

The empirics of agglomeration economies: the link with productivity

Marli Fernandes, Sílvia Santos e Ana Fontoura Gouveia*

Abstract

There is a large branch of literature providing empirical evidence on the positive effects of agglomeration economies on productivity. However, for policy makers it is important to understand the role of agglomeration economies at a more micro level, disentangling the effects across industries, firm-level characteristics and time. The present survey reviews this literature, outlining the econometric approaches and methodological challenges. In general, results show that the magnitude of agglomeration economies differ substantially across industries and point to the presence of non-linear effects, also depending on the industry and product life cycles. The channels through which these effects operate may also differ – resulting from specialization externalities (within industries in the same region) and/or urbanisation externalities (across industries in the same region). Overall, the evidence reviewed in this survey highlights the need for policy makers to follow tailor-made approaches and to complement existing evidence with national level studies, maximizing potential productivity gains.

1. Introduction

Economies of agglomeration are the outcome of both economies of scale and network economies that arise when firms (and people) locate near one another. They thus relate to spatial proximity and, as Glaeser (2010) puts it, may be formulated as a reduction in transportation costs in a broad sense, i.e. transportation costs related to goods but also to people and ideas.

Agglomeration economies are in fact an old concept, advanced first by Alfred Marshall in the 1890s, with his book “The Principles of Economics”. According to Marshall’s view, three factors reduce production costs for agglomerated firms: higher availability of skilled labour, higher specialization of suppliers and larger knowledge inflows from competitors. Concerning the first factor, when in a given region jobs are concentrated in the same industry, unskilled workers are more likely to specialise in order to more easily find a job; in turn, firms will save time and money that they would otherwise spend on training. Relatively to the second factor, suppliers of a cluster are more likely to make industry-specific investments, reducing transportation and coordination costs. Finally, concerning knowledge diffusion, Marshall argues that firms with similar products may have different production processes and some of these processes may be more productive than others; over time these practices tend to diffuse to neighbouring firms, improving efficiency and lowering costs. Therefore, Marshall or specialisation externalities refer to externalities from other plants in the same industry.

In this context, Porter (1990) argues that stronger competition in the same market gives incentives for firms to innovate, accelerating the rate of technical progress and hence of productivity growth (Porter externalities). This higher productivity can also be linked to an extensive literature focusing on firm selection (see, for instance, Melitz, 2003, Syverson, 2004 and Baldwin and Okubo, 2006), meaning that the presence of more firms makes competition fiercer, leading the less productive firms to leave and therefore increasing average productivity of those that remain in the market.

While Marshall defend that benefits of agglomeration economies stem from specialisation or competition externalities, Jacobs (1969) defends urbanisation or diversification externalities. The author argues that

* Marli Fernandes – Nova SBE; Sílvia Santos – GPEARI – Ministry of Finance; Ana Fontoura Gouveia (corresponding author ana.gouveia@gpearl.min-financas.pt) – GPEARI/Ministry of Finance and Nova SBE. The opinions expressed are those of the authors and not necessarily of the institutions. Any errors or omissions are the authors' responsibility.

the most important sources of knowledge spillovers are in fact external to the industry in which the firm operates. Moreover, since these knowledge sources are more diverse in urban areas, cities are the best regions for innovative activity. More diverse industrial fabrics create opportunities to imitate, share and diffuse ideas and practices across industries, facilitating search and experimentation in innovation. Therefore, well-functioning infrastructures of transportation and communication, with good proximity to other markets and good access to specialised services in a diversified local production structure allow these externalities to rise. In this sense, urbanisation or diversification externalities refer to externalities from other plants outside the own industry but within the same region.

Given these types of externalities, the theory tells us that we should expect agglomeration to have a positive effect on firm-level productivity, either through specialization or diversification effects¹. This idea has spurred a vast amount of research on the relationship between agglomeration economies and firm level productivity. In this survey we focus on the differences across industries and firms as indeed agglomeration economies are not uniform. Understanding these (heterogeneous) effects is key to inform policy makers, ensuring well designed, targeted policy measures². Obviously, the idea of agglomeration economies needs to be tackled as part of a broader, comprehensive strategy. In fact, agglomeration is a necessary but not a sufficient condition for productivity growth; other factors, like institutions and metropolitan governance, are determinant.

The remainder of this paper is organized as follows: Section 2 addresses methodological issues, also presenting the main estimation challenges, in particular related to endogeneity; Section 3 considers the empirical evidence of agglomeration effects on firm productivity, highlighting the differences across industries, firms and time; finally, Section 4 concludes.

2. Estimation strategies

It is not possible to discuss the estimation of agglomeration economies without first clarifying the empirical strategies and underlying estimation mechanisms that are followed in the literature³. This section presents these strategies so that we can then correctly interpret and discuss estimates. First, we present the use of Total Factor Productivity (TFP) and labour productivity as measures of productivity and the estimation strategies to assess the impact of agglomeration economies. We then turn to endogeneity issues and the approaches to overcome them.

2.1 Estimating productivity and the relation with agglomeration economies

Empiric studies focus on two widely used productivity measures: TFP and labour productivity. In the next subsections, we present the methodologies to compute them and the strategies used in the different studies to relate them with agglomeration economies.

¹ For a review of Marshallian, Porter and Jacobs' externalities, please refer to Beaudry and Schiffauerova (2009). For an overview of the micro-foundations of agglomeration economies, please see Duranton and Puga (2003). We focus on the agglomeration of firms. For an example of a study focusing on the agglomeration of consumers, see for instance Waldfogel (2010). Also, while we focus on the effects on productivity, some studies look at other outcomes such as entrepreneurship (e.g. Rosenthal and Strange, 2010) or export status (indeed, export spillovers may arise when local firms benefit from the proximity of other exporters and there are opportunities for information or technological spillovers. Evidence of these effects is mixed. While Greenaway and Kneller, 2008 and Koenig, 2009 conclude that they are present for the UK and France, respectively, Barrios *et al.*, 2003 and Bernard and Jensen, 2004 fail to detect export externalities for Spain and the US, respectively).

² Agglomeration economies are a large concept and there exist a vast number of reviews of the empirical literature on agglomeration, exploring a particular angle. Puga (2010) surveys the magnitude and causes of agglomeration economies. See Carlino and Kerr (2015) and Feldman (2006) on the relation between agglomeration and innovation. Combes and Gobillon (2014) cover effects on industrial employment and firms' location choices. See Head and Mayer (2004) on the relation between agglomeration and trade. Fujita and Thisse (2013) survey the effects of agglomeration economies on cities, industrial location and globalisation. For an overview of the productivity effects also covered in this survey, see Rosenthal and Strange (2004).

³ For a comprehensive discussion of methodological issues, see Combes and Gobillon (2014).

A. Total Factor Productivity

During the last two decades, several studies measured the impact of agglomeration economies on TFP, which is a widely used productivity measure. The estimation of TFP starts with the firm's production function of the form:

$$Y_{it} = A_{it}F(L_{it}, K_{it}) \quad (1)$$

where Y_{it} is firm output at time t , A_{it} denotes the firm technological level, L_{it} and K_{it} correspond to labour and capital respectively. The exact form of the production function used in the different studies varies: the most widely used formulations are Cobb-Douglas and Translog⁴.

A.1 Cobb-Douglas Production Function

The Cobb-Douglas production function has been the most empirically used production function (e.g. Mitra, 1999; Greenstone *et al.*, 2010; Antonietti and Cainelli, 2011; Martin *et al.*, 2011; Morikawa, 2011; Neffke *et al.*, 2011; Combes *et al.*, 2012; Di Giacinto *et al.*, 2014), although many authors argue that is extremely restrictive as it sets the elasticity of substitution between factors to one (Eberts and McMillen, 1999). It can be written as:

$$Y_{it} = AK_{it}^{\alpha}(l_{it}L_{it})^{1-\alpha} \quad (2)$$

Where l_{it} corresponds to average labour skills. If we assume that the firm i technological level (A_{it}) depends on a firm component, U_{it} , but also on its environment in terms of specialisation and urbanisation economies, we can decompose A_{it} into:

$$A_{it} = (\text{SPE}_{it}^{jz})^{\delta} (\text{URB}_{it}^{jz})^{\gamma} U_{it} \quad (3)^5$$

where SPE_{it}^{jz} is a measure of specialisation externalities and URB_{it}^{jz} is a measure of urbanisation externalities for firm i in industry j and area z . Output value is obtained by $p_{it}Y_{it}$ where p_{it} is the average income of the firm per unit produced. The logarithm of TFP is derived as:

$$\ln p_{it}Y_{it} - \alpha \ln L_{it} - (1 - \alpha) \ln K_{it} = \ln p_{it}A_{it}l_{it}^{\alpha} \quad (4)^6$$

which is equivalent to:

$$\ln p_{it}Y_{it} - \alpha \ln L_{it} - (1 - \alpha) \ln K_{it} = \ln p_{it}(\delta \text{SPE}_{it}^{jz} + \gamma \text{URB}_{it}^{jz})l_{it}^{\alpha} \quad (5)$$

The logarithm of TFP, in the left-hand side, can be related to some local characteristics which define the means through which agglomeration economies operate, such as the inputs used, K_{it} or the technological level, A_{it} .

A.2 Translog Production Function

One of the most used translog production functions in agglomeration economies is based on an inverse demand framework (e.g. Graham, 2007; Graham and Kim, 2008), proposed by Kim (1992). Let us have the following production function:

$$Y = g(U_{it}, S_{it})f(L_{it}, K_{it}) \quad (7)^7$$

where $g(U_{it}, S_{it})$ are influences on production which are Hick's neutral in nature and comprise the effects that arise from urbanisation (U_{it}) and specialisation externalities (S_{it}). The function $g(U_{it}, S_{it})$ is usually assumed to be log-separable in U_{it} and S_{it} .

The production function described can be represented by the translog approximation:

⁴ The Constant Elasticity of Substitution (CES) is also another form of production function, less restrictive than the Cobb-Douglas production function, but it cannot be used to describe the production function of all the firms in the industry, in particular in a context of technological change.

⁵ From Martin *et al.* (2011).

⁶ From Boschma and Frenken (2011)

⁷ From Graham (2007).

$$\log Y_{it} = \alpha_0 + \beta_U \log U_{it} + \beta_S \log S_{it} + \beta_L \log L_{it} + \beta_K \log K_{it} + \frac{1}{2} \gamma_{LL} (\log L_{it})^2 + \frac{1}{2} \gamma_{KK} (\log K_{it})^2 + \gamma_{LK} \log L_{it} K_{it} \quad (8)$$

The total cost of production to the firm i is:

$$wL_{it} + rK_{it} = i \quad (9)$$

where w is the wage rate and r is the price of capital. Firms maximize output subject to the expenditure constraint expressed in (9), in consequence the Lagrange function is:

$$L = g(U_{it}, S_{it})f(L_{it}, K_{it}) + \lambda[i - wL_{it} - rK_{it}]. \quad (10)$$

Assuming that inputs are rented in competitive markets, the first-order condition from the Lagrange function is:

$$\lambda = \frac{\left(\frac{\partial Y_{it}}{\partial L_{it}} \cdot L_{it}\right) + \left(\frac{\partial Y_{it}}{\partial K_{it}} \cdot K_{it}\right)}{i} \quad (11)$$

and substituting equation (10) back into the first order conditions, after rearrangement, yields the inverse input demand equations:

$$\frac{w}{i} = \frac{\frac{\partial Y_{it}}{\partial L_{it}}}{\left(\frac{\partial Y_{it}}{\partial L_{it}} \cdot L_{it}\right) + \left(\frac{\partial Y_{it}}{\partial K_{it}} \cdot K_{it}\right)} \quad (12)$$

$$\frac{r}{i} = \frac{\frac{\partial Y_{it}}{\partial K_{it}}}{\left(\frac{\partial Y_{it}}{\partial L_{it}} \cdot L_{it}\right) + \left(\frac{\partial Y_{it}}{\partial K_{it}} \cdot K_{it}\right)} \quad (13)$$

The inputs demand equations (12) and (13) can be written in cost share form (C_L^i) and (C_K^i) as

$$C_L^i = \frac{wL}{i} = \frac{\frac{\partial \log Y_{it}}{\partial \log L_{it}}}{\frac{\partial \log Y_{it}}{\partial \log L_{it}} + \frac{\partial \log Y_{it}}{\partial \log K_{it}}} \quad (14)$$

$$C_K^i = \frac{rK}{i} = \frac{\frac{\partial \log Y_{it}}{\partial \log K_{it}}}{\frac{\partial \log Y_{it}}{\partial \log L_{it}} + \frac{\partial \log Y_{it}}{\partial \log K_{it}}} \quad (15)$$

Given equations (14) and (15), differentiation of equation (6) yields the cost share equations:

$$C_L^i = \frac{\beta_L + \gamma_{LL} \log L_{it} + \gamma_{LK} \log K_{it}}{\beta_L + \beta_K + (\gamma_{LL} + \gamma_{LK}) \log L_{it} + (\gamma_{KK} + \gamma_{LK}) \log K_{it}} \quad (16)$$

$$C_K^i = \frac{\beta_K + \gamma_{KK} \log K_{it} + \gamma_{LK} \log L_{it}}{\beta_L + \beta_K + (\gamma_{LL} + \gamma_{LK}) \log L_{it} + (\gamma_{KK} + \gamma_{LK}) \log K_{it}} \quad (17)$$

The translog parameters can be estimated by simultaneously estimating equation (8) and equations (16) and (17) as a nonlinear multivariate regression system.

Estimation of the translog function allows isolating the three different sources of scale economies: internal economies of scale and scale effects associated with specialisation and urbanisation economies. Controlling by the input use and the Hick's neutral agglomeration terms permit to measure the amount by which the production function is shifted outwards, given the volume of input use, as a consequence of agglomeration externalities. Thus, the elasticities associated with the agglomeration variables are interpreted as the change in productivity or TFP relatively to agglomeration.

Although the translog function has the advantage of being less restrictive than the Cobb-Douglas function, it is not a perfect alternative. The number of parameters that need to be estimated imposes hard

constraints, since it may lead to collinearity. The translog function is frequently used as a robustness check, after the Cobb-Douglas production function.

B. Labour Productivity

Alternatively to the use of TFP, one can study the effects of agglomeration economies on labour productivity (e.g. Baldwin *et al.*, 2008; Baldwin *et al.*, 2010; Andersson and Lööf, 2011). The basic model can be expressed in a Cobb-Douglas function such as:

$$Y_{it} = A_{it} K_{it}^{\beta_K} L_{it}^{\beta_L} H_{it}^{\beta_H} \quad (18)^8$$

where L_{it} is low-skilled labour and H_{it} is high-skilled labour. Equation (18) may be rewritten such that labour productivity is a function of capital and labour inputs:

$$LP = \frac{Y_{it}}{L_{it}} = A_{it} K_{it}^{\beta_K} L_{it}^{\beta_L - 1} H_{it}^{\beta_H} \quad (19)$$

Hence, labour productivity is a positive function of the amount of capital employed per production worker, the number of low-skilled workers for each production worker, and the size of the firm as measured by the number of high-skilled workers.

Agglomeration phenomena can be assumed to influence firms' technology such that the measure of the potential for agglomeration economies influences A_{it} . Therefore, A_{it} is modelled as follows:

$$\ln A_{it} = \varphi \ln S_{r,t} + x'_{it} \lambda + \varepsilon_{it} \quad (20)$$

where $S_{r,t}$ is the potential for agglomeration economies (e.g. size of region r at time t) while x'_{it} consists of control variables and ε_{it} is an error term, which can be interpreted as capturing technological shocks. The full model is given by:

$$\ln Y_{it} = \varphi \ln S_{r,t} + x'_{it} \lambda + \beta_K \ln K_{it} + (\beta_L - 1) \ln L_{it} + \beta_H \ln H_{it} + \varepsilon_{it} \quad (21)$$

The variable of main interest is $S_{r,t}$ and the objective is to estimate the parameter φ . Theoretically, φ is expected to have a positive sign, since agglomeration economies are expected to have a positive effect on productivity.

2.2 Endogeneity issues

Combes *et al.* (2010) argue that previous literature on the estimation of agglomeration externalities suffers from serious endogeneity problems, mainly unobserved/unmeasured heterogeneity and simultaneity bias.

A first source of endogeneity is related with the unobserved/unmeasured firm's environment variables. In fact, when the agglomeration effect is estimated from a production function, the error term cannot be correlated with the other regression' variables. However, the difficulty to measure all firms' environment variables may violate this assumption, leading to a biased estimation. For instance, input variables, usually labour and capital, do not include precise information about the "quality" or "frequency of use" of these elements, and information regarding other input variables like land, raw materials or energy are in general lacking.

A second source of endogeneity is simultaneity bias. In fact, an economic shock in a region or sector may have positive (negative) consequences in other firms which can again determine the correlation between the errors and the localisation and urbanisation variables.

Therefore, as explained above, agglomeration effects can raise productivity; but an entrepreneur may also seek the most productive locations turning it into an agglomerated area. The difficulty to determine the direction of causality justifies the need to address the different sources of endogeneity to avoid having wrong estimations.

⁸ From Andersson and Lööf (2011).

In general, the endogeneity problem is very difficult to address, in particular due to data limitations. Therefore, although it is not possible to fully overcome it, a good strategy is to compute several robustness checks, as done by most of the papers covered in this survey. Indeed, in the presence of endogeneity effects, ordinary least squares (OLS) methods may lead to biased estimates. Therefore, several papers use GMM techniques (e.g. Henderson, 2003; Martin *et al.*, 2011) to estimate the specification in first differences while using lagged values of variables as instruments, in order to identify the role of local determinants on local outcomes, both in level and first differences. The specification is written in first difference between t and $t - 1$ to eliminate the firm fixed effect and capture time-invariant firm and local effects. This approach is also prone to criticism as some authors consider the assumption of the lagged values being exogenous too strong (e.g. Combes and Gobillon, 2014).

Otsuka and Goto (2015) propose a new strategy to deal with the endogeneity of the agglomeration effects. They use the Solow residual measurements to determine the degree of agglomeration instead of estimating a production function with specific factors representing externalities.

Another strategy to deal with the problem of an endogenous local determinant is the use of quasi-experiments, i.e. changes that induced a sizeable localised shock on a specific determinant which is not directly related to the outcome variable (e.g. Greenstone *et al.*, 2010; Buenstorf and Guenther, 2010). This is achieved by comparing the outcome average variation in places which have experienced a shock with the outcome average variation in places which have not suffered that shock (control group). There is the need to find a control group similar to the treated group, in such way that their unobserved characteristics would have evolved similarly if the shock has not happened.

Finally, it is important to notice that local determinants of agglomeration economies may be endogenous since some missing variables determine them simultaneously with the local outcome; in particular, when there are missing amenities that affect both productivity and local population. Using local fixed effects can be a strategy to deal with this, when having panel data (e.g.: Henderson, 2003; Holl, 2004; Lall *et al.*, 2004; Syverson, 2004; Baldwin *et al.*, 2008; Davis and Weinstein, 2008; Broersma and Oosterhaven, 2009; Greenstone *et al.*, 2010; Andersson and Lööf, 2011; Martin *et al.*, 2011; Neffke *et al.*, 2011; Di Giacinto *et al.*, 2014); however, this strategy has some important drawbacks as it does not deal with missing variables that evolve over time, e.g. new universities are built or improved over the years considering local demand, including firms demand. Also, time invariant local fixed effects do not solve the endogeneity issue related to reverse causality, such that higher expected wages or productivity in a region attract more firms and workers.

3. Agglomeration effects

The concept of agglomeration economies has prompted an extensive literature focused on the relation with firm level productivity. There is indeed wide evidence that agglomeration economies have a positive impact on productivity (see, for instance Otsuka *et al.*, 2009, for Japan, Greenstone *et al.*, 2010, for the US, Andersson and Lööf, 2011, for Sweden and Di Giacinto *et al.*, 2014, for Italy).

Some authors explore the role of the different types of externalities outlined in Section 1. Martin *et al.* (2011), using firm and plant level data, conclude that agglomeration externalities in France take the form of specialisation economies in the short-run, while urbanisation economies are relevant in the longer-term. Baldwin *et al.* (2008), using data for Canada, conclude that all three of Marshall's agglomeration economies (access to buyer-supplier networks, labour market pooling and knowledge spillovers) are relevant to labour productivity across manufacturing plants. In a later study, Baldwin *et al.* (2010), using plant-level data, find again evidence of the positive effects of different types of agglomeration economies on the labour productivity of manufacturing establishments, independently of the plant and firm characteristics. They add that labour mix, meaning the consistency of the match between the local supply and demand for labour across occupations, is the most important of the three Marshall's agglomeration economies.

However, for policy makers it is crucial to further understand these results, in particular given that these effects are not uniform across industries, firms and time. The remainder of this section presents an overview of this literature.

3.1 Effects across industries

Agglomeration economies have heterogeneous effects across industries, as their strength depend on industries characteristics – while there are some industries that greatly benefit from agglomeration externalities, for others these effects are much more modest.

Henderson (2003), focusing on USA plant level data, finds evidence of specialisation externalities in high-tech industries but not in machinery industries. There is no evidence pointing to urbanisation externalities for any of the two industries.

Mitra (1999), using firm level data for two Indian industries (electrical machinery and cotton and cotton textiles), find evidence of a positive association between technical efficiency and city size, although after a certain threshold level city size worked more as diseconomies than economies of scale. This threshold level is lower for electrical machinery industries than for cotton and cotton textiles. Indeed, Lall *et al.* (2004), focusing on industrial sectors in India, find considerable variation in the magnitudes of agglomeration economies. In particular, they conclude that market access and proximity to transport hubs have positive effects in five industry sectors (machine tools and electronics, computer equipment, cotton textiles, beverages and tobacco), while the benefits from specialisation economies are just significant and positive for two industries (printing and publishing and non-metallic mineral products). Similarly, Mitra (2000), using industry-level data⁹, concludes that eleven industries¹⁰ out of nineteen enjoy agglomeration benefits. Among them, seven¹¹ tend to show a decline in the growth of TFP after total population in the state and the share of manufacturing in total urban employment cross a certain threshold.

Graham and Kim (2008) by using firm level data for UK, obtain different values for agglomeration elasticities among three industries: manufacturing has the lowest level of elasticity, followed by the construction industry; the services sector has the highest level of elasticity. Graham (2009) also studies urbanisation and specialisation economies using firm-level TFP, using two digit manufacturing and service industries data for the UK; he finds evidence of specialisation economies in 13 of the 27 sectors examined¹². However, the author also stresses that these externalities tend to exist over small spatial scales and attenuate rapidly with distance. In addition, he finds evidence of urbanisation externalities in 14 sectors¹³.

Morikawa (2011), using establishment-level data for Japan and by estimating the elasticity of firm TFP in service industries, find that productivity rises between 7% and 15% if the municipality population density doubles. Density has important but distinct effects on manufacturing and retail. Otsuka and Goto (2015), using annual data for 47 Japanese administrative divisions, propose a new approach to measure agglomeration economies, based on the use of the Solow residual. Nonetheless, their results are in line with previous literature. More specifically, social overhead capital, which strengths economies of agglomeration, has a positive effect on productivity growth, greater for manufacturing than non-manufacturing industries.

⁹ Literature using industry-level data is scarce. Three explanations for that can be offered: first, assumptions about economic behaviour at the firm level do not necessarily hold at more aggregate levels; second, aggregation may require the imposition of some restrictive assumptions relatively to production technology and the micro data offer superior spatial detail (Graham and Kim, 2008); third, the lack of availability of local data per industry.

¹⁰ Food products; beverages; cotton textiles; woollen textiles; textile products; rubber, petroleum and coal products; non-metallic minerals; basic metals; metal products; transport and parts; and other manufacturing industries.

¹¹ Food products, beverages, cotton textiles, textiles products, woollen textiles, non-metallic minerals, transport and parts, and other manufacturing industries.

¹² In particular, these are larger for business and management consultancy services and computer and related activities, followed by advertising, food manufacture, architecture & engineering, pulp & paper, financial services, hotels & restaurants, transport, motor vehicle manufacture, construction, chemicals and rubber manufacture.

¹³ Transport services; business and management consultancy; financial services; public services; and significant but less strong for office machinery; radio & TV equipment; food manufacture; wood and wood manufacture; basic metals and fabricated metal manufacture; construction; wholesale and retail trades; hotels & restaurants, land, water, air transport and supporting services, and motion picture and video activities.

3.2 Effects across industry and product life cycle

Some researchers consider that agglomeration externalities do not only vary across firm or industry characteristics, but also across the product or industry life-cycle. This new theory, called 'evolutionary agglomeration theory', defends that agglomeration economies generate increasing or diminishing returns depending on the stage of the product or industry life cycle¹⁴.

Potter and Watts (2011), using metaphors from biological science, evolutionary biology and biogeography, develop a theoretical model called the Agglomeration Life Cycle Model, which illustrates how the incentives to agglomerate and disperse evolve over time and how the industry life cycle changes the relationship between agglomeration economies and economic performance.

The authors divide the industry life cycle in four stages: embryonic, growth, mature and decline stage. During the embryonic stage, firms experience increasing returns from agglomeration economies and diminishing returns from dispersion economies, as in this stage firms start to agglomerate in close geographical proximity to the entrepreneurs of the start-ups within the industry benefiting from the knowledge spillovers, network connectivity and supply chain linkages. The embryonic stage evolves into the growth stage, characterised by a fast rise in the rates of firm entry, start-up, spin-off, survival and a low rate of firm exit from the industry. Since few industries are capable of sustaining growth indefinitely, the growth stage is succeeded by the mature stage, characterised by constant returns of scale, as an increasing number of firms start to experience diminishing returns from agglomeration economies, the increasing agglomeration of firms within a locality causes higher labour costs, greater land rents, congestion costs, pollution and fiercer local competition. The fourth stage of the industry life cycle, decline stage, is characterised by a period of decline of agglomeration benefits that differently affects firms in the industry; the firms that continue to depend on local firms will specialise in outdated technology, replicate established routines, and will be limited to old supply chain networks of outdated and low quality products; in contrast, the other firms, with a higher capacity to adapt, will adjust their routines (geographic relocation, industry diversification, increasing plant size, business mergers and acquisitions).

The authors test the theory empirically using plant-level data from the South Yorkshire city-region in the UK, confirming their theoretical expectations. Neffke *et al.* (2011) also find evidence of these effects in the case of Swedish industries: specialisation externalities increase with the maturity of the industry; in addition, urbanisation externalities are positive for young industries but they decline and even become negative at later stages of the industry life cycle.

It was also suggested that agglomeration externalities vary according to the stage of the industries' product lifecycle. Duranton and Puga (2001) model considers that (i) as more local firms use the same type of production process, the lower will be the cost of using it, due to specialisation economies; and (ii) urban crowding places a limit on city size. They start with the assumption that when a firm decides to produce a new product, it does not have enough knowledge on how to produce it. Firms will take more benefit to locate at this stage in more diversified cities as they will benefit from learning with local types of production processes. Three types of steady-state exist in their model: diversified cities, specialised cities, and both diversified and specialised cities. When mixed configuration exists, diversified and specialised cities, it means that each firm prefers to locate in a diversified city, while searching for its ideal process; and in the future relocate to a specialised city where all firms are using the same production process, avoiding the congestion imposed by the presence of other sectors. Duranton and Puga (2001) find evidence of these effects for the case of France, Pellenberg and Van Steen (2003) for Netherlands and Holl (2004) for Portugal.

3.3. Size and ex-ante productivity level effects

The literature shows that the agglomeration economies differ for more competitive firms. One interesting result from Combes *et al.* (2012), using French establishment level data, is that larger firms (having more workers) and those with higher productivity per se can more easily grasp the benefits of agglomeration. In

¹⁴ See Boschma and Frenken (2011) for a literature review on the empirics of evolutionary economic geography.

this sense, agglomeration will also allow an increased dilation of the distribution of firms' productivities in larger cities.

In contrast, Andersson and Lööf (2011) do not find a relation between firm size, agglomeration economies and productivity in Sweden. Mikkala (2004), assessing three manufacturing industries¹⁵ in Finland, conclude that specialisation economies are actually stronger in regions where the average size of firms is small. The authors argue that the presence of firms in the same industry helps to overcome the limitation of resources that for small firms. Capello (2002), looking at the high-tech sector in the metropolitan area of Milan, demonstrates that specialisation economies affect more heavily small firms, while urbanisation economies are more valued by large firms.

Another interesting result is from Drucker and Feser (2012), who conclude that in the U.S. a more concentrated regional industrial structure (dominated by a few large firms) limits agglomeration economies and diminishes the economic performance of firms in three manufacturing industries (rubber and plastics, metalworking machinery, and measuring and controlling devices), in particular for small firms.

3.4. Agglomeration Economies or Firm Selection?

While a vast research emphasises the role of agglomeration economies on the productive advantages of large cities, an emerging literature offers an alternative argumentation, based on firm selection meaning that the presence of more firms in larger markets makes competition more fierce, leading less productive firms to leave.

Combes *et al.* (2012) develop a framework to distinguish between agglomeration and firm selection in explaining why average productivity is higher in larger cities. Following a generalised version of the firm selection model of Melitz and Ottaviano (2008) and the agglomeration models of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002), the authors nest a model that allows parameterising the importance of agglomeration and selection. Using French establishment level data, they find that stronger selection in larger cities left-truncates the firm productivity distribution as the least productive firms exit, while stronger agglomeration right-shifts and dilates it as agglomeration effects turn firms more productive. The authors show that firm selection cannot explain spatial productivity differences and this productivity differences across urban areas in France are mostly explained by agglomeration.

4. Conclusion and way forward

Despite facing a number of methodological challenges, the existing literature establishes a positive link between agglomeration and productivity. However, these effects are not uniform across industries, firm size and product and industry life cycles.

In fact, the magnitude of agglomeration externalities differs across industries: while in some sectors the externalities are substantial, in others the effects are quite modest. Also, agglomeration economies may operate through different channels, namely within the same industry in the same region (the so-called specialisation externalities) or across industries in the same region (the diversification externalities). Evidence also indicates that there are important non-linear effects, depending, *inter alia*, on industry and product life cycles. Finally, the productivity gains from agglomeration economies also differ across firm size – however, there is no consensus on which firms, smaller or larger, benefit the most.

These results, highlighting the importance of tailor made approaches, are crucial to inform policy makers and allow for targeted and effective policy measures. The heterogeneity of results also point to the need to further develop studies at national level, before devising national policies. In addition, further cross-country studies may shed light on non-linearities and on the role of structural characteristics in mediating the results. Also, empirical applications should disentangle the types of externalities behind agglomeration economies, as they are likely to affect different industries and firms differently (e.g. larger and smaller firms).

¹⁵ Food, beverages and tobacco; wood, paper and pulp, printing and publishing; and basic metal, metal/electric products and transport equipment.

References

- Aitken, B., Hanson, G. and Harrison, A. 1997. Spillovers, foreign investment, and export behaviour. *Journal of International Economics* 43: 103-132.
- Alcácer, J. and Chung, W. 2013. Location strategies for agglomeration economies. *Strategic Management Journal* 35(12): 1749-1761.
- Andersson, M. and Lööf, H. 2011. Agglomeration and productivity: evidence from firm-level data. *Annual Regional Science* 46: 601-620.
- Baldwin, R.; Beckstead, D.; Brown, W. and Rigby, D. 2008. Agglomeration and the geography of localization economies in Canada. *Regional Studies* 42(1): 117-132.
- Baldwin, R.; Brown, W. and Rigby, D. 2010. Agglomeration economies: microdata panel estimates from Canadian manufacturing. *Journal of Regional Science* 50(5): 915-934.
- Baldwin, R. and Okubo, T. 2006. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography* 6(3): 323-346.
- Barrios, S., Görg, H. and Strobl, E. 2003. Explaining firms' export behaviour: R&D, spillovers and the destination market. *Oxford Bulletin of Economics and Statistics* 65: 475-496.
- Beaudry, C. and Schiffauerova, A. 2009. Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy* 38: 318-337.
- Bernard, A.; Jensen, B. and Schott, P. 2003. Falling trade costs, heterogeneous firms and industry dynamics. CEPR Discussion Papers, Centre for Economic Performance, LSE.
- Bernard, A. and Jensen, J. 2004. Why some firms export. *Review of Economics and Statistics* 86: 561-569.
- Boschma, R. and Frenken, K. 2005. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography* 6(3): 273-302.
- Boschma, R. and Frenken, K. 2011. The emerging empirics of evolutionary economic geography. *Journal of Economic Geography* 11: 295-307.
- Buenstorf, G. and Guenther, C. 2010. No place like home? Relocation, capabilities, and firm survival in the German machine tool industry after World War II. *Industrial and Corporate Change* 20(1): 1-28.
- Cabral, L. and Mata, J. 2003. On the evolution of the firm size distribution: facts and theory. *American Economic Review* 93(4): 1075-1090.
- Capello, R. 2002. Entrepreneurship and spatial externalities: Theory and measurement. *The Annals of Regional Science* 36: 387-402.
- Carlino, G. and Kerr, W. 2015. Agglomeration and innovation. In Duranton, Gilles, Henderson, Vernon, and Strange, Will, editors, *Handbook of Urban and Regional Economics*, vol. 5. North-Holland, Amsterdam.
- Combes, P., Gobillon, L. 2014. The empirics of agglomeration economies. In: Henderson, J.V., Duranton, G., Strange, W, editors, *Handbook of Regional and Urban Economics*, vol. 5. North Holland, Amsterdam.
- Combes, P.; Duranton, G.; Gobillon, L. and Roux, S. 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Glaeser, Edward L., editor, *Agglomeration Economies*, pages 15-65. Chicago University Press, Chicago, IL.
- Combes, P.; Duranton, G.; Gobillon, L.; Puga, D. and Roux, S. 2012. The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica* 80(6): 2543-2594.
- Davis, D. and Weinstein, D. 2008. A search for multiple equilibria in urban industrial infrastructure. *Journal of Regional Science* 48(1): 29-65.
- Di Giacinto, V.; Gomellini, M.; Micucci, G. and Pagnini, M. 2014. Mapping local productivity advantages in Italy: industrial districts, cities or both? *Journal of Economic Geography* (2): 365-394.
- Duranton, G. and Puga, D. 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91: 1454-1477.

- Duranton, G. and Puga, D. 2003. Micro-foundations of urban agglomeration economies, NBER working paper 9931.
- Feldman, M. 1999. The new economics of innovation, spillovers and agglomeration: A review of empirical studies. *Economics of Innovation and New Technology* 8: 5-25.
- Fujita, M. and Thisse, J. 2002. *Economics of agglomeration: cities, industrial location, and regional growth*. Cambridge: Cambridge University Press.
- Fujita, M. and Ogawa, H. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2): 161-196.
- Glaeser, E. (2010) Introduction – Agglomeration economies – NBER book, pp 1 – 14, University of Chicago Press.
- Graham, D. 2009. Identifying urbanisation and specialisation externalities in manufacturing and service industries. *Papers in Regional Science* 88(1): 63-84.
- Graham, D. and Kim, H. 2008. An empirical analytical framework for agglomeration economies. *Annals of Regional Science* 42: 267-289.
- Greenaway, D. and Kneller, R. 2005. Exporting, productivity and agglomeration. *European Economic Review* 52: 919-939.
- Greenstone, M.; Hornbeck, R. and Moretti, E. 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(31): 536-598.
- Head, K. and Mayer, T. 2004. The Empirics of Agglomeration and Trade, in J.V. Henderson and J.-F. Thisse, editors, *Handbook of Urban and Regional Economics*, vol. 4. New York, North Holland.
- Helpman, E.; Melitz, M. and Yeaple, S. 2004. Export versus FDI. NBER Working Paper no. 9439.
- Henderson, J. 2003. Marshall's scale economies. *Journal of Urban Economics* 53(1): 1-28.
- Holl, A. 2004. Start-ups and relocations: Manufacturing plant location in Portugal. *Papers in Regional Science* 83: 649-668.
- Jacobs, J. 1969. *The economics of cities*. Random House, New York.
- Kim, H. 1992. The translog production function and variable returns to scale. *Review of Economics and Statistics* 74: 546-552.
- Koenig, P. 2009. Agglomeration and the export decisions of French firms. *Journal of Urban Economics* 66: 186-195.
- Lall, S.; Shalizi, Z. and Deichmann, U. 2004. Agglomeration economies and productivity in Indian industry. *Journal of Development Economics* 73: 643-673.
- Lucas, R. and Rossi-Hansberg, E. 2002. On the internal structure of cities. *Econometrica* 70(4): 1445-1476.
- Marshall, A. 1890. *Principles of Economics: An Introductory Volume*. Macmillan and Co.: London, U.K.
- Martin, P.; Mayer, T. and Mayneris, F. 2011. Spatial concentration and plant-level productivity in France. *Journal of Urban Economics* 69(2): 182-195.
- Martin, P.; Ottaviano, G. 2001. Growth and agglomeration. *International Economic Review* 42(4): 947-968.
- Melitz, M. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6): 1695-1725.
- Melitz, M. and Ottaviano, G. 2008. Market size, trade and productivity. *Review of Economic Studies* 75(1): 295-316.
- Mitra, A. 1999. Agglomeration economies as manifested in technical efficiency at the firm level. *Journal of Urban Economics* 45: 490-500.
- Mitra, A. 2000. Total factor productivity growth and urbanization economies: A case of Indian cities. *Review of Urban & Regional Development Studies* 12(2): 97-108.

- Morikawa, M. 2011. Economies of density and productivity in service industries: an analysis of personal service industries based on establishment-level data. *The Review of Economics and Statistics* 93(1): 179-192.
- Mukkala, K. 2004. Agglomeration economies in the Finnish manufacturing sector. *Applied Economics* 36: 2419-2427.
- Neffke, F.; Henning, M.; Boschma, R.; Lundquist, K. and Olander, L. 2011. The dynamics of agglomeration externalities along the life-cycle of industries. *Regional Studies* 45(1): 49-65.
- Otsuka, A. and Goto, M. 2015. Agglomeration economies in Japanese industries: the Solow residual approach. *Annals of Regional Science* 54: 401-416.
- Otsuka, A.; Goto, M. and Sueyoshi, T. 2009. Industrial agglomeration effects in Japan: Productive efficiency, market access, and public fiscal transfer. *Papers in Regional Science* 89(4): 819-840.
- Ottaviano, G.; Tabuchi, T. and Thisse, J. 2002. Agglomeration and trade revisited. *International Economic Review* 43: 409-436.
- Pellenbarg, P. and Van Steen, P. 2003. Spatial perspectives on firm dynamics in the Netherlands. *Tijdschrift voor Economische en Sociale Geografie*, 94: 620–630.
- Porter, M. 1990. *The competitive advantage of nations*. Macmillan, London.
- Potter, A. and Watts, H. 2011. Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography* 11: 417-455.
- Puga, D. 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50(1): 203-219.
- Puga, D. and Venables, A. 1999. Agglomeration and economic development: import substitution vs. trade liberalisation. *The Economic Journal* 109(455): 292-311.
- Rosenthal, S. and Strange, W. 2004. Evidence on the nature and sources of agglomeration economies. In Henderson, Vernon and Thisse, Jacques-François, editors, *Handbook of Regional and Urban Economics*, volume 4, pages 2119-2171. North-Holland, Amsterdam.
- Rosenthal, S. and Strange, W. 2010. Small establishments/Big effects: Agglomeration, industrial organization and entrepreneurship. NBER book on *Agglomeration Economies*, pp. 277 – 302.
- Syverson, C. 2004. Market structure and productivity: a concrete example. *Journal of Political Economy* 112(6): 1181-1222.
- Waldfogel. 2010, Who benefits whom in the neighborhood? Demographics and retail product geography, Chapter in NBER book on *Agglomeration Economies* edited by E. Glaeser, pp. 181 – 209.