Agglomeration, related variety and vertical integration

Giulio Cainelli
Donato Iacobucci

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. We are interested specifically in assessing the effect of specialization compared to variety, in local production systems, in these decisions. Using a dataset of 24,663 Italian business groups for 2001, we estimate Tobit models to investigate the influence of these forces on the degree of vertical integration. Following a methodology developed by Fan et al. (2009), we construct an index of vertical integration in business groups using information on the activities of firms belonging to the same group, and information from input-output tables on the exchanges between industries. Our evidence supports the hypothesis that vertical integration choices are influenced by agglomeration forces. Moreover, we find that the higher the variety in the local system the lower the need for firms to integrate activities since they can acquire intermediate goods and services from within the local system. Finally, we show that it is not variety ‘per se’ that matters for vertical integration choices, but the presence of technologically related activities.

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

JEL: L2, M2

Giulio Cainelli
Dipartimento di Scienze Economiche “Marco Fanno”, Università degli Studi di Padova
CERIS-CNR, Milan
giulio.cainelli@unipd.it

Donato Iacobucci
Dipartimento di Ingegneria Informatica, Gestionale e dell’Automazione
Università Politecnica delle Marche
iacobucci@univpm.it

c.MET Working paper 5/2010
November 2010

© 2010 by Giulio Cainelli and Donato Iacobucci. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
1. INTRODUCTION

Several recent studies focus on the role of spatial agglomeration in affecting firms’ organization and strategies. In particular, they highlight the relation between geographic concentration of production activities and vertical integration (Brookfield, 2008; Cainelli & Iacobucci, 2009; Diez-Vial & Alvarez-Suescun, 2010; Helsley & Strange, 2007; Holmes, 1999). The common hypothesis in 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, find a negative relationship between geographical concentration of production and vertical integration.

The present paper aims to contribute to this literature. We are interested in assessing the role of specialization as opposed to variety in local production systems, in influencing vertical integration decisions. Unlike previous works, we study inter-industry vertical integration since we consider control of all manufacturing and service activities by the same firm. Our unit of analysis is the business group, i.e. the set of legally autonomous firms that is owned and controlled by the same people. Also, we consider the geographic concentration of firms belonging to the same industry (specialization), as well as other characteristics of the local systems in which the firms are located, such as size (in terms of population), degree of industry variety (unrelated variety) and presence of technologically-related industries (related variety).

Using a dataset of 24,663 Italian business groups for the year 2001, we estimate Tobit models to investigate the influence of these forces on the degree of vertical integration. We take the business group as our unit of analysis since we associate vertical integration with the ownership and control of production activities with input-output linkages. We take the province as our geographical unit since most input-output exchanges for key firm inputs such as labor and capital, take place within provinces. Also, most groups are comprised of firms located in the same province (Cainelli & Iacobucci, 2007).

Following a methodology developed by Fan et al. (2009), we construct an index of vertical integration in business groups, using information on the activities of firms in the same group, and information from input-output tables on the exchanges between industries. Thus, this measure of vertical integration has two components: presence of vertically related industries within the same group; intensity of the exchanges between industries based on the input-output tables.
This paper makes three contributions to the empirical literature on the relationship between agglomeration and vertical integration. First, our evidence supports the hypothesis that vertical integration choices are influenced by agglomeration forces. Second, we find that the higher the variety of the local system the lower will be the need for firms to integrate activities since they can acquire intermediate goods and services within the local system. Third, we show that it is not variety ‘per se’ that matters for vertical integration choices, but the presence of technologically related activities (related variety).

The paper is organized as follows. Section 2 provides a review of the literature on the determinants of vertical integration with specific reference to the role of spatial agglomeration. Section 3 discusses the data and methodology. Section 4 presents the results of the empirical analysis and Section 5 presents the main conclusions.

2. RELATED LITERATURE

There is an abundant and growing literature on the role of space in influencing firm performance. One of the main questions addressed by this stream of literature is the different roles of specialization and variety in influencing the innovation propensity and economic performance of firms (Audretsch & Feldman, 1996; Gilles Duranton & Puga, 2004; Glaeser, Kallal, Scheinkman, & Shleifer, 1992; Porter, 1998; Rosenthal & Strange, 2003). Less attention has been paid to assessing the relations between these forces and the organizational choices made by firms, in this paper, called degree of vertical integration.

The spatial agglomeration literature stresses the importance of market-based relationships between firms located in specialized clusters, as a result of well developed intermediate markets and lower levels of transactions costs. The latter are generally explained as due to lower levels of opportunism among economic agents and reduced information asymmetry. The lower level of opportunism is explained by the homogeneity of clusters in terms of local institutions, “culture”, social capital, language, etc. (Helsley & Strange, 2007; Wood & Parr, 2005), while information asymmetries are reduced by spatial proximity, frequency of face-to-face contacts and local knowledge spillovers (Breschi & Lissoni, 2001). The literature on industrial districts emphasizes the role of social capital and trust in shaping the vertical relationships between independent agents, underlining the co-operative nature of these relationships (Becattini, 1992; Brusco, 1982; Dei Ottati, 1994). In emphasizing the role of co-operation, and absence of opportunistic behavior in vertical relations
between firms, this literature predicts that firms located in industrial clusters will show lower levels of vertical integration than similar non-agglomerated firms.

A few studies investigate the relations between agglomeration and vertical integration using large datasets and econometric methodologies (Brookfield, 2008; Cainelli & Iacobucci, 2009; Diez-Vial & Alvarez-Suescun, 2010; Holmes, 1999). Though referring to different countries and industries and using different econometric approaches, all these studies find a negative relationship between the spatial agglomeration of industries and the degree of vertical integration of firms.

Holmes (1999) studies the relation between agglomeration and vertical integration using plant level data for the U.S. manufacturing sector. To measure vertical integration he uses the Adelman index based on the ratio between purchased inputs and sales. This index captures both intra-industry and inter-industry integration since it refers to all the inputs purchased by firms. Geographic concentration of industries is measured using three different indicators: (i) employees working on the same activity; (ii) employees working on different activities within the same two-digit code; (iii) employees in different industries. Holmes (1999) finds that establishments in areas with high concentration of an industry show lower degrees of vertical integration. Brookfield (2008) investigates the relation between industry concentration and vertical integration based on the Taichung machine tool industrial district in Taiwan. He finds that firms located in the Taichung area are more specialized and less vertically integrated than firms located elsewhere in the island. Li and Lu (2009) replicate Holmes’s (1999) methodology using data on Chinese manufacturing firms. Geographic concentration of industrial activity is measured at the level of the Chinese provinces (31). This is a much larger geographical unit than used by other studies of spatial agglomeration. Li and Lu (2009) confirm the negative role of geographic concentration of industrial activities on the degree of vertical integration of firms. Cainelli & Iacobucci (2009) consider backward vertical integration on the basis of control of production activities rather than their presence in the same physical location (establishment). They use data on Italian business groups to analyze the role of technology in the controlling activities along the production chain, taking account of agglomeration forces. Agglomeration forces are associated with the location of groups within an industrial district. Consistent with Holmes (1999), they find that the role of agglomeration depends on the industry being considered. Finally, 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 measured on the basis of the presence of different meat processing stages in the same establishment. They find that the location of meat establishments within a 2.5 km radius reduces vertical integration.
A more recent strand in the literature introduces the concept of “related variety” at local level, as one of the keys to firm performance and regional growth (Frenken, Van Oort, & Verburg, 2007). This literature argues that a more diversified local structure is preferable because diversity facilitates the generation of new ideas, induces knowledge spillovers, and provides more valuable resources for innovation (G. Duranton & Puga, 2001; Jacobs, 1969). However, several studies stress that it is not only variety that matters, also important is the presence of sectors that are related in terms of shared or complementary knowledge: leading to the concepts of “technological relatedness” and “related variety” (Boschma & Frenken, 2009). Asheim et al. (2007) explicitly differentiate between “knowledge related” variety, and input-output relations between sectors. However, most of knowledge spillovers are generated among firms in vertical relations. Also if technological relatedness is measured on the basis of co-occurrence analysis – i.e. the effective pattern of technological diversification of firms – their role is pre-eminent (Bryce & Winter, 2009; Neffke, Henning, & Boschma, 2009).

For this reason we would suggest that it is not just variety but also the technological relatedness between local industries that influences vertical integration decisions. In particular, the presence in the same province, of firms belonging to industries that are vertically related, increases the probability of obtaining intermediate inputs from the market, rather than integrating activities within the firm.

Thus, we propose that:

\[ H_1: \text{industry specialization at local level is expected to play a negative role in the degree of vertical integration}; \]

\[ H_2: \text{related variety in the local system in terms of input-output linkages, is expected to have a negative impact on the degree of vertical integration}. \]

3. DATA AND METHODOLOGY

3.1 Dataset

For the empirical analysis we use a business groups dataset developed by ISTAT (Italian National Statistical Institute). Business groups are identified through control linkages between pairs of legal units, according to European level operational guidelines (Eurostat, 2003). The dataset was built
considering the population of joint-stock companies in the Italian economy. 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 of the group is determined by the activity of the largest company.

In measuring degree of vertical integration we take account only of firms located in Italy. Previous work shows that almost all firms in a particular group are located in the same province (Cainelli & Iacobucci, 2007). This makes groups well suited to a study of the relations between agglomeration within a local production system and vertical integration.

The 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. The resulting dataset contains observations on 87,688 firms and 24,663 business groups.

3.2 Measuring vertical integration

One of the main problems in assessing degree of vertical integration is the availability of information about the control, by the same firm, of adjacent activities along the production chain (Acemoglu, Aghion, Griffith, & Zilibotti, 2010). To overcome this problem we use data on the production activities controlled by a business group. In taking the group as the organizational form we consider ownership and control of business activities are appropriate to identify vertical integration. We contend that this is an appropriate way to characterize vertical integration since the problems that arise from market transactions refer to contractual relations between different firms, rather than just the physical distance between establishments. What counts for vertical integration is the ownership and control of assets rather than their physical location within the same plant (Baker & Hubbard, 2004; Feenstra & Hanson, 2005).

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).

---

1 Given the increasing level of internationalization of production, this might be considered a major limitation for the study of vertical integration. However, within the aims and scope of this study it is not, and for several reasons. First, the number of companies controlled abroad accounts for only about 5% of the total. Second, our focus is on the role of spatial agglomeration on vertical integration. Third, consideration of domestic companies is more congruent with the input-output table on domestic exchanges.
Given \( v_{ij} = \text{value of input acquired by industry } i \text{ from industry } j \text{ on total output of industry } i, \) \( v_{ji} = \text{share of input acquired by industry } j \text{ from industry } i \text{ on total output of industry } j, \) the vertical relatedness 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.\(^2\) Excluding exchanges within the same sector, we have 1,431 pairs for which we calculate the value of \( V_{ij}. \)\(^3\) 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 vertical relatedness index between the two industries is 0.1439.

To distinguish the presence of vertical integration we use information on the activities controlled by the same business group. A group is associated to an industry on the basis of its primary activity; primary activity is defined as the activity (industry) employing the largest numbers. We prefer to use employees because, in most cases, there are buying-supplying relations that duplicate sales; a commercial company might be defined as the largest company only because it acquires and sells the products or services of other firms in the group. Some groups contain more than one company belonging to the same industry.\(^4\) To calculate the index of vertical relatedness we aggregate the sales of groups by industry. The aggregation reduces the number of different codes in groups to 52,859. Given that the overall number of groups is 24,663, this means there is an average of about 2 codes per group.

Given \( i \) us the primary industry in a group, for each of the other industries \( i \neq j \) controlled by the same group, we construct a measure of the actual presence of vertical integration as the ratio between the employees of the group in industry \( j \) (the supply industry) and the employees in industry \( i \) (the acquiring industry).

\[
w_{ij} = \frac{\text{emp}_j}{\text{emp}_i}
\]

Given that industry \( i \) is, by definition, the primary industry: \( \text{emp}_i > \text{emp}_j. \)

---

\(^2\) The Italian input-output table uses the classification of economic activities NaceRev.1.1.

\(^3\) This is the result of combining 2 industry codes out of 54: \( c^k = \frac{a^k}{k!(n-k)!}. \)

\(^4\) And also because we aggregated the industries to which a company belongs according to the aggregation level in the input-output table.
Moreover, it could be that \( \text{emp}_j / \text{emp}_i > V_{ij} \). For this reason, to measure the effective presence of vertical integration for each pair of industries \( ij \) we build the following index:

\[
w_{2y} = \max \left( \frac{\text{emp}_j}{\text{emp}_i}, V_{ij} \right)
\]

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

\[
\text{Vertical int}_f = \sum_{j=1}^{n} w_{2y}
\]

where \( i \) denotes the primary industry of group \( f \); \( j \neq i \) indicates all the other industries in which group \( f \) has controlling 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

We define different types of agglomeration forces: (i) geographical concentration of industries; (ii) unrelated variety; (iii) related variety.

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

\[
\text{Specialization}_{i,k} = \frac{l_{i,k}}{l_k} \frac{l_k}{l_{i,\text{Italy}}} \frac{l_{i,\text{Italy}}}{l_{\text{Italy}}}
\]

where \( l_{i,k} \) denotes employment of industry \( i \) in province \( k \), \( l_k \) total employment of province \( k \), \( l_{i,\text{Italy}} \) total employment of industry \( i \) in Italy, and finally \( l_{\text{Italy}} \) total manufacturing employment in Italy.

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

\[
\text{Unrelated variety}_k = \sum_{g=1}^{G} S_g \log_2 \left( \frac{1}{S_g} \right)
\]
where \( S_g \) \((g=1, 2, \ldots, G)\) are the shares of employees at 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.\(^5\) The higher the value of the index, the higher the variety of the industries that are located in the same geographical area (the province in our case).

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

\[
\text{Technology related variety}_k = \frac{\sum_{i} \sum_{j} S_{ik} S_{jk} a_{ij}}{\sum_{i} \sum_{j} S_{ik} S_{jk}}
\]

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: i) degree of concentration of activities in a few industries; ii) degree of relatedness between industries.

An interesting aspect of this index is that the “relatedness” factor \( a_{ij} \) can be chosen to represent different dimensions of “relatedness”. Frenken et al. (2007) consider \( a_{ij} \) as an index of the similarities between industries based on their input structure. This is intended to capture technological similarities among sectors because the authors were interested in assessing the level of knowledge spillovers between industries. In our case an appropriate measure of relatedness is degree of vertical relatedness between industries, defined as \( v_{ij} \) by Fan and Lang (2000); this index captures the opportunities for vertical exchanges between sectors. We have 50 industries (excluding agriculture and energy) and 103 provinces. We construct the index taking into consideration the number of employees in 1991 and 2001, and the vertical relatedness index \( v_{ij} \) calculated using the input-output table for 2000. All these measures of agglomeration are computed at the province level. In Italy, the province is an intermediate level between municipality and region. The Italian territory is split into 103 provinces with the average population per province about 600,000 people. The province encompasses the local

\(^5\) The entropy index does not have an upper bound. Its maximum level depends on \( n \):

\[
H_{\text{max}} = \sum_{i=1}^{n} \frac{1}{n} \log_2(n) = n \frac{1}{n} \log_2(n) = \log_2(n).
\]

Given that we consider 50 industries the maximum value of the index is 3.91.
labor system (i.e. the territory within which people commute for working reasons) and industrial districts. Also, most firms within any business group are generally located within the same province. For these reasons the province appears to be an appropriate territorial level to characterize the business environment of firms.

3.4 Controls

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

We introduce the following controls:

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

(ii) the size of the local system (Local system size) captured by the provincial population in 2000 in order to account for the size of the local market;

(iii) the size of the group (Group size);

(iv) 22 industry dummies at the 2-digit level in the estimates for all business groups, and 4 Pavitt sector dummies in the case of manufacturing groups.

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 & Iacobucci, 2009).

3.5 Econometric modelling

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

\[ y_f^* = X_f \beta + \varepsilon_f \]

where \( \varepsilon_f \approx N(0, \sigma^2) \) and \( X_f \) denotes the independent variables defined in Section 3.3.

Table 1 reports some descriptive statistics for these variables. \( y^* \) is a latent variable that is observed for values greater than \( \tau \) and censored otherwise. In our case \( \tau = 0 \) : i.e. 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 \( \beta \) will
be inconsistent. The observed \( y_f \) – our measure of vertical integration – is defined by the following measurement equation:

\[
y_f = \begin{cases} 
  y_f^* & \text{if } y_f^* > 0 \\
  0 & \text{if } y_f^* \leq 0
\end{cases}
\]

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

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

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, vertical integration can also affect agglomeration, thus generating a classical reverse causality problem. In fact, as suggested by Helsley & Strange (2007) “the effect of agglomeration on opportunism and the organization of production is a force that will lead firms to agglomerate” (Helsley & Strange, 2007, p. 57). The presence of (potential) endogeneity – i.e., one or more explanatory variables correlated with the error term – can generate biased and inconsistent estimates of the coefficients under investigation. For this reason, we control for this by assuming, as is usual in the regional economics literature (Glaeser et al., 1992; Henderson, Kuncoro, & Turner, 1995), a time lag between vertical integration and the agglomeration variables. These latter refer to 1991. Also the Local system size variable is lagged of one year in order to account for the simultaneity problem. It refers to 2000.

Table 1 – Descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Year</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std.dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical int.</td>
<td>Group</td>
<td>2001</td>
<td>24,633</td>
<td>.0149</td>
<td>.0211</td>
<td>0</td>
<td>0.214</td>
</tr>
<tr>
<td>Group size</td>
<td>Group</td>
<td>2001</td>
<td>24,633</td>
<td>.995</td>
<td>.487</td>
<td>.693</td>
<td>6.405</td>
</tr>
<tr>
<td>Local system size</td>
<td>Province</td>
<td>2000</td>
<td>103</td>
<td>1,484.615</td>
<td>1,354.979</td>
<td>90,076</td>
<td>3,705.018</td>
</tr>
<tr>
<td>Unrelated variety</td>
<td>Province</td>
<td>1991</td>
<td>103</td>
<td>4.726</td>
<td>0.205</td>
<td>3.745</td>
<td>4.958</td>
</tr>
<tr>
<td>Vertically related variety</td>
<td>Province</td>
<td>1991</td>
<td>103</td>
<td>.062</td>
<td>.0145</td>
<td>.052</td>
<td>.174</td>
</tr>
<tr>
<td>Specialization</td>
<td>Province</td>
<td>1991</td>
<td>103</td>
<td>1.413</td>
<td>1.536</td>
<td>0</td>
<td>16.737</td>
</tr>
</tbody>
</table>
4. EMPIRICAL RESULTS

Tobit estimates are performed for the whole population of Italian business groups, and for manufacturing groups only. The results for the population of business groups are presented in Table 2. We estimate three different econometric specifications by introducing different variables, step by step, to capture specialization, unrelated variety, and the technological relatedness of economic activities at the local level.6

The coefficient of the specialization variable has the expected sign and is significant at the 1% level. This confirms the results of other studies: i.e., 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, the variety of the local system is expected to play a role in facilitating exchanges between industries that are vertically related. The coefficient of unrelated variety does not confirm this prediction: it has a positive sign. The positive sign of the variety variable could be due to the fact that it gives no information on the “types” of industries that are present; i.e. whether they are related along the same production chain, giving scope for vertical exchanges between firms. The indicator measuring degree of vertically related variety has the expected sign and its coefficient is highly statistically significant. This shows that it is not variety *per se* that matters for vertical integration, but the presence of industries that are technologically related in terms input-output exchanges.

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

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 enhance the possibility of market transactions and reduce the need for integrating activities within the firm.

---

6 We do not include all three agglomerative variables simultaneously because they are highly collinear.
Our results are similar when we consider the sub-population of manufacturing groups (see Table 3). However they differ in that the coefficient of unrelated variety for the sub-population of manufacturing firms is not statistically significant, but the coefficients of specialization and vertically related variety show the expected sign, and are both significant. This is further confirmation that specialization matters for vertical integration, as suggested by previous studies (Brookfield, 2008; Diez-Vial & Alvarez-Suescun, 2010; Holmes, 1999), but together with vertically related variety. The two variables capture different factors influencing vertical integration. The first captures the role of geographical concentration in reducing opportunism and thus serves as a substitute for integration (Dei Ottati, 1994; Diez-Vial & Alvarez-Suescun, 2010; Helsley & Strange, 2007). The second underlines that what matters for decisions about vertical integration is not just physical proximity but also the technological relatedness among the industries in the same local system. Also, in the case of manufacturing groups, we introduce geographic and industry dummies. The geographic dummy is the same as we used for the population of groups. Our industry dummies are based on Pavitt’s (Pavitt, 1984) taxonomy which refers to the innovation pattern in manufacturing

### Table 2 – Determinants of vertical integration

<table>
<thead>
<tr>
<th>Estimation Method: Tobit</th>
<th>All business groups</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialization</td>
<td>-0.0005**</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td></td>
<td>[-2.34]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrelated variety</td>
<td>…</td>
<td>0.003**</td>
<td>[1.97]</td>
<td>…</td>
</tr>
<tr>
<td>Vertically related variety</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>-0.047***</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td></td>
<td></td>
<td>[-3.50]</td>
</tr>
<tr>
<td>Local system size</td>
<td>-0.0006***</td>
<td>-0.0009***</td>
<td>-0.0008***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-3.50]</td>
<td>[-4.13]</td>
<td>[-4.45]</td>
<td></td>
</tr>
<tr>
<td>Group size</td>
<td>0.012***</td>
<td>0.012***</td>
<td>0.012***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[17.73]</td>
<td>[17.95]</td>
<td>[17.95]</td>
<td></td>
</tr>
<tr>
<td>Industry dummy</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Geographic dummy</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

N. Obs. 24,633 24,633 24,633
Uncensored Obs. 15,811 15,811 15,811
Censored Obs. 8,882 8,882 8,882
Log pseudolikelihood 27323.8 27317.3 27321.6
Pseudo R² 0.020 0.020 0.020

*** 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.
sectors. This is based on the influence of technological regimes on the vertical integration decisions of firms (Acemoglu, Aghion, Griffith, & Zilibotti, 2004; Cainelli & Iacobucci, 2009).

Table 3 – Determinants of vertical integration in manufacturing groups

<table>
<thead>
<tr>
<th>ESTIMATION METHOD: TOBIT</th>
<th>Only manufacturing groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[1]</td>
</tr>
<tr>
<td>Specialization</td>
<td>-0.0006***</td>
</tr>
<tr>
<td></td>
<td>[-3.75]</td>
</tr>
<tr>
<td>Unrelated variety</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Vertically related variety</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Local system size</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>[-1.16]</td>
</tr>
<tr>
<td>Group size</td>
<td>0.009***</td>
</tr>
<tr>
<td></td>
<td>[13.38]</td>
</tr>
<tr>
<td>Supplier dominated</td>
<td>-0.006***</td>
</tr>
<tr>
<td></td>
<td>[-7.54]</td>
</tr>
<tr>
<td>Scale Intensive</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>[0.61]</td>
</tr>
<tr>
<td>Science based</td>
<td>-0.004***</td>
</tr>
<tr>
<td></td>
<td>[-3.30]</td>
</tr>
<tr>
<td>Specialized supplier</td>
<td>Ref.</td>
</tr>
<tr>
<td>Geographic dummy</td>
<td>Yes</td>
</tr>
<tr>
<td>N. Obs.</td>
<td>8,097</td>
</tr>
<tr>
<td>Uncensored Obs.</td>
<td>5,786</td>
</tr>
<tr>
<td>Censored Obs.</td>
<td>2,311</td>
</tr>
<tr>
<td>Log pseudolikelihood</td>
<td>9945.9</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.024</td>
</tr>
</tbody>
</table>

*** 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

5. CONCLUSIONS

We used a large dataset on Italian business groups to investigate the relationship between agglomeration and inter-industry vertical integration. To measure degree of vertical integration we use the business group as the unit of analysis because we associate vertical integration with the ownership and control of activities along the production chain, rather than the presence of adjacent production phases within the same establishment (Diez-Vial & Alvarez-Suescun, 2010; Holmes,
The choice of the group as a unit of analysis is relevant specifically to this study because in most cases control of different phases in the production chain is achieved through ownership of the different companies rather than integration of activities within the same legal unit. We take the province as the appropriate geographical unit (in the Italian case) as the context of agglomeration forces.

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

We investigated the role of agglomeration forces in vertical integration choices by analyzing the effect of variety. We used two different indicators of variety: the first a general indicator of “unrelated variety” that captures the presence of different industries within the same geographical area; the second a more specific indicator of industry variety that captures the presence of vertically related industries. The relevance of the vertical relations between sectors is measured on the basis of the coefficients in the input-output table.

Our findings show that it is not variety ‘per se’ that matters for vertical integration choices but the technological relatedness of production activities at the local level. In other words, a higher level of vertical related variety significantly reduces the need 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 about control over the production phases rather than their acquisitions through market transactions. We show also that geographical concentration which reduces opportunism affects vertical integration, as underlined in previous work, but also the degree of vertically related variety at the local level, i.e. the technological relatedness between industries located in the same local system, has an effect.
References


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

<table>
<thead>
<tr>
<th>AREAS</th>
<th>N.</th>
<th>%</th>
<th>Employees</th>
<th>N.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>North_West</td>
<td>9368</td>
<td>38.03</td>
<td>1882676</td>
<td>47.54</td>
<td></td>
</tr>
<tr>
<td>North-East</td>
<td>6700</td>
<td>27.2</td>
<td>783272</td>
<td>19.78</td>
<td></td>
</tr>
<tr>
<td>Centre</td>
<td>5263</td>
<td>21.37</td>
<td>1058791</td>
<td>26.73</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>2373</td>
<td>9.63</td>
<td>167328</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>Islands</td>
<td>929</td>
<td>3.77</td>
<td>68406</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>24,633</strong></td>
<td><strong>100.0</strong></td>
<td><strong>3,960,473</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>ATECO</th>
<th>Firms</th>
<th>Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.</td>
<td>%</td>
<td>N.</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>8,230</td>
<td>33.4</td>
</tr>
<tr>
<td>Services</td>
<td>16,403</td>
<td>66.6</td>
</tr>
<tr>
<td><strong>Totale</strong></td>
<td><strong>24,633</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

### Table A.3 – Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Vertical int.</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2] Local system size</td>
<td>0.004</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3] Group size</td>
<td>0.123</td>
<td>0.058</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[4] Specialization</td>
<td>-0.001</td>
<td>-0.090</td>
<td>0.023</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5] Technologically related variety</td>
<td>-0.035</td>
<td>-0.337</td>
<td>-0.045</td>
<td>0.343</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>[6] Unrelated variety</td>
<td>0.042</td>
<td>0.498</td>
<td>0.067</td>
<td>-0.200</td>
<td>-0.900</td>
<td>1.000</td>
</tr>
</tbody>
</table>

[2] Local system size
[4] Specialization
[5] Technologically related variety
[6] Unrelated variety