UNIVERSITÀ CATTOLICA del Sacro Cuore

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

Productivity and earnings at Firms' Local Unit Level: the case of Lombardy urban and non-urban agglomerations

Eleonora Bartoloni
Maurizio Baussola

Quaderno n. 144/novembre 2019

VP VITA E PENSIERO

Università Cattolica del Sacro Cuore

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

Productivity and earnings at Firms' Local Unit Level: the case of Lombardy urban and non-urban agglomerations

Eleonora Bartoloni
Maurizio Baussola

Quaderno n. 144/novembre 2019

VP VITA E PENSIERO

Eleonora Bartoloni, ISTAT, Italian National Institute of Statistics, via Oldofredi, 23, 20124 Milano, and Università di Parma, Dipartimento di Scienze Economiche e Aziendali, via J.F. Kennedy 6 - 43125 Parma.

Maurizio Baussola, Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza.

✉ bartolon@istat.it

✉ maurizio.baussola@unicatt.it

I quaderni possono essere richiesti a:
Dipartimento di Scienze Economiche e Sociali,
Università Cattolica del Sacro Cuore
Via Emilia Parmense 84 - 29122 Piacenza - Tel. 0523 599.342
http://dipartimenti.unicatt.it/dises

✉ dises-pc@unicatt.it

www.vitaepensiero.it

ISBN 978-88-343-4147-6

Abstract

We analyse business performance by using a unique dataset of the universe of Italian local business units. We investigate the pattern of both productivity and profitability by adopting a decomposition technique to document spatial variation across urban and non-urban areas. Aggregate evidence indicates that an urban – non-urban productivity divide exists, but this premium vanishes with respect to profitability. Plant-level estimations using a Hierarchical Linear Model show that area attractiveness positively affects productivity, whereas diseconomies of agglomeration negatively affect profitability. Coping with agglomeration costs is a priority for the regional policy if the productivity gains in urban areas could be transferred into new investment and growth opportunities.

Keywords: local agglomeration, firm performance, decomposition analysis, hierarchical linear model, Lombardy Region
JEL codes: R12; R3; C43; C21; L25

1. Introduction

It is generally acknowledged that firms are highly heterogeneous as concerns economic performance. Both productivity and profitability vary dramatically across firms (Syverson, 2011; Andrews et al., 2015; Foster et al., 2008), and the economic literature has typically re-conducted these differences to firm-specific characteristics and sectoral specificities. Nevertheless, even when accounting for internal and sectoral determinants, the external environment, i.e. geographical location, matters in affecting the performance of firms or plants which operate in proximity areas (Syverson, 2011).

Spatial concentration of economic and social activities has been rising faster in the last three decades: according to the United Nations (2018), nowadays, more than half of the world's population lives in urban areas, and projections indicate that by 2050 urban population will increase to almost 70%, at the expenses of rural population which has grown slowly from the '50s.

At the same time, the vast body of theoretical and empirical literature available so far tends to conclude that agglomeration economies enhance economic performance at the local level, thus allowing faster growth of firms localized in urban areas. Accordingly, empirical research has also identified the presence of an urban-rural divide in terms of business productivity (Rizov and Walsh, 2011; Webber, 2009).

From a historical perspective, these more recent trends contrast with the opposite process of reallocation of manufacturing employment from urban to rural areas that, starting from the late 60s of the last century, has been observed among industrialized countries (Keeble et al., 1983). This latter process has been explained in terms of urban space shortage and higher operating costs compared to rural areas (Fothergill and Gudgin, 1982; Fothergill et al., 1985) which have led large industrial firms to decentralize in order to maintain profit margins. Unfortunately, this debate was almost exclusively focused on the relevance of manufacturing activities, thus failing to take into adequate consideration the increasing importance of the knowledge economy, which today seems to be highly responsible for urban growth.

Despite the enhancements in empirical research, the observed relationship between firm performance and its determinants at the local level still requires a thorough investigation which may also encompass the different approaches used for such analyses and at the same time use more appropriate data sources.

This work explores business performance in the Lombardy region in Italy by using the information at the sub-regional level. Our analysis represents an original contribution to the current debate in many respects. Firstly, it focuses on economic performance of both manufacturing and services firms by providing a comprehensive assessment of both productivity and profitability conditions, the latter reflecting possible firm-specific cost advantages or disadvantages because of localization which do not necessarily reflect efficiency conditions.

Firms' performance is generally considered by looking mainly at productivity or sales growth, thus implicitly assuming that the implied impact of these variables on profitability is positive. In this study, firms' performance is considered as a twofold mechanism simultaneously represented by productivity and profitability. Although it is reasonable to assume that higher productive firms earn larger profits, we want to ascertain whether this mechanism is operational in an empirical framework in which we do account for possible agglomeration costs. Indeed, we show that there exists a significant and positive effect of urbanization on firms' productivity, but the same argument cannot be used for profitability, thus suggesting that dis-economy of agglomeration are at work.

Secondly, this investigation is based on the universe of companies localized in the Lombardy region. A new integrated database developed by the National Institute of Statistics (ISTAT) provides census data at the local unit (plant) level, which are also linked to demographic and economic information. Thirdly, it explores three different types of local clusters reflecting different degrees of urbanization with the municipality (local administrative unit, LAU2) as the basic unit of aggregation. Fourthly, in order to investigate the sources of aggregate performance in urban and non-urban locations, it applies a decomposition (Olley and Pakes, 1996) to spatial variation of productivity and profitability levels in order to distinguish the contribution of firm-level efficiency conditions from the role played by differences in sectoral composition. Finally, the analysis moves from the aggregate evidence to the local unit level perspective, which has the advantage of addressing the role played by individual heterogeneity. We apply a Hierarchical Linear Model approach, which allows us to estimate individual performance by controlling for firm (plant) specific characteristics and agglomeration externality and, at the same time, taking into consideration the nested structure of our data that are grouped into sectoral/geographical clusters. Agglomeration externalities are controlled for by using a data mining technique applied to a broad set of indicators at the municipality level.

The paper proceeds as follows. The next section provides a review of the relevant empirical literature, emphasizing the heterogeneous approaches which have been proposed in analysing business performance at the local level. Section 3 presents the data and geographical clusters of analysis. Section 4 explores aggregate patterns of economic performance, while in Section 5 we develop an econometric analysis at the individual level and discuss the results. Section 6 presents our conclusive remarks.

2. Related literature

The role of agglomeration economies has long been debated starting from the seminal contribution by Marshall (1920), who explained the mechanisms that drive the geographical concentration of economic activities. Firstly, physical proximity allows firms to access a common pool of workers (labour market pooling). Firms may find it easier to access to diversified job experiences thus increasing their skills and indirectly affecting the availability of specialized labour in the agglomeration. Secondly, firms operating in concentrated areas may access to a greater variety of intermediate inputs such as machinery and equipment (input sharing) thus increasing efficiency conditions. Finally, the transmission of knowledge spillovers can be enhanced when firms are located in close geographic proximity. It has been argued that these type of agglomeration externalities, also called *localization economies,* operate within clusters of firms which belong to specific or similar industries.

In a different perspective, Jacobs (1969) argues that the most important sources of external economies do not depend on industry structure. All firms sharing the same location may benefit from general advantages arising because of the size of the area or the density of a vast range of activities which are not solely economic but also social, cultural and political. In this respect, *urbanization economies* may positively affect firm performance in agglomerations which are more populated or with easy access to large metropolitan areas. These firms may also benefit from knowledge spillovers because of the presence of research centres, universities or other kinds of players promoting the circulation of ideas and information.

The distinction between localization and urbanization economies offers the advantage of focusing on different mechanisms affecting economic performance (Rigby and Essletzbichler, 2002), that is, specialization externalities generating within-industry spillovers in the first case and, in the second case, diversity externalities, arising because of the complimentary access to different types of knowledge and skills coming from different agents in the geographical proximity. Nevertheless, empirical evidence about the forms and the specific impacts of agglomeration externalities on firm performance is quite mixed. Beaudry and Schiffauerova (2009) provide a detailed survey of empirical contributions and summarize results in terms of similarities and differences. They suggest that the lack of conclusive evidence highly depends on (i) the type of aggregation used, both at the industry and geographical level, (ii) the proxies used to measure economic performance and (iii) the kind of indicators used to capture different agglomeration mechanisms. Interestingly, results also indicate that when the level of geographical aggregation is small both specialization and diversity externalities are at work, thus suggesting that a rigid distinction between localization and urbanization economies may be misleading when analyzing their impact on firm performance at the local level of aggregation.

Besides positive agglomeration externalities, congestion costs negatively affect firm performance (Henderson, 1974), thus representing centrifugal forces. These negative externalities may be due to traffic jam (Sweet, 2014; Baert and Reynaerts, 2018), pollution, crime, high housing and land rents, increased labour costs because of competition among firms for the skilled workforce (Combes and Duranton, 2006). As urbanization diseconomies cannot be directly measured, some empirical studies investigate the *net* effect of agglomeration economies (Rizov et al, 2012; Martin et al, 2011; Saito and Wu, 2016). This issue is also investigated by Brinkman (2016) who simulated and estimated a structural spatial model in which congestion costs and agglomeration externalities are jointly considered. In particular, policy simulation shows that congestion policy which modifies congestion prices could lead to negative economic outcomes.

A relevant question concerns the measure of performance to be used to assess the impact of agglomeration economies. The regional science literature available so far has extensively used at least three indicators: economic growth (Cingano and Schivardi, 2004; Combes et al. 2004; Audretsch and Dohse, 2007), productivity (Henderson, 2003; Baldwin et al., 2008; Martin et al, 2011; Anderson and Loof, 2011; Webber et al., 2009; Rizov and Walsh, 2011) and innovation (Feldman and Audretsch, 1999; Beaudry and Breschi, 2003; Van Oort, 2002; Boschma, 2005; Carlino and Kerr, 2014). This is not surprising given the main interest of this stream of literature on regional growth implications.

Nevertheless, an emerging body of literature has paid increasing attention to the relationships between the spatial density of economic activities and individual business performance (McCann and Folta, 2011). Indeed, this approach is well-grounded given the micro-foundation of the three underlying mechanisms – sharing, matching, and learning – addressed by the Marshallian theory of agglomeration (Duranton and Puga, 2004; Rosenthal and Strange, 2004).

A micro-founded perspective is recommended to clarify better the complex mechanisms which affect the performance-localization relationship. It has been argued that high productivity levels in high dense areas may occur not because agglomeration economies enhance firm-level productivity, but because of a selection mechanism that induces those firms which are (ex-ante) better endowed with internal attributes to locate in agglomerated areas because of costs savings factors. (Baldwin and Okubo, 2006). To this respect, the adoption of a micro-level approach allows one to control for firm-specific characteristics and, thus, to avoid that differences among areas in terms of firm-specific heterogeneity may induce overestimation of productivity differentials, if not taken adequately into consideration. Indeed, recent investigations have increasingly used firm (plant)-level data in order to investigate the relationship between agglomeration and productivity.

The study by Henderson (2003) was among the first employing plant-level data. By using a US panel over the period 1072-92 for machinery and high-tech US industries he finds a positive effect of localization economies on Total Factor Productivity (??) which is, however, limited to the high-tech industry: a 10% increase in the agglomeration index used, given by the number of other plants of the same industry in the county, increases the productivity of the plant by 0.8%. No significant effect arising from urbanization externalities is found.

Martin et al. (2011) use French plant-level data for both manufacturing and services industries over the period 1996-2004. In the specification of agglomeration externalities, they adopt both localization and urbanization indexes and find that only localization externalities can significantly increase productivity in terms of TFP: a 10% increase of employment in neighbouring firms of the same industry increases firm productivity of 0.5–0.6%.

Baldwin et al. (2008) find similar results for a cross-section of Canadian manufacturing plants. Their analysis considers the impact on labour productivity of different Marshallian mechanisms (labour market pooling, knowledge spillovers, buyer and supplier networks), thus focusing the investigation on the impact of localization externalities. OLS estimations show that, after controlling for individual specific characteristics and sectoral fixed effects, all the three Marshall's externalities previously identified positively affect productivity although the precise impact may be not homogeneous in size and significance across industrial sectors.

Andersson and Loof (2011) propose both static and dynamic estimations based on a panel of Swedish manufacturing firms during the period 1997-2004. Explanatory variables in both specifications include firm-specific attributes and a general proxy for the agglomeration potential in a given location, given by the size (in terms of employment) of the region where the plant is located. In the dynamic specification lagged productivity has been included among the set of regressors in order to control for persistence patterns. Their results support the view that firms located in larger regions are more productive after controlling for firm-specific attributes. The significant impact of agglomeration persists (although reduced in size) after controlling for past productivity. This effect is interpreted as evidence not only of an expected positive correlation between agglomeration and performance but also that agglomeration significantly contributes to enhancing productivity, thus pointing to a learning mechanism implied by this result.

Even when the effects of agglomeration economies are analyzed by using firm (plant) level productivity data, the emerging of a positive impact on a firm's operating efficiency does not imply a positive gain in terms of financial performance, i.e. profitability.

Some recent empirical evidence based on plant-level data supports this hypothesis. By using a sample of Dutch establishments, Jennen and Verwijmeren (2010) suggest that the density of local areas negatively affects firms' profitability due to agglomeration costs that outweigh the benefits. On the basis of a variance decomposition analysis applied to a large sample of manufacturing firms located in EU regions, Stavropoulos and Skuras (2016) find that inter-regional differences explain only a small proportion of total profit variance, thus indirectly supporting the view that agglomeration economies, although not directly included in the model, do not affect firm profitability. In a longitudinal study on a small sample of Indian firms operating in the semiconductor and pharmaceutical industries, Kukalis (2010) does not find any profit premium for clustered firms compared to the non-clustered one in the early stages of the industry life-cycle. However, when the late stages are reached, non-clustered firms outperform the clustered ones.

This empirical evidence suggests that the firm performance-agglomeration relationship is somewhat complex and far from being completely understood. It requires both accurate information at the plant level and improved methodological tools. Moreover, the selection of geographical aggregations has a crucial role as different levels of clustering could be affected by different dimensions which may tend to obscure the real impact of agglomeration economies.

Research at the sub-regional level has also identified the presence of an urban-rural productivity divide. Decomposition techniques have been used to analyse spatial differences at the regional level (Rice and al., 2006; Oosterhaven and Boersna, 2007). Within this line of analysis, Rizov and Walsh (2011) consider the classification in urban areas, rural less sparse areas and rural sparse areas defined by the UK Department for Environment, Food and Rural Affairs (DEFRA) and adopt a decomposition methodology to UK plant-level data in order to unveil spatial variation of *business* productivity concerning of TFP. This allows one to distinguish between variation due to plant-level productivity, resulting in different productivity averages across sectors and variation due to different industry composition in each location.

Results indicate that a rural-urban productivity divide does exist, which is due to both differences in plant-level efficiency conditions and localization effects, thus suggesting that the non-urban locations are characterized by both low productivity conditions and not adequate industrial mix. Webber et al. (2009) adopt the same DEFRA urban-rural classification to estimate labour productivity differentials at the plant-level by using an OLS regression analysis, which has the advantage to control for a set of firm-specific, sectoral and area-level characteristics. Results are in line with Rizov and Walsh (2011), showing the presence of an urban-rural divide in business labour productivity. They also show that plant-level productivity in sparse areas is lower compared to urban and less-sparse areas not only because of different industrial structure (i.e. localization economies) but also because of firm-specific characteristics in terms of ownership structure and capital stock endowment.

3. Geographical aggregations and data

3.1 Location clusters in the Lombardy region

We begin our analysis by considering three types of local aggregations: (i) The Eurostat Urban-Rural classification (DEGURBA); (ii) The PSR-EAFRD classification and (iii) the Istat Labour Market Areas (SLL, Sistemi Locali del Lavoro). All the proposed classification criteria adopt the municipality (local administrative unit, LAU2) as the basic unit of aggregation and are based on underlying functional characteristics. In particular, while the first two classifications are used as policy instruments, the SLL classification is a statistical classification developed by the National Institute of Statistics.

The degree of urbanization classification (DEGURBA) was set up by Eurostat and adopted within various surveys conducted at the European level. Based on a combination of criteria of geographical contiguity and minimum population threshold, this type of aggregation produces a map of the regional territory by a grid square cell of 1 km^2 to avoid distortions caused by using local administrative units varying in size and/or shape. This methodology creates a classification of municipalities in three groups: high densely populated areas (at least 50% of the population lives in high-density urban centres), intermediate urban areas (at least 50% of the population lives in urban clusters) and rural areas (at least 50% of the population lives in rural grid cells)[1].

An alternative classification was adopted in Italy as part of the Rural Development Program 2014-20 (PSR) which is sponsored by the European Agricultural Fund for Rural Development (EAFRD). This classification is applied to the Italian municipalities and adopts a criterion which combines population density and agricultural and forestry extension. Within this framework, Lombardy region has classified its municipalities into four types of areas: urban and peri-urban areas, rural areas with intensive agriculture, intermediate rural areas, within which diversified areas fall, and rural areas with development problems[2].

The last type of geographical agglomerations considered in the present analysis are the so-called Labour Market Areas (LMA, *Sistemi Locali del Lavoro* in Italy). These aggregations are functional areas which correspond to the geographical space where people live and work, thus moving to their working places within a relatively small commuting distance[3]. Given these characteristics, which take into consideration both social and employment patterns, LMAs can be defined as economically integrated spatial units. The map of LMAs is defined and periodically updated by applying a sophisticated algorithm to the commuting data derived from the Italian population census. Lombardy Labour Market Areas have been further characterized based on their degree of urbanization and industrial specialization (ISTAT, 2015). This classification allows the first distinction between those LMAs which have a manufacturing specialization and those which have not this kind of specialization. Further, it allows one to distinguish within the manufacturing LMAs those which have a specialization in the so-called

---

[1] For further details about the methodology used see https://agriregionieuropa.univpm.it/it/content/article/31/35/le-aree-rurali-nella-nuova-programmazione.

[2] For further details see https://ec.europa.eu/eurostat/web/degree-of-urbanisation/background.

[3] For a more precise definition see: https://www.istat.it/en/archive/142790. It is worth noting that the number of such LMA has reduced through time according to the latest calculations, as the commuting distance increases significantly over time.

Made in Italy productions from those which are characterized by a specialization in the heavy manufacturing[4]. Finally, among the LMAs

Table 1 - *Geographical classifications: basic characteristics (totals and % by aggregations)*

	Munici-palities	Population (resident inhabitants)[a]	Sur-face Km2	Firms' local units
Lombardy region	1,529	10,002,569	23,864	849,533
A Urban-Rural (Eurostat)				
Urban High-density	8%	40%	8%	49%
Intermediate	47%	48%	34%	42%
Rural	45%	12%	58%	9%
B PSR-EAFRD (Lombardy Region)				
Urban	6%	37%	6%	47%
High-spec-rural	44%	39%	43%	32%
Rural	39%	21%	33%	19%
Critical-rural	11%	3%	18%	2%
C Labour Market Areas (SLL Istat)				
Made in Italy	31%	24%	27%	21%
heavy industry specialization	47%	34%	47%	32%
tourism specialization	16%	41%	14%	46%
urban	6%	1%	12%	1%

[a] *Residents at the 1th of January 2015*

without a manufacturing specialization, this type of aggregation distinguishes between the urban systems, i.e. areas with a resident population of at least 500,000 inhabitants, whose business fabric is characterized by a significant degree of heterogeneity with a large presence of services activities, and those systems.

[4] Made in Italy identifies six traditional manufacturing sectors: clothing, food products, furnishing, footwear, eyewear and jewelry. The methodological approach used to define this group of productions is described in Centro Studi Confindustria e Prometeia (2014), p. 141.

3.2 Data description

We use a novel data set set up by the Italian National Institute of Statistics which is represented by the *FRAME Territoriale SBS,* the Italian business register which integrates firms' microdata at the plant-level by using (i) the FRAME-SBS register, representing the main data source at the firm-level on structural and economic characteristics for the total population of Italian enterprises and (ii) the ASIA-UL archive, the statistical register of active businesses, at the local unit (plant) level[5].

Firms' plants may be located in different geographical areas, which are identified at the municipality level (LAU2). These plants are referred to as local units. These units may operate in different manufacturing and services activities, excluding the financial sector as well as some personal and household services. The industrial activity to which a local unit is assigned corresponds to its principal economic activity according to the Ateco 2007 (Nace Rev. 2) classification. We have more than 4.7 million plants in Italy, generating almost 716 billion euros of value-added.

From the main dataset, we selected the subsample of plants localized in the Lombardy region. We have almost 850,000 plants in Lombardy region generating 187 billion euros of value-added corresponding to 26% of the national amount. Descriptive statistics by type of geographical aggregations are reported in Table 2.

---

[5] For further details: Istat (2018), *Risultati Economici delle Imprese a Livello Territoriale: Ampliamento del Dettaglio di Analisi*, Statistiche Report, 13 Giugno 2013. The main economic variables at the plant level are estimated by using a methodology which also exploits the information deriving from the extended register on labor costs at the local-unit (RACLI Territoriale). The *FRAME Territoriale SBS* is annually updated and this study is based on the first release, that refers to the year 2015.

Figure 1 - *Local aggregations in Lombardy*

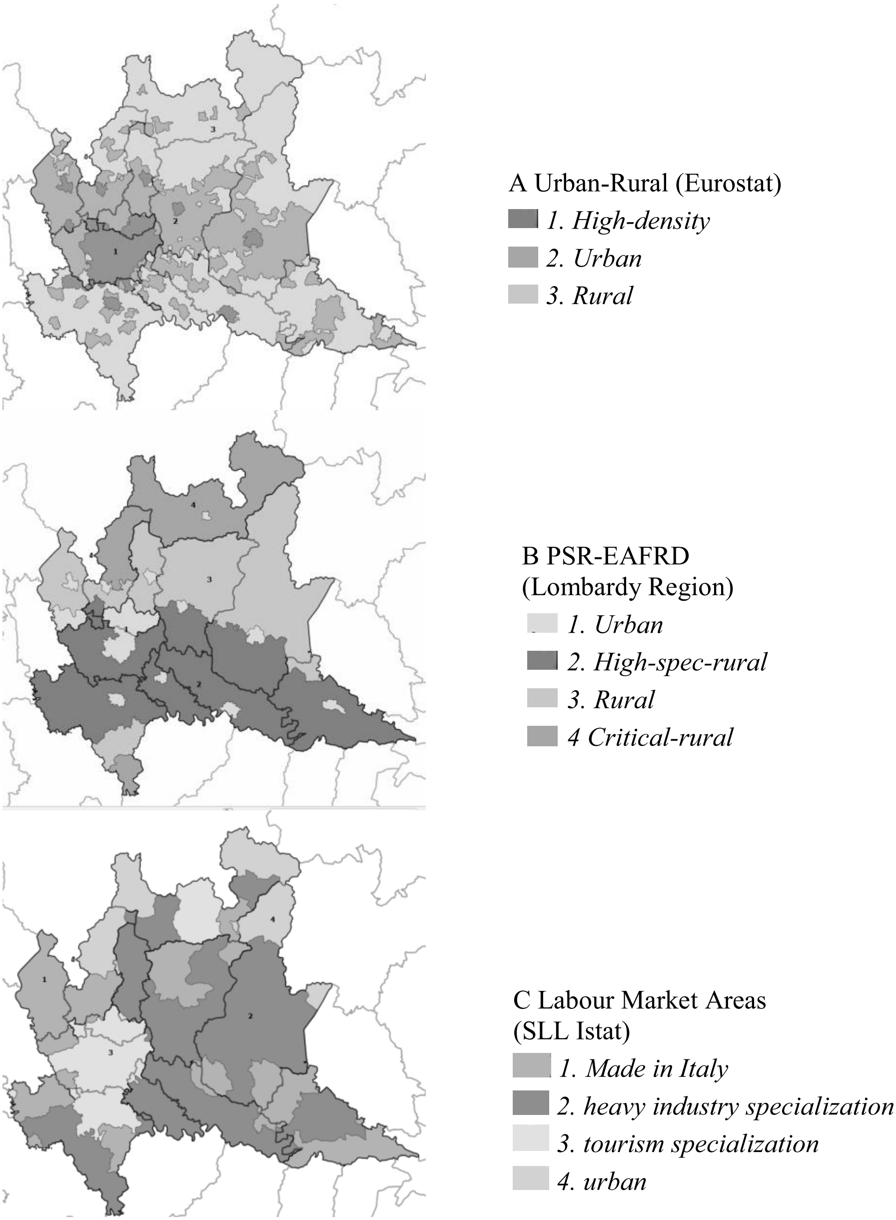

Table 2 - *Descriptive statisitcs by type of geographical aggregation*

	Local units	Emplo-yees	Value added (millions)	Labour producti-vity	Gross mar-gin ratio
	All activities				
	Lombardy totals and local area shares			Average values	
Lombardy region	849,533	3,351,781	187,000	55,791	51.7
A Urban-Rural (Eurostat)					
Urban High-density	0.49	0.51	0.57	63,066	51.1
Intermediate	0.42	0.42	0.36	48,642	53.2
Rural	0.09	0.08	0.06	45,947	49.2
B PSR-EAFRD (Lombardy Region)					
Urban	0.47	0.46	0.51	61,142	52.1
High-spec-rural	0.32	0.34	0.32	53,403	49.9
Rural	0.19	0.18	0.15	48,047	53.7
Critical-rural	0.02	0.02	0.02	42,332	52.0
C Labour Market Areas (SLL Istat)					
Made in Italy	0.21	0.20	0.17	47,699	52.2
heavy industry specialization	0.32	0.31	0.27	48,770	51.9
tourism specialization	0.46	0.48	0.55	64,405	51.1
urban	0.01	0.01	0.01	38,002	59.1

	Local units	Emplo-yees	Value added (millions)	Labour producti-vity	Gross mar-gin ratio
	Manufacturing sector				
	Lombardy totals and local area shares			Average values	
Lombardy region	89,635	881,243	60,400	68,540	49.0
A Urban-Rural (Eurostat)					
High-density	0.34	0.30	0.34	78,266	46.6
Urban	0.55	0.59	0.55	64,708	51.5
Rural	0.11	0.11	0.10	62,922	43.0
B PSR-EAFRD (Lombardy Region)					
Urban	0.33	0.26	0.29	75,074	49.9
High-spec-rural	0.39	0.44	0.44	68,034	46.8
Rural	0.25	0.28	0.26	63,778	51.6
Critical-rural	0.02	0.02	0.02	58,170	52.8
C Labour Market Areas (SLL Istat)					
Made in Italy	0.26	0.27	0.25	63,220	50.4
heavy industry specialization	0.37	0.41	0.40	66,366	49.8
tourism specialization	0.36	0.32	0.35	76,458	46.9
urban	0.01	0.01	0.00	49,336	53.1

Source: ISTAT, FRAME Territoriale SBS

4. Aggregate performance at the urban-non urban level

In this section we analyse the spatial variation in aggregate economic performance by using a decomposition technique (Rice, 2006; Rizov and Walsh, 2011) which allows us to disentangle the contribution to aggregate performance at the local area level due to the average performance at the firm level from that due to the specific industry composition.

Let p_a^i be the average economic performance level in sector i (Ateco classification, 2 digits) and area a and $\bar{p}^i$ the average level of performance for sector i in Lombardy region (i.e., aggregating all the sub-regional areas a). In the case of labour productivity, the average value is given by the ratio of sectoral value added to employment in the specific area, while in the case of profitability this is given by the ratio of sectoral profit margin to value-added (gross margin ratio). All aggregates are computed by summing up plant-level values for employment, productivity and gross margin.

Let λ_a^i be the share of industry i in area a. In the case of productivity, this is the share of employment, while in the case of profitability this is the share of value-added. In both cases we can denote industry size in area a as $S_a = \sum_i s_a^i$ and $\lambda_a^i = s_a^i / S_a$. Finally, the share of industry i in Lombardy region is given by $\bar{\lambda}^i = \sum_a s_a^i / \sum_a S_a$.

Aggregate economic performance (component A in the tables of results) in area a is a weighted average of industry performances using the appropriate industry shares as weights. It can be decomposed as follows:

$$p_a = \sum_i p_a^i \lambda_a^i = \sum_i p_a^i \bar{\lambda}^i + \sum_i \bar{p}^i \lambda_a^i + \sum_i (p_a^i - \bar{p}^i)(\lambda_a^i - \bar{\lambda}^i) - \sum_i \bar{p}^i \bar{\lambda}^i$$

The first term on the right-hand side of the above equation (component B) is the average level of firm performance in the area, which is conditional on the industry composition being the same as for the Lombardy region. The second term (component C) is the average level of performance in the area given its industry composition, but assuming that the average performance of each industry equals the value observed at the regional level. The third term (component D) measures residual covariance between industry performance and industry composition in each area, while the last term is aggregate regional performance. Results are normalized by dividing each component by the aggregate regional performance so that we obtain a straightforward interpretation in terms of indexes. Tables 3 and 4 show for each type of local aggregations the results.

Table 3 - *Labour productivity decomposition by type of geographical aggregation*

	all firms			
	A - Aggregate average productivity	B - Firm productivity index	C - Industry composition index	D - Residual covariance
A Urban-Rural (Eurostat)				
Urban High-density	113.2	110.9	101.8	0.5
Intermediate	87.2	85.6	98.0	3.6
Rural	82.4	77.9	99.1	5.4
B PSR-EAFRD (Lombardy Region)				
Urban	109.5	110.2	99.3	0.0
High-spec-rural	95.8	89.8	102.5	3.5
Rural	86.2	83.1	97.7	5.3
Critical-rural	75.9	76.2	95.5	4.2
C Labour Market Areas (SLL Istat)				
Made in Italy	85.6	84.8	96.6	4.2
heavy industry specialization	87.4	83.0	98.8	5.6
tourism specialization	68.8	69.3	89.0	10.5
urban	115.0	111.3	102.5	1.3

	manufacturing firms			
	A - Aggregate average productivity	B - Firm productivity index	C - Industry composition index	D - Residual covariance
A Urban-Rural (Eurostat)				
Urban High-density	114.3	105.4	107.0	1.9
Intermediate	94.3	96.1	96.5	1.6
Rural	91.8	89.7	99.4	2.6
B PSR-EAFRD (Lombardy Region)				
Urban	109.6	106.7	101.8	1.1
High-spec-rural	99.3	96.9	102.0	0.4
Rural	93.1	95.3	95.2	2.5
Critical-rural	84.9	87.2	97.2	0.4
C Labour Market Areas (SLL Istat)				
Made in Italy	92.0	95.1	94.4	2.5
heavy industry specialization	96.8	94.3	99.5	3.0
tourism specialization	72.5	68.9	94.4	9.2
urban	111.5	103.4	105.6	2.5

Table 4 - *Profitability decomposition by type of geographical aggregation*

	A - Aggregate average profitability	B - Firm profitability index	C - Industry composition index	D - Residual covariance
	all firms			
A Urban-Rural (Eurostat)				
Urban High-density	98.8	95.6	102.1	1.1
Intermediate	102.8	106.0	98.7	-1.8
Rural	95.2	106.7	89.1	-0.6
B PSR-EAFRD (Lombardy Region)				
Urban	100.9	97.8	102.3	0.7
High-spec-rural	96.7	94.6	97.0	5.0
Rural	104.0	108.3	98.6	-2.9
Critical-rural	101.0	105.8	98.4	-3.2
C Labour Market Areas (SLL Istat)				
Made in Italy	100.9	104.1	97.5	-0.6
heavy industry specialization	100.5	103.0	97.2	0.3
tourism specialization	113.1	117.3	96.8	-1.1
urban	99.3	96.0	102.2	1.1
	manufacturing firms			
	A - Aggregate average profitability	B - Firm profitability index	C - Industry composition index	D - Residual covariance
A Urban-Rural (Eurostat)				
Urban High-density	95.2	94.3	99.9	0.9
Intermediate	105.5	106.8	103.0	-4.3
Rural	87.6	103.4	84.4	-0.2
B PSR-EAFRD (Lombardy Region)				
Urban	101.7	96.7	104.5	0.5
High-spec-rural	95.4	99.9	95.5	0.1
Rural	105.3	107.1	102.5	-4.2
Critical-rural	107.9	109.8	103.5	-5.4
C Labour Market Areas (SLL Istat)				
Made in Italy	103.1	105.0	102.0	-3.8
heavy industry specialization	101.7	102.0	98.2	1.5
tourism specialization	104.8	103.1	99.4	2.3
urban	95.7	93.4	100.6	1.7

Source: FRAME Territoriale SBS

High-density urban areas present a level of labour productivity higher than the regional benchmark. This result is robust to the type of geographical aggregation we use: the premium in terms of efficiency is equal to +13 p.p. by using the Eurostat-DEGURBA classification, +9.5 p.p. by using the PSR-EAFRD classification and +15 p.p. if we consider the LMAs classification. We obtain similar results when confining the analysis to the labour productivity of the manufacturing firms.

Conversely, the average level of efficiency of firms placed in the other locations (intermediate urban clusters and rural) is below the regional average, although one should note that manufacturing firms show, on average, lower level of inefficiency compared to the regional benchmark.

In the urban high-density areas, both firm-specific performance and industry structure positively affect the average level of efficiency with the most important role played by firm-specific efficiency if we consider the full set of activities. However, if we limit the analysis to the manufacturing activities, we obtain less clear-cut results given that industry composition is the most relevant to average efficiency in both the Eurostat DEGURBA and LMAs classifications, but is less relevant in the urban areas defined according to the PSR-EAFRD classification.

Conversely, in the intermediate and rural locations both components contribute negatively to the average productivity, although results indicate that, in general, these areas suffer more for the lack of firm-specific efficiency conditions than for less favourable industry mix.

These results seem to suggest the existence of a premium in terms of operating efficiency, as measured by labour productivity for those firms which are localized in the urban areas compared to other locations and this premium is highly related to firm-specific efficiency, at least when we consider the whole set of activities. Nevertheless, the results we obtain from decomposing our profitability index suggest a much more controversial picture.

In general, we can notice that the profitability premium gained by firms located in urban areas is considerably reduced if it does not disappear completely. We also notice that firm-specific profitability conditions are the main responsible for this pattern, while industry structure tends to counterbalance the negative impact of the firm profitability index. Conversely, other areas which do not exhibit a high density character seem to show a positive premium which is even higher when considering the manufacturing activities: this is the case of the intermediate urban clusters according to the Eurostat DEGURBA classification, the rural areas (excluding those with intensive agriculture) according to the PSR EAFRD classification and the non-urban LMAs which exhibit a different specialization. Furthermore, the higher performance in terms of profitability that we observe in these non-urban clusters is fundamentally guided by firms' specific conditions (i.e., the profitability index is, in general, higher than the regional benchmark), and not by their industry structure which, indeed, plays a negative role if we consider the whole set of economic activities.

This decomposition allows us to derive three considerations. The first is that firms' efficiency conditions do not necessarily imply financial efficiency in terms of profitability, this latter being negatively affected by operating costs that also may depend on the extent to which localization (congestion) factors are at work. The second consideration is that firms localized in the urban areas earn a premium in terms of productivity, but their profitability conditions seem to be constrained compared to other locations. The third is related to the specific role played by firm-specific efficiency in affecting aggregate performance at the local level. Results suggest that firms' operational efficiency is crucial in determining aggregate productivity of high-density urban areas, while it is in the less urbanized locations that the financial efficiency at the plant level contributes most to aggregate profitability conditions.

5. Plant-level analysis

5.1 The empirical model

In this section, we explore the evidence provided in Section 4 further by assuming a micro-level perspective. Results at the aggregate level show that an urban - non-urban divide exists regardless of the geographical classification used; thus we decide to concentrate the analysis at the plant level on the Eurostat DEGURBA classification which has the advantage of being internationally harmonized. Our specific aim is that to understand the extent to which plant-specific differences in terms of both productivity and profitability are affected, on the one hand, by individual heterogeneity and, on the one hand, by the sectoral and/or geographical context in which they operate.

We will apply a Hierarchical Linear Models approach which allows us to model the means of our focus variables by taking into consideration the nested structure of our data that are grouped into sectoral/geographical clusters and, thus, may exhibit nonconstant variability. Thus, the empirical model can be written as:

$$y = X\beta + Zu + \epsilon \qquad 1)$$

Where y is an *Nx1* column vector, the plant-level performance indicator, X is an *Nxp* matrix of the p explanatory variables, β is a *px1* column vector of the regression coefficients, Z is an *Nxq* design matrix for the q random effects, i.e. the random complement to the fixed effects X, u is a *qx1* vector of the random effects. We assume that $\varepsilon \sim N(0, \sigma)$ is an *Nx1* column vector of residuals and $u \sim N(0, G)$, where G is the variance-covariance matrix of the random effects. The total number of observations at the plant level can be expressed as follows:

$$N = \sum_{a=1}^{3} \sum_{s=1}^{75} n_{si} \qquad 2)$$

where i identifies the plant, s_i identifies the sector at the 2 digit Ateco classification and a identifies the DEGURBA areas. Thus, the value of q depends on the hierarchical structure *(sa)* that we decide to adopt with respect the sectoral and geographical levels of analysis.

Equation 1) may be expressed in explicit notation as follows:

$$y_{i(sa)} = \beta_{0(sa)} + \sum_{j=1}^{p} \beta_j X_{ij} + \varepsilon_{i(sa)} \qquad 3)$$

where:

$$\beta_{0(sa)} = \gamma_{00} + u_{0(sa)} \qquad 4)$$

Thus, we assume a grouping structure of our data and that every group of business units has its intercept, given by expression 4) while the β_p coefficients are fixed across clusters.

Firstly, the estimation of an "unconditional" model, i.e. without fixed effects, allows us to consider the total variability across economic units as being determined, on the one hand, by between-group variations and, on the other hand, by between plants (within-group) variations. The former component is variation due to the specific characteristics of the group within the firm operates, being these groups represented by the sector of activity, or the area of localization, or a combination of these two characteristics; the latter component represents unexplained residual variance because it is due to factors directly linked to firm-specific heterogeneity conditions which determine variations with respect to the mean performance. The empirical literature available so far tends to stress the fact that individual-specific differences (within-group variance) are the most relevant factor and conclusions at the aggregate level from the previous section are in line with this view. Nevertheless, the econometric analysis at the individual level preformed in this section is aimed at specifying the most suitable random effect structure. Secondly, by considering the impact of selected fixed effects in the model it is possible to estimate a "conditional" specification in order to assess the relevance of specific covariates in reducing the amount of unexplained residual variance.

5.2 The explanatory variables: firm-level characteristics and spatial determinants

Fixed effects are assumed to be both firms' specific characteristics in terms of age, size, corporate group membership and internationalization. This information is derived from the *Frame territoriale SBS.* Descriptive statistics of the full set of variables used in the econometric analysis are reported in Table 5.

In addition to firm-specific characteristics, we include in our model two additional regressors which aim at exploring the role played by agglomeration economies. Measuring agglomeration economies and their impact on a firm's economic performance is not an easy task. The empirical literature available so far have proposed different approaches, which are very often constrained by data availability and, thus, provide controversial results. Much of the analyses have tried to distinguish localization economies from urban economies. The former have been frequently identified in terms of location quotient or own-industry size, in absolute and relative terms, while the measurement of the latter rests on a more heterogeneous set of indicators which may capture industrial diversity, in terms of industry concentration (Hirschman–Herfindahl index) or inequality (Gini index) or the spatial scale in terms of population, employment or other socio-economic characteristics of the area.

In our analysis, we propose a more comprehensive approach. By using a broad set of municipality-level indexes we apply a data mining technique in order to find the best set of variables which discriminate among area groups. More specifically, we use a multi-group discriminant analysis approach (Fisher, 1936), which allows us to obtain a model to predict local area membership. These variables are latent dimensions which were derived from a large set of statistical information at the municipality level. For space constraints here and for the sake of completeness we describe in the Appendix the data used for the discriminant model, and we provide a detailed discussion of the results.

Table 5 - *Plant-level analysis: descriptive statistics*

Variable	type	description	Mean	Std. Dev.	Min	Max
size	c	number of employees at the plant level	3.95	27.18	0	12175
age	categ	1: 0-2 yrs; 2: 3-5 yrs; 3: 6-10 yrs; 4: 11-15 yrs; 5: 16-20 yrs; 6: 21-25 yrs; 7: 26-30 yrs; 8: 31+ yrs	4.03	2.29	1	8
internationalization	0/1	1 if the firm sells its products in the international market	0.09	0.29	0	1
productivity	c	the ratio of the plant-level productivity to the mean productivity at the sectoral level (Ateco 2 digits)	1.32	1.38	-1.62	17.59
profitability	c	the ratio of the plant-level profitability to the mean productivity at the sectoral level (Ateco 2 digits)	0.89	0.70	-12.64	17.85
High density	0/1	1 if the plant is localized in the high density areas (DEGURBA classification)	0.49	0.50	0	1
Urban	0/1	1 if the plant is localized in the urban areas (DEGURBA classification)	0.42	0.49	0	1
Rural	0/1	1 if the plant is localized in the rural areas (DEGURBA classification)	0.09	0.29	0	1

5.3 Results

5.3.1 The unconditional specification

The analysis of variance reported in Table 6 allows us to determine whether a multilevel approach is justified and then if this is the case, which type of hierarchical structure is appropriate for our data. Firstly, we consider a two-level structure, where the groups are represented, alternatively, by the sectors of activity and the localization areas. We can note that the within-group component explains the most part of total variability for both productivity and profitability. The between-area variance is not significant at the conventional level for both productivity and profitability, while the between-sector variance significantly explains a not negligible part of total variance which is more than doubled when considering profitability (5.4%).

By cross-classifying the sectoral and geographical dimension, we obtain our preferred structure, given that both characteristics are expected to affect data variability jointly. This signifies that the between variability may be further decomposed in three components: the variance between sectors averaged over all areas (σ^2_{0s0}), the variance between regions averaged between all sectors (σ^2_{00a}) and the cross-classified sectors/areas variance (σ^2_{0sa}). The residual part consists of the within sectors and areas individual variability ($\sigma^2_{\varepsilon sa}$). We can note that the variance contribution associated with interaction between sectors and areas is significant for both the productivity and profitability models (respectively 1.1% and 2.1%). Given that the contribution due to the between-area variance continues to be not significant we decided to drop it in the last model structure on the right, where the between-group structure is represented by two random components, one for the sectors and the other one for the combination of sectors and areas. Also, note that the cross classified structure is appropriate as it reduces the contribution of unexplained plant-level variability, while both sectoral and joint sectoral/localization random effects significantly affect our data structure.

This latter cross-classified specification represents the starting structure for developing the plant-level model in terms of both productivity and profitability. In the following section, we will discuss the full models that, according to the formalization expressed in Equation 3) will include the explanatory variables previously described.

5.3.2 Fixed effects specifications

The productivity model

In order to identify the best specification for productivity, we discuss three alternative models which allow us to compare the results and, thus, test for the robustness of our findings.

The first two models assume the cross-classified specification (2) provided by Table 7, where two random intercepts are included: one in order to take into account variance between sectors averaged over areas and one in order to take into consideration variance between cross-classified sectors and areas. Model 1 and Model 2 differ in that the first model is augmented with strictly firm-specific covariates, while the second one also includes the two proxies for agglomeration economies.

In Model 1 the inclusion of fixed effects increases the fit of our model given that both the restricted Log Likelihood and the AIC test are reduced with respect to the unconditional specification. We can also note that the share of residual unexplained variance ($\sigma^2_{\varepsilon sa}$) reduces from 96.9% to 86.3%, a signal that the set of firm-specific variables are appropriate to capture firm-specific variability within groups. Also, it is worth noting that the proportion of variance between sectors is substantially increased when adding fixed effects to the model, thus suggesting that the full model is more able to capture the expected correlation in productivity between plants which operate in the same sector of activity. A further improvement in the

Table 6. *Economic performance at the plant level: hierarchical linear regression (unconditional specification)*

ReML estimations of variance components	Two level structure				cross-classified structure (1)				cross classified structure (2)		
	between areas	within areas	between sectors	within sectors	between sectors over areas	between areas over sectors	between cross-classified sectors/ areas	within areas and sectors	between sectors over areas	between cross-classified sectors/ areas	within areas and regions
Variance components	σ^2_{0a}	σ^2_{ε}	σ^2_{0s}	σ^2_{ε}	σ^2_{0s0}	σ^2_{00a}	σ^2_{0sa}	$\sigma^2_{\varepsilon(sa)}$	σ^2_{0s0}	σ^2_{0sa}	$\sigma^2_{\varepsilon(sa)}$
% of total variance explained											
Productivity	1.1	98.9	2.5	97.5	1.5	0.7	1.1	96.7	1.3	1.9	96.9
Profitability	0.2	99.8	5.4	94.6	4.0	0.0	2.1	93.9	4.0	2.1	93.9
FIT STATISTICS											
-2 Res Log Likelihood	2,786,998		2,777,904		2,772,633				2,772,680		
AIC (Smaller is Better)	2,787,002		2,777,908		2,772,633				2,772,680		

All variance components are significant at the 1% level, except for the between areas variances σ_{0a} and σ_{00a} which are not significant at the conventional levels of significance

Table 7. *Productivity at the plant level: the mixed model (fixed and random effects)*

Conditional model	Model 1			Model 2			Model 3		
FIXED EFFECTS	Coeff.	St. err.	Pr>|t|	Coeff.	St. err.	Pr>|t|	Coeff.	St. err.	Pr>|t|
intercept	0.3869	0.05973	<.0001	0.4058	0.05934	<.0001	0.4116	0.06055	<.0001
area actractiveness				0.08557	0.01064	<.0001	0.09095	0.0107	<.0001
(dis)economies of agglom.				-0.00215	0.01088	0.8431	-0.00152	0.01095	0.8896
size	0.1998	0.002587	<.0001	0.2019	0.00259	<.0001	0.2027	0.00258	<.0001
group	0.9092	0.006035	<.0001	0.8965	0.00604	<.0001	0.8933	0.00603	<.0001
internazionalization	0.4027	0.006138	<.0001	0.3999	0.00613	<.0001	0.3964	0.00613	<.0001
age	0.06345	0.000679	<.0001	0.064	0.00068	<.0001	0.06458	0.00068	<.0001
High density							0.01026	0.00441	0.0199
Rural							-0.01053	0.0059	0.0745
RANDOM EFFECTS	Coeff.	St. err.	Pr>|Z|	Coeff.	St. err.	Pr>|Z|	Coeff.	St. err.	Pr>|Z|
Models 1-2: σ^2_{0s0}; Model 3: σ^2_{0s}	0.2498	0.0445	<.0001	0.2496	0.04368	<.0001	0.2645	0.04527	<.0001
Models 1-2: σ^2_{0sa}	0.02161	0.003236	<.0001	0.01205	0.00192	<.0001			
Models 1-2: $\sigma^2_{\varepsilon(sa)}$; Model 3: σ^2_{ε}	1.7049	0.002695	<.0001	1.7018	0.00269	<.0001	1.7063	0.0027	<.0001
FIT STATISTICS									
-2 Res Log Likelihood		2,700,269			2,698,761			2,700,624	
AIC (Smaller is Better)		2,700,275			2,698,767			2,700,628	
pseudo R^2		0.109			0.111			0.108	
Number of observations		849,533			849,533			849,533	
Number of sectors		75			75			75	
Number of areas		3			3			3	

Table 8. *Profitability at the plant level: the mixed model (fixed and random effects)*

Conditional model	Model 1			Model 2			Model 3			Model 4		
FIXED EFFECTS	Coeff.	St. err.	Pr>|t|	Coeff.	St. err.	Pr>|t|	Coeff.	St. err.	Pr>|t|	Coeff.	St. err.	Pr>|t|
intercept	1.300	0.032	<.0001	1.298	0.032	<.0001	1.296	0.032	<.0001	1.410	0.033	<.0001
area actractiveness				0.005	0.005	0.360	0.006	0.005	0.276	0.008	0.005	0.134
(dis)economies of agglom.				-0.018	0.005	0.001	-0.019	0.005	0.001	-0.018	0.005	0.001
size	-0.215	0.001	<.0001	-0.216	0.001	<.0001	-0.215	0.001	<.0001	-0.337	0.002	<.0001
group	-0.126	0.003	<.0001	-0.124	0.003	<.0001	-0.127	0.003	<.0001	-0.298	0.004	<.0001
internazionalization	-0.073	0.003	<.0001	-0.073	0.003	<.0001	-0.075	0.003	<.0001	-0.280	0.005	<.0001
age	-0.003	0.000	<.0001	-0.003	0.000	<.0001	-0.003	0.000	<.0001	-0.012	0.001	<.0001
size*group										0.117	0.002	<.0001
size*internazionalization										0.120	0.003	<.0001
size*age										0.009	0.000	<.0001
High density							-0.005	0.002	0.032	-0.006	0.002	0.004
Rural							0.022	0.003	<.0001	0.022	0.003	<.0001
RANDOM EFFECTS	Coeff.	St. err.	Pr>|Z|	Coeff.	St. err.	Pr>|Z|	Coeff.	St. err.	Pr>|Z|	Coeff.	St. err.	Pr>|Z|
Models 1-2 σ^2_{0s0}; Models 3-4: σ^2_{0s}	0.070	0.014	<.0001	0.070	0.014	<.0001	0.077	0.014	<.0001	0.076	0.014	<.0001
Models 1-2: σ^2_{0sa}	0.010	0.001	<.0001	0.010	0.001	<.0001						
Models 1-2: $\sigma^2_{\varepsilon(sa)}$; Model 3-4: σ^2_{ε}	0.449	0.001	<.0001	0.449	0.001	<.0001	0.450	0.001	<.0001	0.446	0.001	<.0001
FIT STATISTICS												
-2 Res Log Likelihood		1,729,004			1,728,872			1,730,827			1,722,662	
AIC (Smaller is Better)		1,729,010			1,728,878			1,730,831			1,722,666	
pseudo R^2		0.070			0.071			0.069			0.077	
Number of observations		849,533			849,533			849,533			849,533	
Number of sectors		75			75			75			75	
Number of areas		3			3			3			3	

model fit is obtained in Model 2, where we have also added the two proxies for area attractiveness and (dis)economies of agglomeration.

Although the interaction between sectors and areas significantly contribute to total variability, its influence is quite negligible (1.9% in the unconditional model, which reduces to less than 1% in Model 2). Based on this consideration, we decide to include Model 3 which differs from the previous two specifications on the ground that here we assume a two-level structure, with the random intercept capturing the variability between sectors. The localization areas are thus removed from the random components and included among the set of explanatory variables. This change is equivalent to assuming that localization, which is introduced in Model 3 by adding a set of dummy variables (High density, Urban and Rural, with Urban as reference), can affect the predicted mean of productivity but not the correlation structure of our model. This assumption does not change the model fit substantially compared to the Model 2 specification, and the unobserved within sectors variability does not reduce. As it stands, we are interested in both the Model 2 and Model 3 specifications as both can help in clarifying the role of firm localization to explain plant-level performance.

Once described the model structure, we can move to discuss the fixed effects estimates. Firm-specific characteristics significantly affect productivity with the expected signs and coefficients, which are quite stable across specifications. Larger plants which belong to older firms are expected to be more efficient, while a positive effect is also expected for those productive units included in a corporate group or which take part of an internationalized company, as measured by export propensity at the firm level. The impact of agglomeration economies is positive and significant when considering the first factor (area attractiveness), while we do not observe any significant effect coming from the second factor, which summarizes the role of (dis)economies of agglomeration.

When localization is included among the set of explanatory variables (Model 3), we can note that the results confirm the previous finding at the aggregate level, that is, the productive units which are localized in high-density areas can gain a premium in terms of productivity with respect those units which are localized in the urban areas. Conversely, the plants which are localized in rural areas show a loss in terms of productivity, as suggested by the negative and significant impact of the “Rural” dummy.

The profitability model

We firstly model plant-level profitability by using the same set of explanatory variables used for the productivity estimates, which allows us to have a purposive base of comparison. Thus Models from 1 to 3 in Table 8 correspond to the same specifications previously discussed. Much of the conclusions concerning the appropriateness of a random effect structure, which also includes an interacted effect at the sector and area level, are confirmed. In particular, by assuming a two-level structure at the sector level and including the localization areas among the set of explanatory variables (Models 2 and 3) we can notice that financial plant efficiency is negatively related to the *High density* dummy, thus confirming the possible relevance of negative externalities related to more congested locations. Conversely, and in line with the analysis performed at the aggregate level, being localized in scattered (rural) areas may reduce these negative externalities as suggested by the negative and highly significant coefficient associated to the *Rural* dummy. This line of reasoning is also reinforced if we move to consider the impact of the two additional factors aimed at capturing the extent of agglomeration economies. Results suggest that the role of area attractiveness is not significant, while the latent factor that captures the impact of (dis)economies of agglomeration turn out negative and significant. This evidence, when associated with those obtained for the productivity model, suggests that, although the presence of economies of agglomerations may positively affect individual efficiency and, possibly, offset the adverse effects of congestion costs, the impact on firms’ profitability is much more controversial given that the presence of agglomeration costs may counterbalance the potential benefits arising from localization economies.

Previous empirical investigations underline and estimate the impact of urban costs. In particular, Combes et al. (2018) estimate the elasticity of such costs to the city population in French urban areas. These costs are driven by land costs which also reflect commuting cost and other urban (dis-)amenities.

In general, agglomeration matters as it encourages firms and, particularly, new businesses to locate in more clustered areas (Artz et.al (2016), but at the same time one has to take into consideration the counter effect provided by increasing urban costs, which may counterbalance agglomeration economies in the long run in the absence of any policy intervention.

The comparison between profitability and productivity conditions at the plant level highlights other relevant differences which are related to the role played by firm-specific heterogeneity. We can note that plant size, firm's age and the other characteristics in terms of group membership and internationalization propensity all harm profitability. As concern size, one would expect a positive impact given that larger units are expected to be more efficient and, thus, able to grow and capture larger market shares. Nevertheless, our evidence suggests that this mechanism is not operational, thus confirming previous evidence on the Italian manufacturing firms which also emphasizes the inability of small and medium-sized firms to effectively expand their size (Bartoloni, 2013; Bartoloni and Baussola, 2019).

One should also investigate further the negative impact of the other individual characteristics. For this reason, we estimate Model 4 which also includes among the set of fixed effects interacted coefficients, to gain insights about the possible moderating role that plant size may play on the other individual characteristics included in this specification. Results suggest that this hypothesis is correct as the interacted coefficients are all positive and highly significant while, at the same time their inclusion substantially increases the model fit. This evidence reinforces the importance of firms'growth determinants and, thus, plant size in enhancing profitability conditions directly and, also, via the moderating effect that they can cause on other firm-specific characteristics. In particular firm's age, which can capture learning and selection mechanisms (Jovanovic, 1982; Ericson and Pakes, 1995;

Pakes and Ericson, 1989) exerts a positive effect only for larger plants, while profitability is penalized for those productive units which are younger and mature.

The propensity to belong to a corporate group (group) may positively affect individual efficiency conditions, while the propensity to sell products in international markets (internationalization) is expected to enhance earning conditions by increasing market shares and increase turnover. Nevertheless, these mechanisms exert a positive effect on individual profitability limited to larger plants, while smaller units are likely to be affected negatively by operating costs associated with the attempt to set up more structured and complex ownership, which may also entail a deeper international orientation.

6. Conclusion

The external environment in which firms operate does affect businesses' performance. The role of agglomeration externalities in determining firms' efficiency conditions has been widely analysed. In this framework, an urban – non-urban divide turns out as a possible limitation to economic growth at the regional level.

This study explores the patterns of business performance at the local level by using a census source of firms' microdata. Productivity and profitability conditions have been analysed by adopting both an aggregate perspective at the sub-regional level and a micro-econometric approach at the plant level. The former approach is aimed at exploring aggregate patterns of business performance by adopting a decomposition technique to document spatial variation of both productivity and profitability. The micro-econometric approach instead investigates firms' performance and its determinants.

Evidence at the aggregate level indicates that an urban – non-urban divide exists regarding operative efficiency regardless of the spatial aggregation considered. When the Eurostat DEGURBA classification is adopted, firms localized in the high-density urban centres gain a productivity premium equal to +13 p.p. compared to the regional average, which is even higher when we consider specifically the local manufacturing units. However, this premium vanishes when we consider the profitability conditions at the local level: firms localized in high-density urban areas are less profitable compared to the regional average (-1.2 p.p.) and only firms localized in the intermediate areas gain a profitability premium which is equal to +2.8 p.p. and raises to +2.5 p.p. for the manufacturing local units.

Localization factors in high-density urban areas, as proxied by the industry composition index, play a significant role in affecting business performance - particularly concerning profitability. A more favourable industry mix tend indeed to counterbalance the negative contribution derived from firm-specific profitability conditions partially.

Firm-specific performance is the crucial determinant of aggregate productivity in high-density urban locations: our estimations suggest that the productivity premium gained because of firm-specific productivity conditions is equal to +11 p.p. However the contribution derived from firm-specific profitability conditions is also crucial as concerns aggregate profitability in the intermediate areas, and even more in rural sub-areas, where it contributes greatly to counterbalance the negative effect of a less favourable industry mix.

The micro-econometric analysis pinpoints the factors that most contribute to explain firms' performance observed at the aggregate level. We find that firm localization significantly explains both productivity and profitability conditions when it is included as a random effect in conjunction with sectoral effects. In addition, when localization is included among the set of fixed effects, results confirm the aggregate evidence that being localized in a high density urban (rural) location positively (negatively) affects business productivity while the opposite is observed when profitability is considered. Also, localization economies turn out to be significant determinants, albeit with contrasting effects. Indeed, an area attractiveness exerts a positive and significant impact on firm productivity at the local level, whereas diseconomies of agglomeration negatively affect profitability conditions. This result is quite relevant for a regional policy perspective as it points out the need to enhance economic growth at the local level not only by reducing the efficiency gap that we observe among areas with different level of urbanization, but also by creating those structural conditions that facilitate and encourage earnings and thus profitability.

Tackling agglomeration costs becomes, therefore, a priority if the productivity gains enjoyed by firms localized in urban areas could be entirely transferred into higher profitability and thus, new investment opportunities and ultimately growth. This factor is crucial as it underlines how relevant may be endogenous conditions (firms' profitability) vs exogenous factors, say, public investment, to affect local areas' growth. Thus, firms' profitability appears is crucially affected by agglomeration costs and other negative externalities that our study, in line with the theoretical predictions, has documented.

Among individual heterogeneity conditions we have shown that the size of the local unit and other competitive factors such as group membership, the propensity to internationalization, firm's age - as a proxy of knowledge accumulation- play a positive role in enhancing efficiency at the plant level. As concerns profitability, the picture which emerges is more complex, and the same factors enhancing productivity conditions turn out to be relevant for financial efficiency only when the size of the local unit is adequate to spread out their positive effects.

Thus, from a policy perspective, actions devoted to stimulating firm growth at the plant level are recommended in order to take advantage of competitive opportunities. These actions may entail focusing on a balance between regional and sub-regional tax policy, as this latter may be more focused on cost reduction – thus directly affecting firms' profitability – whereas the former may be focused on public investment devoted to creating economic, social and physical infrastructures.

Appendix- Defining agglomeration economies: a discriminant approach

The model consists of a set of canonical linear discriminant functions (the maximum number is equal to the number of groups minus one) which are orthogonal and, thus, their contributions to the discrimination between groups cannot overlap.

The discriminant analysis was applied to the Eurostat Urban-Rural classification (DEGURBA) which consists of three local areas: high-density urban centers, intermediate urban clusters and rural areas. We have two discriminant dimensions, as follows:

$$F_1 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + \varepsilon$$
$$F_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \mu$$

where, $F_{1,2}$ are the latent variables, $X_2, \ldots X_n$ are the *n* independent variables, ε and μ are the error terms which are independent of each other and α_i and β_i are the discriminant coefficients. Each independent variable is available at the municipality level and for simplicity in the notation we omit the municipality index.

These functions may be interpreted as latent dimensions of the observed group structure (the qualitative depended variable) which are able to capture the most part of total variation in the original set of data at the municipality level and that may be more conveniently used and interpreted in terms of agglomeration (dis)economies. In order to derive the latent factors, the sets of α_i and β_i coefficients are used to calculate the function score for each observation (municipality). These scores are calculated in the same matter as a predicted value from a linear regression by using standardized coefficients and standardized variables.

We report in Table A1 the list of variables used as relevant dimensions to be included in the discriminant analysis together with descriptive statistics. A preliminary selection from a larger set of variables was performed in order to discard redundant information. Results reported in Table A2) indicate that there are two latent dimensions, both of which are statistically significant in discriminating among the three groups of local areas. Function 1 and Function 2 may be interpreted as projections of the original set of independent variables that best separate between sub-regional locations. Canonical correlations, eigenvalues and variance proportions are strictly related and provide an indication of the discriminating power of the discriminant functions: the ratio of the two eigenvalues is equal to three, thus suggesting that the first dimension is three-time more

powerful compared to the second dimension in discriminating among groups.
The canonical correlations reported in the table identifies the strength and direction of the relationship between observed variables and the two latent factors and, thus provide the information which is needed for capturing the underlying meaning to be assigned to each latent factor. We can note that much of the information provided by the set of independent variables at the municipality level is highly and positively associated with the first factor. This information includes the density of economic activities (v1, v2), the skill potential (v5, v6, v10), the relative size of main economic sectors (v8, v9) and the local attractivity which may be analysed both in terms of the density of construction activity (v3) and building expansion (v4, v16) and in terms of commuting flows (v15). Variables associated with these dimensions present high factor loadings and, thus, we decide the label this factor "area attractivity". The other factor, although significant, has lower explanatory power, as suggested by the factor loadings reported in the table. Nevertheless, we can note that some characteristics included in the set of explanatory variables are highly associated with this latent factor (and not to the other one). These characteristics deal with mobility aspects and possible congestion costs: mobility with private vehicles (v11) correlates negatively while collective mobility together with long mobility (v12, v13), which may reduce traffic intensity, correlates positively. Both density and relative size of non-industrial activities (v2, v9) correlate positively, while the relative size of industry (v8) correlates negatively. Also note that the relative size of the agricultural sector (v7) presents a positive, although mild, loading on factor 2 but is highly negative on factor 1. In addition, the density and share of business services (v2, v9) correlate positively as well as the variable reflecting the contribution to value-added of services expenditures undertaken by firms at the municipality level (v17). These loadings and their signs allow us to label factor 2 "economies of agglomeration" which become diseconomies when the sign is negative.

Area attractiveness and economies of agglomeration are the two latent factors that we use in the econometric investigation in order to capture the role of external economies in affecting economic performance at the local level.

Table A1. *Variables at the municipality level used for the discriminant analysis*

Variables	Mean	Std. Dev.	Min	Max
Density of industrial activities[a]	11.6	15.4	0.0	116.4
Density of advanced business services[a]	10.4	21.6	0.0	475.9
Density of construction activities[a]	5.9	7.7	0.0	62.6
Index of building expansion in centers and residential areas	12.1	8.8	0.0	66.4
Share of adults with high school diploma or degree	49.2	9.0	16.5	83.5
share of young adults with a bachelor's degree	17.7	7.7	0.0	66.7
Share of agricultural employment	4.5	5.1	0.0	54.1
Share of industrial employment	38.8	9.3	3.5	75.0
Share of non-commercial tertiary employment	38.7	8.1	16.5	74.4
Share of high and medium skilled employment	27.3	6.7	1.2	61.2
Mobility with private vehicles	68.0	7.4	26.5	92.9
Mobility with collective transport services	11.1	3.7	0.0	36.8
Long mobility	6.8	3.3	0.0	30.4
Per capita income	15246.0	2436.8	3845.3	30427.9
Attractivity index	27.1	12.8	0.0	92.3
Average house prices (m2)[b]	1057.0	329.4	0.0	3640.2
Costs for the purchase of services to value added[a]	0.7	0.5	-1.3	13.3

Data sources: ISTAT, 8000 Census and A misura di comune;
[a] *Our calculations on ISTAT, Frame SBS territoriale*
[b] *The National Revenue Agency*

Table A2. *Canonical linear discriminant analysis*

Factor	Canonical correlation	Eigenvalue	Variance proportion	Variance cumulated	Likelihood Ratio
1	0.73	1.13	0.75	0.75	0.34
2	0.52	0.38	0.25	1.00	0.73

	F statistic	df1	df2	Prob>F
1	63.42	34	3,020	0.0000
2	35.51	16	1,511	0.0000

Ho: this and smaller canon. corr. are zero

Canonical correlations

Variables	Factor 1	Factor 2
Density of industrial activities -v1	0.6886	0.1474
Density of advanced business services - v2	0.5134	0.4292
Density of construction activitiesa - v3	0.6739	0.1555
Index of building expan. in centers and residential areas - v4	0.1775	-0.0140
Share of adults with high school diploma or degree -v5	0.3357	0.3258
share of young adults with a bachelor's degree - v6	0.3716	0.1557
Share of agricultural employment - v7	-0.5088	0.1594
Share of industrial employment - v8	-0.0254	-0.5658
Share of non-commercial tertiary employment - v9	0.3561	0.5283
Share of high and medium skilled employment - v10	0.4963	0.1480
Mobility with private vehicles - v11	0.1519	-0.4570
Mobility with collective transport services - v12	0.2230	0.6113
Long mobility - v13	-0.0451	0.2484

Variables	Factor 1	Factor 2
Per capita income - v14	0.3978	0.2569
Attractivity index - v15	0.3560	-0.0089
Average house prices (m^2) - v16	0.2260	0.1145
Costs for the purchase of services to value added - v17	0.1214	0.1197
Group means on canonical variables		
High density	1.9241	1.7288
Urban	0.7402	-0.4901
Rural	-1.1308	0.1999

Classification results

Original (total)/predicted (cells)	High density	Urban	Rural	Total
High density	**92**	29	4	125
	73.6	23.2	3.2	100
Urban	70	**536**	114	720
	9.72	**74.44**	15.83	100
Rural	8	94	**582**	684
	1.17	13.74	**85.09**	100
Total	170	659	700	1,529
	11.12	43.1	45.78	100
Priors	0.3333	0.3333	0.3333	

Acknowledgements

This work is a result of a collaboration developed within the project "Business demography during the great recession: patterns of resilience and productivity dispersion", which is part of a three-year strategic research plan promoted by the National Institute of Statistics (ISTAT). The empirical part of this research was carried out at the ISTAT Regional Office for Lombardy in Milano. We would like to acknowledge the support of Roberto Monducci and Stefano Menghinello, Directors for Economic Statistics and Rosalia Coniglio, Director of the Regional Office for Lombardy. The views expressed in this paper are those of the authors and do not necessarily reflect the views of ISTAT.
We would like to thank Elena Longoni, Alberto Vitalini and Lorena Viviano for precious help and the participants at the 2019 RSA Annual Conference, Santiago de Compostela, Spain, 5-7 June, for useful comments and suggestions.

References

Andersson, M. and H. Lööf. (2011) Agglomeration and productivity: evidence from firm-level data. *The annals of regional science*, 46(3), 601-620.

Andrews, D., C. Criscuolo and P.N. Gal. (2015) Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries, *OECD Productivity Working Papers* 2, OECD Publishing.

Archibugi, D., A. Filippetti and M. Frenz. (2013) Economic crisis and innovation: Is destruction prevailing over accumulation? *Research Policy*, 42(2), 303-314.

Artz, G. M., Kim, Y., & Orazem, P. F. (2016) Does agglomeration matter everywhere?: new firm location decisions in rural and urban markets. *Journal of Regional Science*, 56(1), 72-95.

Audretsch, D.B. and D. Dohse. (2007) Location: A neglected determinant of firm growth. *Review of World Economics*, 143(1), 79-107.

Baert, L. and J. Reynaerts (2018) Agglomeration, congestion and plant level productivity in Belgium. *VIVES discussion papers*, 4 March, Research Center for Regional Economics, KU Leuven.

Baldwin, J.R., D. Beckstead, W. Mark Brown and D.L. Rigby. (2008) Agglomeration and the geography of localization economies in Canada. *Regional Studies*, 42(1), 117-132.

Baldwin, R.E. and T. Okubo. (2006) Agglomeration, offshoring and heterogenous firms. *CEPR discussion paper*

Bartoloni, E. (2013) Profitability and innovation: new empirical findings based on italian data 1996-2003. *Rivista Internazionale di Scienze Sociali*, 2, 137-170.

Baussola, M., E. Bartoloni and L. Bagnato (2018) *Waiting for Godot: the Failure of SMEs in the Italian Manufacturing Industry to Grow*. Università Cattolica del Sacro Cuore, Dipartimento di Scienze Economiche e Sociali, Quaderno n.132/marzo 2018, Ed. Vita e Pensiero. p. 1-31, ISBN: 9788834335895.

Beaudry, C. and S. Breschi. (2003) Are firms in clusters really more innovative? *Economics of innovation and new technology*, 12(4), 325-342.

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

Boschma, R. (2005) Proximity and innovation: a critical assessment. *Regional Studies*, 39(1), 61-74.

Brinkman, J.C. (2016) Congestion, agglomeration, and the structure of cities. *Journal of Urban Economics*, 94, 13-31.

Carlino, G. and W.R. Kerr. (2014) Agglomeration and innovation. *NBER Working Paper* No. 20367 August 2014.

Centro Studi Confindustria e Prometeia. (2014). Esportare la dolce vita. Il bello e ben fatto italiano nei nuovi mercati, N.5 maggio.

Cingano, F. and F. Schivardi. (2004) Identifying the sources of local productivity growth. *Journal of the European Economic association*, 2(4), 720-742.

Combes, P. P., Duranton, G., & Gobillon, L. (2018) The costs of agglomeration: House and land prices in French cities. *The Review of Economic Studies*, 86(4), 1556-1589.

Combes, P. and G. Duranton. (2006) Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1), 1-28.

Combes, P., T. Magnac and J. Robin. (2004) The dynamics of local employment in France. *Journal of Urban Economics*, 56(2), 217-243.

Duranton, G. and D. Puga. (2004). Micro-foundations of urban agglomeration economies. In Anonymous *Handbook of regional and urban economics*: Elsevier, 2063-2117.

Ericson, R. and A. Pakes. (1995) Markov-perfect industry dynamics: A framework for empirical work. *The Review of economic studies*, 62(1), 53-82.

Feldman, M.P. and D.B. Audretsch. (1999) Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2), 409-429.

Filippetti, A. and D. Archibugi. (2011) Innovation in times of crisis: National Systems of Innovation, structure, and demand. *Research Policy*, 40(2), 179-192.

Fisher, R.A. (1936) The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.

Fothergill, S., & Gudgin, G. (1982). Unequal growth: urban and regional employment change in the UK. *Heinemann Educational Publishers.*

Fothergill, S., Kitson, M., & Monk, S. (1985). Urban Industrial Change: The causes of the urban-rural contrast in manufacturing employment trends. *HM Stationery Office.*

Foster, L., J. Haltiwanger and C. Syverson. (2008) Reallocation, firm turnover, and efficiency: selection on productivity or profitability? *American Economic Review*, 98(1), 394-425.

Hall, B.H. (2011) Innovation and productivity. *NBER Working Paper* No. 17178 August 2011.

Henderson, J.V. (2003) Marshall's scale economies. *Journal of Urban Economics*, 53(1), 1-28.

Henderson, J.V. (1974) The sizes and types of cities. *The American Economic Review*, 64(4), 640-656.

ISTAT (2015) *La nuova geografia dei sistemi locali,* 1-204*, ISBN: 978-88-458-1857-8.*

ISTAT (2018) *Risultati Economici delle Imprese a Livello Territoriale: Ampliamento del Dettaglio di Analisi*, Statistiche Report, 13 Giugno 2018.

Jennen, M. and P. Verwijmeren. (2010) Agglomeration effects and financial performance. *Urban Studies*, 47(12), 2683-2703.

Jacobs, J. (1969). *The economy of cities*. New York: Random House.

Jovanovic, B. (1982) Selection and the Evolution of Industry. *Econometrica: Journal of the Econometric Society*, 649-670.

Kukalis, S. (2010) Agglomeration economies and firm performance: the case of industry clusters. *Journal of Management*, 36(2), 453-481.

Manyika, J, Chui, M et al. (2017) A future that works: automation, employment, and productivity, *McKinsey Global Institute*, San Francisco.

Marshall, A. (1920) *Principles of Economics*, London: Macmillan and Co. 8th ed.

Martin, P., T. Mayer and F. Mayneris. (2011) Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2), 182-195.

McCann, B.T. and T.B. Folta. (2011) Performance differentials within geographic clusters. *Journal of Business Venturing*, 26(1), 104-123.

Oosterhaven, J. and L. Broersma. (2007) Sector structure and cluster economies: a decomposition of regional labour productivity. *Regional Studies*, 41(5), 639-659.

Pakes, A. and R. Ericson. (1998) Empirical implications of alternative models of firm dynamics. *Journal of Economic Theory*, 79(1), 1-45.

Paunov, C. (2012) The global crisis and firms' investments in innovation. *Research Policy*, 41(1), 24-35.

Rice, P., A.J. Venables and E. Patacchini. (2006) Spatial determinants of productivity: analysis for the regions of Great Britain. *Regional science and urban economics*, 36(6), 727-752.

Rigby, D.L. and J. Essletzbichler. (2002) Agglomeration economies and productivity differences in US cities. *Journal of Economic Geography*, 2(4), 407-432.

Rizov, M., A. Oskam and P. Walsh. (2012) Is there a limit to agglomeration? Evidence from productivity of Dutch firms. *Regional Science and Urban Economics*, 42(4), 595-606.

Rizov, M. and P.P. Walsh. (2011) Is there a rural–urban divide? Location and productivity of UK manufacturing. *Regional Studies*, 45(5), 641-656.

Robertson, T.S. and H. Gatignon. (1998) Technology development mode: a transaction cost conceptualization. *Strategic Management Journal*, 19(6), 515-531.

Rosenthal, S.S. and W.C. Strange. (2004). Evidence on the nature and sources of agglomeration economies. In Anonymous *Handbook of regional and urban economics*: Elsevier, 2119-2171.

Saito, H. and J. Wu. (2016) Agglomeration, congestion, and US regional disparities in employment growth. *Journal of Regional Science*, 56(1), 53-71.

Stavropoulos, S. and D. Skuras. (2016) Firm profitability and agglomeration economies: An elusive relationship. *Tijdschrift voor economische en sociale geografie*, 107(1), 66-80.

Sweet, M. (2014) Traffic congestion's economic impacts: Evidence from US metropolitan regions. *Urban Studies*, 51(10), 2088-2110.

Syverson, C. (2017) Challenges to mismeasurement explanations for the US productivity slowdown. *Journal of Economic Perspectives*, 31(2), 165-186.

Syverson, C. (2011) What determines productivity? *Journal of Economic literature*, 49(2), 326-365.

United Nation (2018) *World Urbanization Prospects*, DESA Population Division

Van Oort, F. (2002) Innovation and agglomeration economies in the Netherlands. *Tijdschrift voor economische en sociale geografie*, 93(3), 344-360.

Webber, D., N. Curry and A. Plumridge. (2009) Business productivity and area productivity in rural England. *Regional Studies*, 43(5), 661-675.

Printed by
Gi&Gi srl - Triuggio (MB)
November 2019

9788834341476