few industries; ii) level of input-output exchanges between industries.[11] This index captures the opportunities for vertical exchanges between industries at local level. We have 52 industries (excluding agriculture and energy) and 103 provinces. We construct the index taking account of the number of employees in 1991 and 2001; the input-output index v_{ij} is calculated using the input-output table for 2000.

These measures of agglomeration are computed at province level. The Italian territory is split into 103 provinces (an intermediate level between municipality and region) that encompass the 686 local labor systems as defined by ISTAT (2006). Local labor systems are delimited on the basis of daily commuting of workers; when exchanges of goods and services are considered, the province is a more appropriate geographical unit. This is why we consider the province as the appropriate territorial level to characterize local production systems.

For the robustness checks we also consider two absolute measures of industry concentration at local level: i) number of firms in an industry in a province (*Industry firms)*; ii) number of employees in an industry in a province *(Industry employment)*. Both variables are considered in logarithmic terms.

---

[10] The entropy index does not have an upper bound. Its maximum level depends on n: $H_{\max} = \sum_{i=1}^{n} \frac{1}{n} \log_2(n) = n \frac{1}{n} \log_2(n) = \log_2(n)$. Given that we consider 50 industries the maximum value of the index is 3.91.

[11] Because Frenken et al. (2007) were interested in assessing the level of knowledge spillovers between industries, they consider v_{ij} as an index of the similarities between industries based on their input structure. This is intended to capture technological similarities among sectors.

3.4 Controls

As well as the agglomeration variables our econometric specifications include several controls to account for unobserved heterogeneity. These are:

(i) three geographic dummies: *North-West* (Piemonte, Lombardia, Valle d'Aosta, Liguria), *North-East-Center* (Trentino Alto-Adige, Veneto, Friuli Venezia Giulia, Emilia-Romagna, Toscana, Umbria, Marche, Lazio), *South* (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna). These dummies are introduced to capture differences between Italian areas in terms of infrastructures endowments, efficiency of the legal system, and other institutional features (Fan et al. 2009);

(ii) size of the local system (*LSsize*) captured by the provincial population;

(iii) size of the group captured by the number of controlled firms (*Group size*);

(iv) 24 industry dummies in the estimates for all business groups (see Table A.7), and 4 Pavitt sector dummies in the case of manufacturing groups (see Table A.6).

The Pavitt sector dummies account for the different technological regimes that are supposed to influence the level of asset specificity in input-output exchanges (Acemoglu et al. 2010; Cainelli and Iacobucci 2009).

The variables used in the econometric estimates are listed in Table 1.

Table 1 – Variables used in the econometric estimates

Variable	**Description**	**Year**
VI_f	Continuous measure of vertical integration of business groups	2001
$Spec_{i,k}$	Index of specialization of industry *i* in province *k*	1991, 2001
$Uvariety_k$	Index of unrelated variety in province *k*	1991, 2001
$VRvariety_k$	Index of vertically related variety in province *k*	1991, 2001
$LSsize_k$	Size of province *k* in terms of population	1991, 2001
Group size	Size of the group in terms number of firms controlled	2001
Industry $firms_{i,k}$	Number of firms in industry *i* per km^2 in province *k*	1991
Industry $employmeent_{i,k}$	Number of emplpoyees in industry *i* per km^2 in province *k*	1991
North-West North-East-Center South	Geographic dummies	
$Nace_i$	Industry dummies (see Table A.7 in the appendix)	
Pavitt	Pavitt sectors dummies (see Table A.6 in the appendix)	

3.5 Econometric modelling

In our econometric investigation, we estimate a Tobit model as follows:

$$y_f^* = X_f\beta + \varepsilon_f$$

where $\varepsilon_f \approx N\left(0,\sigma_f^2\right)$ and X_f denotes the independent variables defined in section 3.3. Table 2 reports some descriptive statistics for these variables. y* is a latent variable that is observed for values greater than τ and censored otherwise. In our case $\tau = 0$: that is, data are censored at 0 since the degree of vertical integration is not observed for all business groups, given the presence of groups that are not vertically integrated. In the presence of censored observations, OLS estimates of β would be inconsistent. The observed y_f – our measure of vertical integration – is defined by the following measurement equation:

$$y_f = y_f^* \text{ if } y_f^* > 0$$

$$y_f = 0 \text{ if } y_f^* \leq 0$$

As is well known, the log-likelihood function for the Tobit model when $\tau = 0$ is given by:

$$\ln L = \sum_{f=1}^{N}\left\{d_f\left(-\ln\sigma + \ln\phi\left(\frac{y_f - X_f\beta}{\sigma}\right)\right) + \left(1-d_f\right)\ln\left(1-\Phi\left(\frac{X_f\beta}{\sigma}\right)\right)\right\}$$

This overall log-likelihood consists of two parts. The first corresponds to the classical regression for the uncensored observations, and the second corresponds to the relevant probabilities that an observation is censored.

In our estimates we assume that agglomeration affects vertical integration. However, the reverse can apply that vertical integration can affect agglomeration, generating a classical reverse causality problem. Helsley & Strange (2007), in fact, suggest that "the effect of agglomeration on opportunism and the organization of production is a force that will lead firms to agglomerate" (Helsley and Strange 2007, p. 57). The presence of (potential) endogeneity – that is, one or more explanatory variables correlated with the error term – can generate biased and inconsistent estimates of the coefficients under investigation. We control for this assuming, as is usual in the regional economics literature (Glaeser et al. 1992; Henderson, Kuncoro, and Turner 1995), a time lag between vertical integration and the agglomeration variables. These latter refer to 1991. Also, the variable for local system size is lagged and refers to 1991. Finally, we also consider agglomeration variables and system local size at 2001 to test the robustness of our results (see section 5).

To assess the overall goodness of fit of our econometric specifications we use two different measures: (i) the McKelvey and Zavoina R_2 and the Bayesian information criterion (BIC) proposed by Raftery (Long and Freese 2000). The BIC is generally used to measure and compare the overall fit among nested and non-nested models. Following Long and Freese (2000) BIC is defined as $BIC = D(M) - df \ln N$ where *df* is the degrees of freedom associated with the deviance. This information measure can be interpreted as follows: (i) in absolute terms, the more negative the value of this measure, the better the fit; (ii) in comparative terms, the difference between the BIC values obtained from two different specifications (BIC_1 and BIC_2) indicates which model is more likely to have generated the observed data. In particular, if BIC_1-BIC_2<0 the first specification is preferred to the second, and if BIC_1-BIC_2>0 the second specification is preferred to the first. Using this information criterion we detect that the specifications that include agglomeration variables are preferred to the base line specification, which considers only controls; in several cases the specifications that include specialization, and specialization and related-variety, perform better in terms of goodness of fit.

Table 2 – Descriptive statistics

Variable	*Unit*	*Year*	*Obs.*	*Mean*	*Std.dev.*	*Min*	*Max*
VI_f	Group	2001	24,633	.0149	.0211	0	0.214
Group $size_f$	Group	2001	24,633	3.5	9.9	2	605
LS $size_k$	Province	1991	103	551,340	609,933	91,942	3,761.067
$Uvariety_k$	Province	1991	103	4.726	0.205	3.745	4.958
$VRvariety_k$	Province	1991	103	.062	.0145	.052	.174
$Spec_{i,k}$	Province	1991	103	1.413	1.536	0	16.737

4. EMPIRICAL RESULTS

We perform Tobit estimates for the whole population of Italian business groups, and for the manufacturing groups only. The results for the population of business groups are presented in Table 3. We estimate several econometric specifications by introducing the variables, one at a time, to capture specialization, unrelated variety, and the vertical relatedness of economic activities at the local level.[12]

The coefficient of the specialization variable has the expected sign and is statistically significant in all the specifications. This confirms the results of other studies: that is, the role of geographic concentration of activities in reducing the level of vertical integration of firms. Since we use a measure of inter-industry integration, this confirms that the agglomeration of firms belonging to the same industry also promotes the location of auxiliary and complementary industries thus reducing the need to control them.

[12] We do not include all three agglomerative variables simultaneously because they are highly collinear (see Table A.3).

The coefficient of unrelated variety has a positive sign. This unexpected result is due to the nature of the unrelated variety indicator which cannot account for vertical relatedness between industries. On the other hand, the coefficient of the vertically related variety indicator has the expected (negative) sign and is highly statistically significant in the two specifications that include this variable. This shows that it is not variety *per se* that matters for vertical integration, but the presence of industries that are vertically related in terms of input-output exchanges.

Table 3 – Determinants of vertical integration. Tobin estimates for all business groups in Italy

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$Spec_{i,k,1991}$	…	-0.0005** [-2.45]	…	…	-0.0004** [-2.17]	-0.0004*** [-4.28]
$Uvariety_{k,1991}$	…	…	0.004*** [2.93]	…	0.003*** [2.67]	…
$VRvariety_{k,1991}$	…	…	…	-0.058*** [-4.86]	…	-0.042*** [-3.28]
$\ln(\text{LS size}_{k,1991})$	-0.0007*** [-2.73]	-0.0008*** [-3.03]	-0.001*** [-5.54]	-0.001*** [-4.79]	-0.001*** [-4.84]	-0.0009*** [-4.28]
$\ln(\text{Group size}_{f,2001})$	0.012*** [17.77]	0.012*** [17.63]	0.012*** [17.91]	0.012*** [17.89]	0.012*** [17.72]	0.012*** [17.70]
North_West	0.001** [2.51]	0.001*** [2.99]	0.00002 [0.03]	0.0006 [1.16]	0.0005 [0.89]	0.001** [1.99]
North_East_Center	0.0002 [0.43]	0.0004 [0.70]	-0.0008 [-1.41]	-0.0004 [-0.87]	-0.0004 [-0.82]	-0.0001 [-0.30]
South	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Industry dummy[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	24,633	24,633	24,633	24,633	24,633	24,633
Uncensored Obs.	15,811	15,811	15,811	15,811	15,811	15,811
Censored Obs.	8,882	8,882	8,882	8,882	8,882	8,882
Log pseudolikelihood	27311.5	27319.2	27314.7	27318.7	27319.6	27322.4
McKelvey-Zavoina R^2	0.090	0.091	0.091	0.091	0.091	0.091
BIC	-303434.9	-303440.2	-303420.3	-303439.3	-303419.8	-303436.6

*** Significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at province level (103 units) with *t* statistics in parentheses

The variable for local system size shows a negative and statistically significant relation to vertical integration. This variable captures the size of the local market which will influence the possibility of market transactions and will reduce the need to integrate activities within the firm.

Group size has the expected positive sign. The larger the size of the firms the greater the scope for vertical integration because larger firms are more able fully to exploit the economies of scale in different stages of the production process.

We obtain similar results when we consider the sub-population of manufacturing groups (see Table 4).

Table 4 - Determinants of vertical integration. Tobit estimates for manufacturing groups.

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$Spec_{i,k,1991}$	…	-0.0008*** [-4.55]	…	…	-0.0007*** [-3.34]	-0.0006*** [-2.69]
$Uvariety_{k,1991}$	…	…	0.008** [2.56]	…	0.003 [1.38]	…
$VRvariety_{k,1991}$	…	…	…	-0.090*** [-3.84]	…	-0.036 [-1.16]
$\ln(LS\ size_{k,1991})$	-0.0002 [-0.42]	-0.0007 [-1.11]	-0.001* [-1.88]	-0.0009 [-1.49]	-0.001* [-1.73]	-0.0009 [-1.43]
$\ln(Group\ size_{f,2001})$	0.009*** [13.37]	0.009*** [13.27]	0.009*** [13.30]	0.009*** [13.28]	0.009*** [13.24]	0.009*** [13.21]
North_West	-0.0001 [-0.10]	0.001 [0.75]	-0.002 [-1.24]	-0.0009 [-0.64]	-0.00006 [-0.04]	0.0005 [0.37]
North_East_Center	-0.001 [-1.16]	-0.0005 [-0.46]	-0.003** [-2.42]	-0.002** [-2.06]	-0.001 [-1.20]	-0.001 [-0.89]
South	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Supplier dominated	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Scale intensive	0.006*** [6.03]	0.005*** [5.34]	0.005** [6.11]	0.005*** [6.00]	0.005*** [5.39]	0.005*** [5.37]
Science based	0.001 [0.97]	0.0006 [0.58]	0.001 [0.96]	0.001 [0.91]	0.0006 [0.64]	0.0006 [0.62]
Specialized supplier	0.004*** [4.23]	0.003*** [3.57]	0.003*** [4.10]	0.003*** [4.06]	0.003*** [3.55]	0.003*** [3.57]
N. Obs.	8,097	8,097	8,097	8,097	8,097	8,097
Uncensored Obs.	5,786	5,786	5,786	5,786	5,786	5,786
Censored Obs.	2,311	2,311	2,311	2,311	2,311	2,311
Log pseudolikelihood	9909.7	9922.8	9913.3	9916.9	9921.2	9923.6
McKelvey-Zavoina R^2	0.030	0.033	0.031	0.032	0.034	0.034
BIC	-92605.4	-92622.8	-92593.4	-92610.9	-92600.4	-92615.1

*** significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses

In this case also, the coefficient of specialization shows the expected sign and is statistically significant in all the specifications. However, the results for manufacturing groups show some differences with the results for the entire population of groups. The coefficient of vertically related variety is statistically significant when introduced in isolation, but loses its significance when it is included jointly with specialization. Also, the variable capturing local system size is not significant. The differences between the whole population and the manufacturing groups can be attributed to the fact that in the case of manufacturing activities, the agglomeration of firms operating in the same industry is more relevant in vertical integration decisions than industry variety at the local level. Also, the size of the local system, which can be considered a proxy for local market size, is more influential in the case of services than manufacturing activities. For this reason it plays a role in the vertical integration decisions of groups in the service sector but not in the case of manufacturing groups.

In the estimates for manufacturing groups we include industry dummies based on Pavitt's (1984) taxonomy which refers to the innovation patterns in manufacturing sectors. This taxonomy captures the influence of technological regimes on vertical integration decisions (Acemoglu et al. 2004; Cainelli and Iacobucci 2009). As expected, groups operating in the scale intensive and specialized supplier sectors tend to be more vertically integrated than groups in supplier dominated (i.e. traditional) sectors. This is because of the higher level of transaction specific investments characterizing the input-output relations in these sectors.

When interpreting the results for the manufacturing groups, we need to take account of the geographical heterogeneity of the Italian manufacturing industry. The literature on the Italian case, generally identifies three macro-areas (Bagnasco 1977): North-West, North-East and Centre. and the South. The North-West – the so-called 'Industrial Triangle', which includes Piemonte, Lombardia and Liguria, was the first area to experience industrialization, in the late 19th and early 20th centuries. This area continues to be characterized by the presence of local systems based on large and medium sized firms, mostly operating in scale intensive sectors (ISTAT 2005). The North-East and Center area – the so-called NEC or "Third Italy" - includes Friuli Venezia-Giulia, Veneto, Emilia Romagna, Toscana, Umbria and Marche (Fuà 1983). In the 1970s and 1980s, these regions experienced rapid industrialization based mostly on small firm start ups in agglomerated areas (industrial districts). The prevalence of local systems based on small firms and the spread of activities over the territory reduced the need for population migrations and avoided the urban congestion often associated with the industrialization process. Also, the spatial agglomeration of small firms provided efficiency gains and promoted innovation. From the mid 1980s, several researchers, in Italy and other countries, focused on analyzing this phenomenon (Storper 1993; Becattini 1989; Pyke, Becattini, and Sengenberger 1992; Brusco 1986; Piore and Sabel 1984; Bellandi, De Propris, and Becattini 2009) of the advantages of spatial agglomeration of specialized firms and the role of interactions between economic, social and cultural factors at local level (Harrison 1992; Wilkinson and You 1994; Harrison 1994; Dei Ottati 1994). During the 1990s most local systems in the North-East and Center experienced a process of concentration of activities and saw the emergence of medium sized firms better able to organize and control production activities within the district (Iacobucci 2004; Cainelli and Zoboli 2004; Brioschi, Brioschi, and Cainelli 2002). Despite these changes, the regions of the 'Third Italy' are still characterized by a high level of industrial activity based on specialized local systems in which small and medium firms predominate (ISTAT 2005).

It should be noted also that there is a greater disparity in the size of local systems in the North West area because of the presence of large metropolitan areas (Milan, Turin and Genoa being the largest), than in the North-East and Center regions which are characterized by the presence of small and medium sized towns, with the result that there is less variance in the size of local systems.

The Southern regions – the so-called Mezzogiorno – is the least developed area of the country. Despite attempts in the 1960s and 1970s to encourage investment in this region, most of the area is characterized by a scant presence of manufacturing activities. The few industrial local systems that exist are characterized by the presence of large firms, mostly controlled by Italian and foreign groups. This is the result of the investment attraction policies implemented by Italian governments during the second half of the 20th century aimed at promoting the industrial development of this part of the country. These policies failed to stimulate a subsequent process of endogenous industrial development.

The geographical distribution of business groups within these three macro-areas is presented in Table A.1. Only a small percentage of business groups is located in the South (13.4%); the remaining groups are split between the other two macro areas. As expected, the average size of groups in the North-West regions is higher than the average size of groups located in the North-East and Center regions. In order to account for this spatial heterogeneity, we replicate the empirical analysis by splitting the sample for these macro-areas. Also, we exclude the Southern regions since the number of groups located in this area is so limited. Disaggregating our analysis according to these macro-areas provides quite interesting results.

In the case of the North-West regions, agglomeration forces do not play a significant role in affecting the vertical integration choices of firms (see Table 5). This is consistent with the prevalence of larger and older firms in the area and the idea that the strategic choices of these firms is not influenced by local level agglomeration forces. The coefficients of the control variables show the expected signs.

In the case of the 'Third Italy', the results obtained confirm the results for the whole country (see Table 6). The role of agglomeration forces in influencing the vertical integration choices of firms is consistent with the feature of this area, which is characterized by specialized local systems based on small and medium sized firms (industrial districts).

Table 5 - Determinants of vertical integration. All business groups located in North-West.

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$Spec_{i,k,1991}$	…	0.0001 [0.22]	…	…	0.0001 [0.20]	0.0001 [0.29]
$Uvariety_{k,1991}$	…	…	-0.0007 [-0.32]	…	-0.0005 [-0.27]	…
$VRvariety_{k,1991}$	…	…	…	-0.012 [-0.49]	…	-0.017 [-0.71]
$\ln(LS\,size_{k,1991})$	-0.0005** [-2.17]	-0.0005** [-2.19]	-0.0004 [-1.28]	-0.0006** [-2.26]	-0.0004 [-1.38]	-0.0006** [-2.21]
$\ln(Group\,size_{f,2001})$	0.010*** [17.12]	0.010*** [16.94]	0.010*** [17.12]	0.010*** [17.12]	0.010*** [16.94]	0.010*** [16.94]
Industry dummy(a)	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	9,368	9,368	9,368	9,368	9,368	9,368
Uncensored Obs.	6,221	6,221	6,221	6,221	6,221	6,221
Censored Obs.	3,147	3,147	3,147	3,147	3,147	3,147
Log pseudolikelihood	10887.5	10887.6	10887.5	10887.5	10887.6	10887.6
McKelvey-Zavoina R^2	0.077	0.077	0.077	0.077	0.077	0.077
BIC	-107217.2	-107208.2	-107208.1	-107208.2	-107199.0	-107199.3

*** significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses

Table 6 – Determinants of vertical integration. All business groups, Third Italy.

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$Spec_{i,k,1991}$	…	-0.0008*** [-4.34]	…	…	-0.0006*** [-3.70]	-0.0006*** [-3.38]
$Uvariety_{k,1991}$	…	…	0.007*** [5.52]	…	0.005*** [4.50]	…
$VRvariety_{k,1991}$	…	…	…	-0.082*** [-7.17]	…	-0.055*** [-4.43]
$\ln(LS\,size_{k,1991})$	-0.001** [-2.37]	-0.001*** [-3.39]	-0.001*** [-6.62]	-0.001*** [-5.31]	-0.001*** [-6.43]	-0.001*** [-5.21]
$\ln(Group\,size_{f,2001})$	0.012*** [25.92]	0.012*** [25.75]	0.012*** [25.76]	0.012*** [25.90]	0.012*** [25.64]	0.012*** [25.78]
Industry dummy(a)	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	11,963	11,963	11,963	11,963	11,963	11,963
Uncensored Obs.	7,629	7,629	7,629	7,629	7,629	7,629
Censored Obs.	4,334	4,334	4,334	4,334	4,334	4,334
Log pseudolikelihood	13202.9	13214.4	13210.9	13213.6	13216.4	13218.5
McKelvey-Zavoina R^2	0.089	0.091	0.091	0.091	0.092	0.092
BIC	-138498.7	-138512.4	-138494.8	-138510.5	-138496.3	-138510.9

*** Significant at 1%; ** significant at 5%; * significant at 10%; regressions include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses.

Both the variables capturing specialization and vertically related variety at the local level have a statistically significant effect on the degree of vertical integration of firms. They are statistically significant when introduced separately (models 2 and 4) and jointly (model 6). Note also that the coefficient of the variable capturing unrelated variety shows a positive sign. This reinforces the idea that it is not variety per se that matters for vertical integration, but the presence of vertically related industries at the local level. The results obtained for the overall population of business groups are confirmed also for manufacturing groups. In the case of manufacturing groups located in the northern part of the country the variables capturing local level specialization, unrelated variety and vertically related variety are not statistically significant in terms of influencing the degree of inter-industry vertical integration (Table 7).

Table 7 - Determinants of vertical integration. Manufacturing groups – North West.

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$\text{Spec}_{i,k,1991}$	…	-0.0003 [-0.60]	…	…	-0.0003 [-0.67]	-0.0003 [-0.57]
$\text{Uvariety}_{k,1991}$	…	…	-0.001 [-0.38]	….	-0.003 [-0.87]	…
$\text{VRvariety}_{k,1991}$	…	…	…	-0.011 [-0.33]	…	0.007 [0.21]
$\ln(\text{LS size}_{k,1991})$	-0.001** [-2.31]	-0.001** [-2.37]	-0.001 [-1.39]	-0.001** [-2.06]	-0.001 [-1.34]	-0.001** [-1.97]
$\ln(\text{Group size}_{f,2001})$	0.008*** [8.02]	0.008*** [7.98]	0.008*** [8.02]	0.008*** [8.00]	0.008*** [7.99]	0.008*** [7.94]
Supplier dominated	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Scale intensive	0.006*** [4.37]	0.006*** [4.38]	0.006*** [4.32]	0.006*** [4.33]	0.006*** [4.38]	0.006*** [4.44]
Science based	0.001 [1.10]	0.001 [1.14]	0.001 [1.11]	0.001 [1.10]	0.001 [1.15]	0.001 [1.14]
Specialized supplier	0.005*** [3.59]	0.005*** [3.43]	0.005*** [3.57]	0.005*** [3.55]	0.005*** [3.44]	0.005*** [3.43]
N. Obs.	3,351	3,351	3,351	3,351	3,351	3,351
Uncensored Obs.	2,437	2,437	2,437	2,437	2,437	2,437
Censored Obs.	914	914	914	914	914	914
Log pseudolikelihood	4198.7	4199.2	4198.8	4198.7	4199.4	4199.2
McKelvey-Zavoina R^2	0.028	0.028	0.028	0.028	0.028	0.028
BIC	-35540.7	-35533.6	-35532.7	-35532.6	-35525.9	-35525.5

*** significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses

Given that in this area the local systems are characterized by the presence of older and larger groups, these results support the idea that as firms grow, their strategic choices are less influenced by the local environment. However, in the case of manufacturing groups located in the North-East and Center (Third Italy), the variables capturing local level specialization and vertically related variety are both statistically significant in influencing the degree of inter-industry integration of manufacturing groups (Table 8).

This result is particularly interesting as the North-East and Center is the part of Italy where manufacturing activities are organized mainly in specialized local systems (industrial districts). Among the factors influencing the organization of firms in industrial districts are the high level of specialization along the production chain and the use of market transactions for intermediate goods and services (Bellandi, De Propris, and Becattini 2009). The results of our analysis are further confirmation of the influencing role of agglomeration forces in these areas in firms' decisions about inter-industry vertical integration.

Table 8 – Determinants of vertical integration. Manufacturing groups – Third Italy

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$\text{Spec}_{i,k.1991}$	…	-0.001*** [-6.77]	...	…	-0.0007*** [-2.93]	-0.0008*** [-3.01]
$\text{Uvariety}_{k,1991}$	…	…	0.014*** [9.49]	…	0.008*** [3.08]	…
$\text{VRvariety}_{k,1991}$	…	…	…	-0.125*** [-7.85]	…	-0.056** [-2.26]
$\ln(\text{LS size}_{k,1991})$	0.0009 [0.64]	-0.0002 [-0.21]	-0.0008 [-1.19]	-0.0002 [-0.28]	-0.0009 [-1.15]	-0.0005 [-0.51]
$\ln(\text{Group size}_{f,2001})$	0.009*** [14.54]	0.009*** [14.50]	0.009*** [13.82]	0.009*** [14.22]	0.009*** [14.09]	0.009*** [14.42]
Supplier dominated	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Scale intensive	0.005*** [3.27]	0.003** [2.54]	0.004*** [3.22]	0.004*** [3.10]	0.003*** [2.72]	0.003*** [2.65]
Science based	-0.0002 [-0.14]	-0.001 [-0.80]	-0.0002 [-0.17]	-0.0004 [-0.26]	-0.001 [-0.63]	-0.001 [-0.71]
Specialized supplier	0.003** [1.99]	0.001 [1.20]	0.002 [1.51]	0.002 [1.58]	0.001 [1.13]	0.001 [1.18]
N. Obs.	3,917	3,917	3,917	3,917	3,917	3,917
Uncensored Obs.	2,767	2,767	2,767	2,767	2,767	2,767
Censored Obs.	1,150	1,150	1,150	1,150	1,150	1,150
Log pseudolikelihood	4749.7	4764.0	4765.5	4759.9	4764.3	4765.4
McKelvey-Zavoina R^2	0.028	0.037	0.035	0.034	0.039	0.038
BIC	-41847.2	-41867.6	-41848.9	-41851.2	-41850.5	-41862.0

*** significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses

Similar to the case of the whole population of groups, the variable capturing vertically related variety matters, while the variable capturing related variety has the opposite sign for effect on the degree of vertical integration. Note also that in the case of manufacturing groups located in the North-West the size of the local system is significant in influencing the degree of vertical integration, but this is not the case for groups located in the Third Italy. This could be attributable to the fact that in the North-West there is wide variability in the size of local systems due to the presence of large urban areas. The North-Eastern and Center regions are characterized by the absence of large urban areas, resulting in reduced variability in the size of local systems.

5. ROBUSTENSS CHECKS

We perform the estimates with different specifications in order to check the robustness of our results. Specifically, all estimates are replicated using the explanatory variables calculated for the same year of the independent variable (2001) and using different proxies for local level specialization.

Table 9 reports the results of the estimates for the overall population of business groups obtained by considering specialization and variety indicators at 2001. The results are similar to those obtained from the basic estimates. Moreover, the coefficients of the specialization and vertically related variety indicators show higher degrees of statistical significance.

Table 10 presents the result of the estimates obtained using different proxies for specialization at local level. We consider two new measures of absolute concentration: i) number of firms in an industry in a province (*Industry firms)*; ii) number of employees in an industry in a province *(Industry employment)*. The variables Industry firms and Industry employment are considered in logarithmic term and calculated for 1991 (Table 9), and 2001: the year of the dependent variable (Table 10).

The econometric results do not change using these measures of absolute concentration at the local level, which confirms the robustness of the relation between vertically related variety and vertical integration. Our results for manufacturing groups and disaggregation for geographical areas are also robust.

Table 9 – Determinants of vertical integration. Tobin estimates for all business groups in Italy

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$\text{Spec}_{i,k,2001}$	…	-0.0005***	…	…	-0.0004***	-0.0004***
		[-3.59]			[-3.21]	[-2.82]
$\text{Uvariety}_{k,2001}$	…	…	0.005***	…	0.004***	…
			[3.86]		[3.49]	
$\text{VRvariety}_{k,2001}$	…	…	…	-0.092***	…	-0.063***
				[-5.97]		[-3.63]
$\ln(\text{LS size}_{k,2001})$	-0.0007***	-0.0008***	-0.0009***	-0.0008***	-0.0009***	-0.0008***
	[-2.69]	[-3.14]	[-5.92]	[-4.27]	[-5.52]	[-4.10]
$\ln(\text{Group size}_{f,2001})$	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***
	[17.77]	[17.63]	[17.93]	[17.90]	[17.73]	[17.70]
North_West	0.001**	0.001***	-0.0001	0.0004	0.0003	0.0008*
	[2.51]	[2.97]	[-0.28]	[0.75]	[0.63]	[1.73]
North_East_Center	0.0002	0.0004	-0.001*	-0.0007	-0.0005	-0.0002
	[0.44]	[0.74]	[-1.78]	[-1.28]	[-1.08]	[-0.49]
South	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Industry dummy[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	24,633	24,633	24,633	24,633	24,633	24,633
Uncensored Obs.	15,811	15,811	15,811	15,811	15,811	15,811
Censored Obs.	8,882	8,882	8,882	8,882	8,882	8,882
Log pseudolikelihood	27311.4	27319.6	27316.7	27320.3	27321.1	27323.3
McKelvey-Zavoina R^2	0.090	0.091	0.091	0.091	0.091	0.091
BIC	-303434.7	-303430.1	-303424.2	-303442.6	-303411.8	-303427.2

*** Significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses.

Table 10 – Determinants of vertical integration; all business groups

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$\ln(\text{Industry firms}_{i,k,1991})$	-0.0007** [-2.12]	-0.001*** [-4.63]	-0.0009*** [-4.07]	…	…	…
$\ln(\text{Industry employment}_{i,k,1991})$	…	…	…	-0.0001 [-0.61]	-0.0005* [-1.81]	-0.0009*** [-4.07]
$\text{Uvariety}_{k,1991}$	…	0.004*** [2.94]	…	…	0.003* [1.78]	…
$\text{VRvariety}_{k,1991}$	…	…	-0.057*** [-4.84]	…	…	-0.057*** [-4.87]
$\ln(\text{Group size}_{f,2001})$	0.012*** [17.77]	0.012*** [17.90]	0.012*** [17.88]	0.012*** [18.15]	0.012*** [18.21]	0.012*** [17.88]
North_West	0.001*** [2.76]	0.0004 [0.60]	0.0009* [1.76]	0.001** [2.06]	0.0002 [0.32]	0.0009* [1.78]
North_East_Center	0.0005 [0.93]	-0.0004 [-0.71]	-0.00008 [-0.14]	0.0003 [0.51]	-0.0004 [-0.55]	-0.00008 [-0.14]
South	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Industry dummy[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	24,633	24,633	24,633	24,633	24,633	24,633
Uncensored Obs.	15,811	15,811	15,811	15,811	15,811	15,811
Censored Obs.	8,882	8,882	8,882	8,882	8,882	8,882
Log pseudolikelihood	27308.1	27314.0	27315.0	27305.7	27305.5	27315.0
McKelvey-Zavoina R^2	0.090	0.091	0.091	0.090	0.091	0.091
BIC	-303417.1	-303418.8	-303420.9	-303423.3	-303418.8	-303420.9

*** Significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses.

Table 11 – Determinants of vertical integration; all business groups

ESTIMATION METHOD: TOBIT	[1]	[2]	[3]	[4]	[5]	[6]
$\ln(\text{Industry firms}_{i,k,2001})$	-0.0007** [-2.21]	-0.001*** [-4.76]	- 0.0008*** [-3.31]	…	…	…
$\ln(\text{Industry employment}_{i,k,2001})$	…	…	…	-0.0001 [-0.40]	- 0.0007*** [-3.11]	-0.0003* [-1.78]
$\text{Uvariety}_{k,2001}$	…	0.006*** [4.03]	…	…	0.007*** [3.44]	…
$\text{VRvariety}_{k,2001}$	…	…	-0.093*** [-5.86]	…	…	- 0.096*** [-5.26]
$\ln(\text{Group size}_{f,2001})$	0.012*** [17.79]	0.012*** [17.94]	0.012*** [17.96]	0.012*** [18.23]	0.012*** [18.12]	0.012*** [18.22]
North_West	0.001*** [2.68]	0.00008 [0.12]	0.0006 [1.18]	0.001* [1.91]	-0.0002 [-0.34]	0.0003 [0.56]
North_East_Center	0.0005 [0.87]	-0.0007 [-1.30]	-0.0004 [-0.78]	0.0003 [0.52]	-0.001 [-1.42]	-0.0006 [-0.90]
South	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Industry dummy[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
N. Obs.	24,633	24,633	24,633	24,633	24,633	24,633
Uncensored Obs.	15,811	15,811	15,811	15,811	15,811	15,811
Censored Obs.	8,882	8,882	8,882	8,882	8,882	8,882
Log pseudolikelihood	27307.2	27315.6	27316.3	27305.5	27311.6	27314.6
McKelvey-Zavoina R^2	0.090	0.091	0.091	0.090	0.090	0.090
BIC	-303415.2	-303422.0	-303423.3	- 303423.0	-303413.9	- 303431.0

*** Significant at 1%; ** significant at 5%; * significant at 10%; regressions also include a constant term; standard errors are clustered at the province level (103 units) with *t* statistics in parentheses

6. CONCLUSIONS

This study used a large dataset of Italian business groups to investigate the relationship between agglomeration and inter-industry vertical integration. We define vertical integration as the ownership and control of activities along the production chain. Our unit of analysis to measure the degree of vertical integration is the business group – that is, the set of firms controlled by the same owner. The choice of business group as the unit of analysis is appropriate for this study because control over the different phases in the production chain is often achieved through the ownership of legally independent companies rather than integration of activities within the same legal unit. To account for agglomeration forces and other characteristics of the local system, our geographical unit is the Italian province.

We investigated the influence of agglomeration forces on the vertical integration choices of firms by analyzing the effects of specialization and variety in the local system. We used two different indicators of variety: the first is an indicator of unrelated variety and captures the presence of different industries within the same geographical area; the second is a more specific indicator of industry variety that captures the presence of vertically related industries.

Our results support the hypothesis that the firm's choice to integrate activities along the production chain is influenced by the geographic concentration of activities.

For industry variety at the local level, our findings show that it is not variety per se that matters for vertical integration choices, but the degree of local level vertical relatedness of production activities. For instance, a higher level of vertical related variety significantly reduces the need for firms to integrate activities along the production chain. This means that spatial proximity matters for vertical integration choices, because the distance over which firms are able to provide intermediate inputs significantly influences their choice to control of the production of intermediate inputs rather than acquire them through market transactions. In line with the literature, we found that the degree of specialization of the local system reduces the level of vertical integration; this confirms the effect of agglomeration forces in reducing opportunism at the local level and inducing firms to rely on market transactions. The concentration of specialized activities in a specific area is likely to attract firms engaged in complementary activities thus favoring the possibility of inter-industry outsourcing and reducing the need to control activities along the production chain. Therefore, specialization matters in the case of infra-industry vertical integration, as shown by previous studies, and also in the case of inter-industry integration.

Our results show also that the importance of agglomeration forces is not homogeneous across the country and depends on the characteristics of firms and local systems. In the North-West regions, which are characterized by the presence of large firms and established industrial urban areas, local system level specialization and variety do not significantly influence the degree of vertical integration. However, in the North-Eastern and Center or Third Italy, which are characterized by the presence of small firms organized in specialized systems or industrial districts both specialization and vertically related variety are significant for firms' decisions to control activities along the production chain.

The novelty of our study lies in its consideration of inter-industry vertical integration and the business group as a unit of analysis. We provide a new index of vertically related variety to measure the presence of vertically related industries at the local level.

Overall, our results confirm the results of previous studies on the role of agglomeration in influencing the degree of firms' vertical integration. Our study shows that in analyzing the role of variety at the local level it is necessary to take account of the type of variety relevant specifically to the phenomenon under investigation. In the case of vertical integration it is the presence of industries with potential input-output relations that matters rather than variety per se.

References

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti. 2004. Vertical integration and technology. Theory and evidence. *Nber Working Paper Series* (10997):1-46.

———. 2010. Vertical Integration and Technology: Theory and Evidence. *Journal of the European Economic Association* 8 (5):989-1033.

Adelman, M. A. 1955. Concept and Statistical Measurement of Vertical Integration. In *Business Concentration and Price Policy*, edited by G. J. Stigler. Princeton: Princeton University Press.

Audretsch, David B., and M. P. Feldman. 1996. R&D spillovers and the geography of innovation and production. *American Economic Review* 86:630-640.

Bagnasco, A. 1977. *Tre Italie. La problematica territoriale nello sviluppo italiano*. Bologna: IL Mulino.

Becattini, G. 1989. Sectors and/or Districts: Some Remark on the Conceptual Foundation of Industrial Economics. In *Small Firms and Industrial Districts in Italy*, edited by E. Goodman, J. Bamford and P. Saynor. London: Routledge.

Bellandi, Marco, Lisa De Propris, and Giacomo Becattini. 2009. *A handbook of Industrial Districts*. Cheltenham: Edward Elgar Publishing.

Bishop, Paul, and Peter Gripaios. 2010. Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies* 44 (4):443 - 454.

Boschma, Ron. 2005. Editorial: Role of Proximity in Interaction and Performance: Conceptual and Empirical Challenges. *Regional Studies* 39 (1):41 - 45.

Boschma, Ron, Rikard Eriksson, and Urban Lindgren. 2009. How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography* 9 (2):169-190.

Boschma, Ron, and Koen Frenken. 2009. Technological relatedness and regional branching. In *Papers in Evolutionary Economic Geography*. Utrecht: Utrecht University.

Boschma, Ron, and Simona Iammarino. 2009. Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography* 85 (3):289-311.

Boschma, Ron, A. Minondo, and M. Navarro. 2010. Related variety and regional growth in Spain. In *Papers in Evolutionary Economic Geography*. Utrecht: Urban & Regional Research Centre Utrecht

Boschma, Ron, and Rik Wenting. 2007. The spatial evolution of the British automobile industry: Does location matter? *Industrial and Corporate Change* 16 (2):213-238.

Brioschi, Francesco, Maria Sole Brioschi, and Giulio Cainelli. 2002. From the industrial district to the district group: an insight to the evolution of local capitalism in Italy. *Regional Studies* 36 (9):1037-1052.

Brookfield, Jonathan. 2008. Firm Clustering and Specialization: A Study of Taiwan's Machine Tool Industry. *Small Business Economics* 30 (4):405-422.

Brusco, S. 1986. Small Firms and Industrial Districts: The Experiece of Italy. In *New Firms and Regional Development Europe*, edited by i. W. F. (Ed.). London: Academic Press.

Bryce, David J., and Sidney G. Winter. 2009. A General Interindustry Relatedness Index. *Management Science* 55 (9):1570-1585.
Cainelli, Giulio, and Donato Iacobucci. 2007. *Agglomeration, Technology and Business Groups*. Cheltenham: Edward Elgar.
———. 2009. Do Agglomeration and Technology Affect Vertical Integration? Evidence from Italian Business Groups. *International Journal of the Economics of Business* 16 (3):305-322.
Cainelli, Giulio, and Roberto Zoboli, eds. 2004. *The Evolution of Industrial Districts. Changing Governance, Innovation and Internationalization of Local Capitalism in Italy*. Heidelberg: Physica-Verlag.
Dei Ottati, Gabi. 1994. Trust, interlinking transactions and credit in the industrial district. *Cambridge Journal of Economics* 18 (6):529-546.
Diez-Vial, Isabel, and Emilio Alvarez-Suescun. 2010. Geographical Agglomeration as an Alternative to Vertical Integration. *Review of Industrial Organization* 36 (4):373-389.
Duranton, Gilles, and Diego Puga. 2001. Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review* 91 (5):1454-1477.
Enright, Michael J. 1995. Organization and Coordination in Geographically Concentrated Industries. In *Coordination and Information: Historical Perspectives on the Organization of Enterprise*, edited by N. R. Lamoreaux and D. M. G. Raff. Chicago and London: The University of Chicago Press.
Eurostat. 2003. Recommendation Manual on Business Registers. Luxembourg.
Fan, Joseph P. H., Jun Huang, Randall Morck, and Bernard Yeung. 2009. Vertical integration, institutional determinants and impact: Evidence from China. *NBER Working Paper* (14650).
Fan, Joseph P. H., and Larry H. P. Lang. 2000. The Measurement of Relatedness: An Application to Corporate Diversification. *Journal of Business* 73 (4):629-660.
Frenken, Koen, Frank Van Oort, and Thijs Verburg. 2007. Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies* 41 (5):685 - 697.
Fuà, G. 1983. L'industrializzazione del Nord Est e del Centro. In *Fuà G., Zacchia C. (a cura di), Industrializzazione senza fratture*. Bologna: Il Mulino.
Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer. 1992. Growth in Cities. *Journal of Political Economy* 100 (6):1126.
Harrison, B. 1992. Industrial Districts: Old Wine in New Bottles? *Regional Studies* 26 (5):469-483.
———. 1994. The Italian Industrial Districts and the Crisis of the Cooperative Form: Part I. *European Planning Studies* 2(1):3-22.
Helsley, Robert W., and William C. Strange. 2007. Agglomeration, opportunism, and the organization of production. *Journal of Urban Economics* 62 (1):55-75.
Henderson, Vernon , Ari Kuncoro, and Matt Turner. 1995. Industrial Development in Cities. *Journal of Political Economy* 103 (5):1067-1090.
Hidalgo, C. A., B. Klinger, A.-L. Barabasi, and R. Hausmann. 2007. The Product Space Conditions the Development of Nations. *Science* 317 (5837):482-487.
Holmes, Thomas J. 1999. Localization of Industry and Vertical Disintegration. *Review of Economics and Statistics* 81 (2):314-325.
Iacobucci, Donato. 2004. Groups of small and medium-sized firms in industrial districts in Italy. In *The Evolution of Industrial Districts. Changing Governance, Innovation and Internationalization of Local Capitalism in Italy*, edited by G. Cainelli and R. Zoboli. Heidelberg: Physica-Verlag.
ISTAT. 2005. I sistemi locali del lavoro. Censimento 2001. Dati definitivi. Roma.
———. 2006. *Distretti industriali e sistemi locali del lavoro 2001*. Roma: ISTAT.
Jacobs, Jane. 1969. *The economy of cities*. New York,: Random House.
Li, Ben, and Yi Lu. 2009. Geographic concentration and vertical disintegration: Evidence from China. *Journal of Urban Economics* 65 (3):294-304.

Long, Scott J., and Jeremy Freese. 2000. Scalar Measures of Fit for Regression Models. *STATA Technical Bulletin* 10:34-40.
Los, B. 2000. The empirical performance of a new inter-industry technology spillover measure. In *Technology and knowledge: from the firm to innovation systems*, edited by P. Saviotti and B. Nooteboom. Cheltenham, UK ; Northampton, MA: Edward Elgar.
Mariotti, Sergio, and M. Mutinelli. 2005. *Italia multinazionale 2005. Le partecipazioni italiane all'estero ed estere in Italia*. Roma: Istituto Nazionale per il Commercio Estero.
Marshall, A. 1920. *Principles of Economics*. London: MacMillan and Co.
Neffke, Frank, Martin Henning, and Ron Boschma. 2009. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. In *Papers in Evolutionary Economic Geography*. Utrecht: Utrecht University.
Neffke, Frank, and Martin Svensson Henning. 2008. Revealed Relatedness: Mapping Industry Space. In *DRUID Working Paper No. 08-18*.
Nooteboom, Bart. 2000. *Learning and innovation in organizations and economies*. Oxford ; New York: Oxford University Press.
Nooteboom, Bart, Wim Van Haverbeke, Geert Duysters, Victor Gilsing, and Ad van den Oord. 2007. Optimal cognitive distance and absorptive capacity. *Research Policy* 36 (7):1016-1034.
Ono, Yukako. 2007. Market thickness and outsourcing services. *Regional Science and Urban Economics* 37 (2):220-238.
Pavitt, K. 1984. Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory. *Research Policy* 13:343-373.
Piore, M.J., and C.F. Sabel. 1984. *The Second Industrial Divide*. New York: Basic Books.
Porter, Michael E. 1998. Clusters and the New Economics of Competition. *Harvard Business Review* (November-December):77-90.
———. 1998. Clusters and the new economics of competition. *Harvard Business Review* (November-December).
———. 2003. The Economic Performance of Regions. *Regional Studies* 37 (6):549 - 578.
Pyke, F., G. Becattini, and W. Sengenberger. 1992. *Industrial Districts and inter-firm cooperation in Italy*. Geneva: International Institute for Labour Studies.
Storper, M. 1993. Regional 'Worlds' of Production: Learning and Innovation in the Technology Districts of France, Italy and USA. *Regional Studies* 27 (5):433-455.
Teece, D.J., R. Rumelt, G. Dosi, and S. Winter. 1994. Understanding Corporate Coherence. Theory and Evidence. *Journal of Economic Behavior and Organization* 23:1-30.
Wilkinson, F., and J. You. 1994. Competition and Cooperation: Toward understanding industrial districts. *Review of Political Economy* July.
Wood, Gavin A., and John B. Parr. 2005. Transaction Costs, Agglomeration Economies, and Industrial Location. *Growth and Change* 36 (1):1-15.

Table A.1 – Distribution of business groups by geographic areas

AREAS	Number of groups		Employees		Average size (employees per group)
	N.	*%*	*N.*	*%*	
North_West	9,368	38.0	1,882,676	47.5	201
North-East-Centre	11,963	48.6	1,842,063	46.5	154
South and Islands	3,302	13.4	235,734	5.0	71
Total	**24,633**	**100.0**	**3,960.473**	**100.0**	**161**

Table A.2 – Distribution of business groups by macro-sector

ATECO	Firms		Employees	
	N.	*%*	*N.*	*%*
Manufacturing	8,230	33.4	1,709.699	43.2
Services	16,403	66.6	2,250.774	56.8
Totale	**24,633**	**100.0**	**3,960.473**	**100.0**

Table A.3 – Correlation matrix – all business groups

Variables	[1]	[2]	[3]	[4]	[5]	[6]
[1]	1.000					
[2]	0.123	1.000				
[3]	0.004	0.057	1.000			
[4]	-0.001	0.023	-0.091	1.000		
[5]	0.042	0.067	0.495	-0.200	1.000	
[6]	-0.035	-0.045	-0.334	0.343	-0.900	1.000

[1] $VI_{f,2001}$

[2] $\ln(\text{Group size}_{f,2001})$

[3] $\ln(\text{LSsize}_{k,1991})$

[4] $\text{Spec}_{i,k,1991}$

[5] $\text{Uvariety}_{k,1991}$

[6] $\text{VRvariety}_{k,1991}$

Table A.4 – Correlation matrix –manufacturing groups

Variables	[1]	[2]	[3]	[4]	[5]	[6]
[1]	1.000					
[2]	0.057	1.000				
[3]	0.003	0.040	1.000			
[4]	-0.048	0.020	-0.183	1.000		
[5]	0.030	0.055	0.553	-0.386	1.000	
[6]	-0.035	-0.032	-0.364	0.532	-0.911	1.000

[1] $VI_{f,2001}$

[2] $\ln\left(\text{Group size}_{f,2001}\right)$

[3] $\ln\left(\text{LSsize}_{k,1991}\right)$

[4] $Spec_{i,k,1991}$

[5] $Uvariety_{k,1991}$

[6] $VRvariety_{k,1991}$

Table A.5 – Correlation matrix – all business groups

Variables	[1]	[2]	[3]	[4]	[5]	[6]
[1]	1.000					
[2]	0.976	1.000				
[3]	-0.200	-0.222	1.000			
[4]	-0.167	-0.184	0.900	1.000		
[5]	0.343	0.362	-0.900	-0.816	1.000	
[6]	0.315	0.328	-0.783	-0.883	0.906	1.000

[1] $Spec_{i,k,1991}$

[2] $Spec_{i,k,2001}$

[3] $Uvariety_{k,1991}$

[4] $Uvariety_{k,2001}$

[5] $VRvariety_{k,1991}$

[6] $VRvariety_{k,2001}$

Table A.6 – Correspondence between Ateco_1991 classification and Pavitt's sector

Description	*Ateco_1991*	*Pavitt's sector*
Food products and beverages	15	SD
Tabacco	16	SD
Textile	17	SD
Clothing and furs	18	SD
Leather and footwear	19	SD
Wood and cork	20	SD
Paper	21	SD
Publishing, graphic arts, and reproduction	22	SI
Manufacture of coke oven products, petroleum refining	23	SI
Chemistry	24	241+247 SI; 242+ 244 SB; 243 SS; 245+246 SD
Rubber and plastic	25	*SD*
Non-metallic mineral products	26	*SD*
Iron metallurgic products	27	*SD*
Non-iron metallurgic products	*28*	281+285 SI; 282+284 SS; 286+287 SS;
Mechanical machinery and equipment	29	291+292 SS; 293 SI; 296+297 SB
Office machines and computers	30	SB
Electrical material, machinery and components	31	SS
Radio apparatus, TV and communication	32	SS
Medical, precision and optical instruments	33	SB
Motor vehicles	34	SS
Other transport manufacturing	35	SS
Other manufacturing industries	36	SD

SB = Science based

SD = Suppliers dominated

SI = Scale intensive

SS = Specialized suppliers

Table A.7 – Classification codes of economic activities

Code	Description
CB	Other mining activity
DA	Food products, beverages and tobacco
DB	Textiles, wearing apparels, dressing and dyeing of fur
DC	Leather and leather products
DD	Lumber and wood products
DE	Pulp, paper and paper products, Printing and publishing
DF	Petroleum refining and related industries
DG	Chemicals and allied products
DH	Rubber and plastic products
DI	Non-metallic mineral products
DJ	Primary metal industries, Fabricated metal products
DK	Industrial machinery and home appliances
DL	Office machinery and computers, Electrical and electronic equipment, Measuring and controlling instruments
DM	Automobiles and components, Other transportation vehicles
DN	Other manufacturing industries, Refuse systems, scrap and waste materials
E	Energy (electric power and gas), Water supply industry
F	Construction
G	Wholesale and retail distribution
H	Hotels and restaurants
I	Railways and road transport, Sea, air and other transport services, postal services and telecommunications
J	Banking and financial intermediation, Insurance and pension funds
K	Business services
M	Education
N	Health and social work

Source: ISTAT.