

Estimating dynamic localization economies: the inadvertent success of the specialization index and the location quotient

ANDREA FRACASSO & GIUSEPPE VITTUCCI MARZETTI

*Department of Economics and Management & School of International Studies,
University of Trento, Via Inama 5, 38122 Trento, Italy. E-mail: andrea.fracasso@unitn.it*

*Department of Sociology and Social Research, University of Milano-Bicocca,
Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy. E-mail: giuseppe.vittucci@unimib.it*

After addressing definitional issues on the concepts of concentration and specialization, the paper reviews the justifications for and the interpretation of some indicators of localization economies used in the empirical literature on agglomeration economies: specialization indexes and location quotients. A simulation exercise shows under what conditions certain specifications lead to biased estimations of dynamic localization (MAR) externalities. The results suggest that applied researchers can choose among the size of the local industry, the specialization index and the location quotient to proxy for these externalities as far as they also encompass a correct proxy for the size of the local economy.

Key words: Agglomeration economies; Localization economies; Location quotient; MAR externalities; Specialization index.

JEL classification: C21; O18; R12; R15

INTRODUCTION

Since MARSHALL (1890), economists have devoted great attention to the effects of spatial agglomeration on the levels and the growth rates of productivity, wages, output and employment. One of the oldest debate in the empirical literature on agglomeration economies is their industrial scope (ROSENTHAL and STRANGE, 2004), usually (although rather improperly) framed as an assessment of the relative importance of ‘localization economies’ (also known as ‘economies of localized industries’) vs. ‘urbanization economies’. While MARSHALL (1890) emphasized the importance of localization economies, i.e., intra-industry positive externalities associated with knowledge spillovers, input sharing and labor market pooling, JACOBS (1969) focused on the latter, stressing the importance of cross-fertilization of ideas (see, among the others, BEAUDRY and SCHIFFAUEROVA, 2009; MELO *et al.*, 2009; DE GROOT *et al.*, 2009, 2015, for recent extensive meta-analysis).

The papers that re-initiated this strand of the literature, that is GLAESER *et al.* (1992) and HENDERSON *et al.* (1995), focused in particular on dynamic agglomeration externalities, which are among the alleged determinants of persistent differentials in regional growth rates, as they refer to growth-related concepts such as knowledge, technology and innovation. Static agglomeration externalities refer instead to one-time efficiency gains produced by spatial concentration and help to account for agglomeration in a homogeneous space, as in HENDERSON (1974, 1986) and FUJITA *et al.* (1999).

Dynamic localization economies are also known as ‘MAR externalities’ after the seminal contributions of MARSHALL (1890), ARROW (1962) and ROMER (1986). In the empirical literature, a number of indicators have been used to operationalize the concepts associated with such externalities. While all the indicators reflect the two main features characterizing agglomeration economies, i.e. industry and geographical location, they differ in many respects. The indicators can be divided into two groups: size-based indexes (e.g. own-industry employment, number of industry plants, employment in related industries) and share-based indexes (e.g. share of the industry in the region, measured in terms of employment, output, or R&D expenditures; location quotient). Share-based indexes have been employed in almost half of the studies surveyed by BEAUDRY and SCHIFFAUEROVA

(2009) and the location quotient (LQ) is the most frequently used index (about 42% of the contributions covered in BEAUDRY and SCHIFFAUEROVA, 2009).

The popularity and the diffusion of the LQ are likely due to its appearance in the influential paper by GLAESER *et al.* (1992), where the LQ of industry employment is claimed to be used to capture localization economies associated with local specialization. The widespread use of share-based indexes, and of the LQ in particular, could seem surprising if one considers that ‘explicit theories of the microfoundations of agglomeration economies have nearly always been based on the idea that an increase in the *absolute scale* of activity has a positive effect [on the local economy, while they do] not make direct predictions regarding the impact of the *industry’s share of employment in a particular city or regarding the city’s share in the industry relative to other cities*’ (ROSENTHAL and STRANGE, 2004, p.2135, emphasis added). Some of these issues have been touched upon by KEMENY and STORPER (2015): while discussing the mechanisms through which industrial specialization can affect the productivity of the local industries, they tackle the theoretical underpinnings of various indexes of specialization and portray the LQ as ‘an indicator in search of a theory’ (2015, p.5).

Notwithstanding the widespread use of share-based indexes in applied works investigating agglomeration economies, much confusion remains as to the relationship between the concepts of MAR externalities, the size of the local industry, and the indexes of specialization and concentration. Even recent and valuable attempts at tackling related issues, such as the above mentioned contribution by KEMENY and STORPER (2015), fail to distinguish properly concepts and measures. By clarifying various issues regarding the use of agglomeration indexes in different empirical specifications, this work might inform the literature and encourage a more theory-consistent use of concepts and indicators, breaking what appears to be a chain of cross-references which often induces the authors to overlook important aspects.

In this note, we address the relationship between the concepts of MAR externalities, the size of the local industry, and the indexes of specialization and of concentration. The remainder of the paper is organized in three main sections. In the first section, we reveal

the confused use in the literature of the concepts of scale, concentration and specialization, as well as of the indicators adopted to capture MAR externalities in multiple regressions. We discuss the reasons why adopting indexes of specialization to capture localization externalities is theoretically incorrect, but can be justified from an empirical perspective when explanatory variables capturing the size of local economic activity are also included in the specification. In the second section, we focus our attention on the LQ, first used by GLAESER *et al.* (1992), to measure MAR externalities. We provide, respectively, theoretical and empirical reasons that may justify its adoption in multiple regressions aimed at capturing the determinants of region-industry economic activity. By means of a simulation, the last section shows that, in accordance with the theoretical analysis, certain specifications may lead to biased estimations of dynamics localization externalities and that the bias depends on region natural advantages and static agglomeration externalities driving the distribution of firms across regions.¹ A short conclusion summarizes the findings.

LOCALIZATION ECONOMIES: RELATED CONCEPTS AND MEASURES

Specialization vs concentration

The two dimensions characterizing agglomeration economies are industry and geographical location. To capture the interplay between them, scholars have typically resorted to the concepts of ‘concentration’ and ‘specialization’.

Spatial concentration relates to the location of an industry across regions (e.g. ELLISON and GLAESER, 1997; ARBIA *et al.*, 2010; MARCON and PUECH, 2010). Specialization, instead, refers to the distribution of industries in a region and it is inversely related to the degree of industrial diversification in the structure of the local economy (e.g. QUIGLEY, 1998; DURANTON and PUGA, 2000, 2001).

For instance, loosely speaking, we say that: a certain industry s is spatially concentrated if the great majority of the workers of s are located in relatively few regions; the industry s is concentrated in a certain region r if a significant fraction of the employment of s

is located in r ; a certain region r is specialized if relatively few industries account for a large fraction of the total employment in r ; the region r is specialized in industry s if a significant fraction of the total workers in r is employed in s .²

Accordingly, while the concept of concentration regards the observed tendency of an industry to concentrate in few locations within a broader area (i.e, a region within a country), the notion of specialization is associated with the industrial structure of one given area.

According to the theory, firms and workers in industries for which localization economies are large tend to concentrate as much as possible in the same area because the larger the size of the industry in the area, the larger localization externalities. This implies that the latter have a theoretical link with the concept of spatial concentration, but not with the specialization of a region. Accordingly, concentration appears as the measure that better captures the relationship between localization externalities, on the one hand, and the size of the industrial activities concentrated in a location, on the other hand.

In line with this reasoning, most researchers investigating the extent of localization externalities across industries have employed synthetic indexes of concentration (e.g., Gini, Herfindahl, Isard, Theil, Ellison-Glaeser), all of which look at the spatial distribution of a given industry (measured alternatively in terms of employment, output, or number of establishments) across a broader territory (see COMBES *et al.*, 2008, Ch.10).

The terms concentration and specialization have been used properly by a number of scholars (e.g. KRUGMAN, 1991; BRAKMAN *et al.*, 2002; COMBES *et al.*, 2008), who distinguish the economic concepts to which each of them respectively refers and adopt appropriate indicators to operationalize them. On the contrary, a number of researchers, especially in the economic geography literature, used the term specialization, instead of concentration, to refer to the relative importance of a region for a given industry, or treated the terms specialization and concentration as synonyms (e.g. BRÜLHART, 1998; COMBES, 1999; BRÜLHART and TRAEGER, 2005; DE GROOT *et al.*, 2009, 2015).³

KEMENY and STORPER (2015) calculate the degree of spatial concentration to distinguish tradable and nontradable industries and, by focusing exclusively on the former, they discuss

the implications of the size of local industry and of the industrial specialization of the local economy on the growth of sectoral wages at the local level. While the authors correctly indicate that an indicator of industrial specialization is the industry share in the local economy, they claim that the level of local employment in the industry can be used as a measure of ‘absolute’ specialization, in turn capturing the productivity-enhancing mechanisms pointed out in the New Economic Geography (NEG) literature. We shall come back on the questionable use of the terms ‘absolute’ and ‘relative’ specialization proposed by KEMENY and STORPER (2015). For now, it is important to notice that, as in most of the NEG literature, KEMENY and STORPER (2015) mix the concept of localization externalities due to the scale of the local industry with the idea that local specialization may guarantee greater industry-specific inputs.

Multivariate specifications to capture localization externalities

Many studies investigated localization externalities by estimating the impact of agglomeration on certain dimensions of the local economy. The most common approach, recently reviewed by COMBES and GOBILLON (2015), is the estimation of multiple regressions where the regressand is an economic variable that has both a location and an industry dimension. A very general specification to capture dynamic agglomeration externalities looks as follows:

$$\ln \left(\frac{y_{rst1}}{y_{rst0}} \right) = \alpha + \mathbf{x}'_{rst0} \boldsymbol{\beta} + \epsilon_{rs} \quad (1)$$

where y_{rst} is any of the variables, measured in region r and industry s at time t , used in the literature to detect the effect of agglomeration, \mathbf{x}'_{rst0} contains a number of explanatory variables, among which local and sectoral variables, and ϵ_{rs} is the region-industry pair-specific error term.

Following the seminal papers by GLAESER *et al.* (1992) and HENDERSON *et al.* (1995), in several studies y_{rst} has been the level of employment; other works have focused instead on the level of productivity (e.g. CICCONE and HALL, 1996; DEKLE, 2002; HENDERSON, 2003; CINGANO and SCHIVARDI, 2004; MARTIN *et al.*, 2011; CAINELLI *et al.*, 2015), real wages (e.g. COMBES *et al.*, 2011), births of new establishments (e.g. ROSENTHAL and

STRANGE, 2003), or rents (e.g. DEKLE and EATON, 1999) (on the direct and indirect approaches for the estimation of agglomeration economies see ROSENTHAL and STRANGE, 2004).

While synthetic indexes of concentration can be useful to assess the extent to which localization economies have influenced the localization of various industries in a broad area (e.g., a country), they cannot be included among the regressors in specifications like (1): by construction, any of these synthetic indexes ‘averages out’ the spatial dimension and therefore an index of concentration calculated over the entire national economy cannot but be equal across all locations. If inserted in an equation like (1), synthetic indexes of concentration would not allow to identify the impact of localization economies on the variable of interest. In this kind of specifications, one should rather include measures of agglomeration that vary across regions as well as industries.

From a theoretical viewpoint, as localization economies are proportional to the size of the local industry, the most appropriate solution would be to include among the regressors a measure of the size of the local industry (calculated in terms of employment or number of establishments). The corresponding specification would look like:

$$\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right) = \alpha + \lambda \ln n_{rst_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs}$$

where n_{rst_0} is the employment (or the number of plants, as in HENDERSON, 2003) in industry s and region r at time t_0 , and \mathbf{z}'_{rst_0} includes any other explanatory variable. The strength of MAR externalities is measured by the parameter λ .

As noticed by COMBES and GOBILLON (2015), the inclusion of a variable measuring the size of the local industry is in fact insufficient to distinguish localization and urbanization economies because the size of the local industry may grow together with the size of the location. An empirical specification that contains terms capturing both MAR externalities and dynamic urbanization economies, or JACOBS’ (1969) externalities, would then look like:

$$\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right) = \alpha + \lambda \ln n_{rst_0} + \theta \ln n_{rt_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (2)$$

where n_{rt_0} ($\equiv \sum_s n_{rst_0}$) is the total employment (or the total number of plants) in region r at time t_0 , and the strength of MAR externalities is again measured by the parameter λ .

An alternative specification, which can be easily derived from equation (2) by setting $\gamma = \lambda + \theta$, is one that includes the share of an industry in the local economy in place of the overall size of the local economy:

$$\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right) = \alpha + \lambda \ln \sigma_{rst_0} + \gamma \ln n_{rt_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (3)$$

where σ_{rst_0} ($\equiv n_{rst_0}/n_{rt_0}$) is the ratio of own industry employment (or establishments) to total local employment (or establishments) at time t_0 . As before, the effect of MAR externalities is measured by λ .

What is worth stressing is that the local industry share σ_{rst_0} can be used in specifications like (3) not because localization economies are theoretically linked to the local degree of industrial specialization (measured by σ_{rst_0}), but because the simultaneous inclusion of the index of specialization and of a measure of the size of the local economy allows to discriminate between localization and urbanization economies.

In their seminal work, HENDERSON *et al.* (1995) use employment data for eight manufacturing industries in 1970 and 1987 and estimate a specification where σ_{rst_0} is expected to capture MAR externalities and n_{rst_0} is the size of local activity in the sector:⁴

$$\ln n_{rst_1} = \alpha + \psi \sigma_{rst_0} + \phi \ln n_{rst_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (4)$$

Both the functional form and the interpretation of σ_{rst_0} are in fact at odds with our observations above. As to the interpretation, HENDERSON *et al.* (1995) describe σ_{rst_0} as ‘the degree of past concentration’ (1995, p.1074) and argue that ‘concentration ... may better represent the potential for MAR externalities since concentration facilitates spillover or “network” information flows among relevant firms and the development of location-specific knowledge, relative to a location with diffuse economic activity’ (1995, p.1071). Although the authors correctly mention the theoretical relationship between concentration and localization economies, they adopt an indicator that in fact captures specialization. HENDERSON *et al.* (1995) refer to CICCONE and HALL (1996) to motivate their use of

a measure of specialization to capture MAR externalities, but, in fact, CICCONE and HALL (1996) focus on local employment ‘density’, and the notion of density is in fact closer to that of concentration than to specialization: in a constant geographical area, any increase (decrease) of local industry scale makes average density increase (decrease), while no such relationship exists between density and specialization. As to the functional form, it is worth noting that the specification (4) raises issues due to the lack of a control for urbanization economies and to the inclusion of the size of local industry (n_{rst_0}) as a control.

Since n_{rst_0} appears as a control in the specifications in HENDERSON *et al.* (1995) (equation 4) and, as we shall see in the following main section, in GLAESER *et al.* (1992) (equation 6), and such inclusion is made without practically any discussion, an interesting question is why these authors used a specialization index to capture the extent of localization economies and also included a measure of local industry scale as an additional control. We believe that this choice was mainly due to the fact that the measure of local industry scale (n_{rst_0}) was encompassed in their specifications with a view to accounting for ‘mean reversion effects’ and this forced them to look for a different variable to capture MAR externalities. Indeed, given their interest in the evolution of local industry employment, these authors faced the problem of a double potential use of the local industry scale in the specification: one to capture dynamic localization economies and one to address a sort of convergence effect. Notably, as pointed out by COMBES (1999, 2000), addressing mean reversion issues by adding a measure of local industry scale into the specification did not only prevent them from using the same indicator so as to capture MAR externalities, but it also had other serious drawbacks. First, endogenous growth models that enshrine the MAR perspective are not consistent with mean reversion effects because β -convergence stems from the assumption of decreasing marginal productivity of inputs, and this is in turn inconsistent with the presence of agglomeration economies.⁵ Second, when one explicitly controls for the local industry employment by including n_{rst_0} among the regressors, as in equation (4), the only possibility for the local industry share σ_{rst_0} to increase (decrease) is that local total employment n_{rt} decreases (increases). This is equivalent to over-imposing

the restriction that MAR externalities increase (decrease) whenever local total employment decreases (increases) while local industry employment remains constant. This setting is very distant from the theoretical models GLAESER *et al.* (1992) and HENDERSON *et al.* (1995) refer to and does not seem a tenable framework.

It follows that only two alternative specifications are consistent with the theory of localization externalities, in that they focus on the impact of the size of the local economy, and permit to distinguish localization and urbanization economies:

- (a) Equation (2), that includes the size of the entire local economy (n_{rt_0}), measured in terms of employment or number of plants, and the local scale of an industry (n_{rst_0}) among the regressors;
- (b) Equation (3), that encompasses, together with the size of local economy (n_{rt_0}), the share of the industry in the local economy (σ_{rst_0}).

It is worth noting that the inclusion of region dummies in place of n_{rt_0} in the previous specifications would also lead to unbiased (although less efficient) estimations of λ . Such inclusion however would not allow to gauge the strength of urbanization economies or of any other regional factor.

To these equations, one can add another specification consistent with the theory that encompasses the LQ as a measure to capture dynamic localization economies. Given the diffusion of this index, it will be the object of a dedicated analysis in the second main section.

Absolute and relative indexes

Specialization (concentration) indexes can be either ‘absolute’ or ‘relative’, according to whether the local distribution of industries (the geographical distribution of an industry) is, respectively, benchmarked or not.

In the case of specialization, for instance, the distribution of industries in the region of interest can be compared with the average distribution of industries in the whole country or in one reference region, as in the Krugman specialization index (KRUGMAN, 1991; MIDELFART-KNARVIK *et al.*, 2000; PALAN and SCHMIEDEBERG, 2010), or even with

a purely theoretical distribution. Absolute specialization, on the contrary, is based on measures that are not assessed against a benchmark: the share of an industry in a region (σ_{rst}), for instance, is an index of absolute specialization.

Clearly, indexes of relative specialization should not be confused with share-based indexes. While it is necessary to develop measures based on shares in order to account for the industrial specialization of a region, share-based indexes of specialization can still be differentiated in ‘absolute’ and ‘relative’ according to the presence of a benchmark or the lack thereof. For instance, one can compare σ_{rst} , an index of absolute specialization for region r and sector s at time t , with the same ratio computed for another reference region ($\sigma_{r'st}$ for region r') or for different levels of geographical aggregation, e.g., σ_{st} ($\equiv n_{st}/n_t \equiv \sum_r n_{rst}/\sum_{r,s} n_{rst}$) for the entire economy. In theory, it would be possible to look at the ratio calculated for the same area in other periods of time or, alternatively, for summary statistics (mean, median, etc.) of the underlying values across all the regions. By doing so, one obtains some alternative indexes of relative specialization.

It is worth noting that the appropriateness of distinguishing size- and share-based measures has been recently pointed out also by KEMENY and STORPER (2015), who use the expression ‘relative specialization’ for what we call a (share-based) absolute specialization index (σ_{rst}) and the expression ‘absolute specialization’ for what we define as a size-based measure of the industry (n_{rst}). We do not adopt the terminology chosen by KEMENY and STORPER (2015) as we believe it is unfortunate for two reasons. First, the expression ‘absolute specialization’ is used to indicate a variable considered in levels, even though the concept of specialization cannot but be relative (as in the ratio between the size of a industry and the size of all industries in a region). Second, the term ‘relative’ is used as a synonym of ‘share-based’, while we think it can be more conveniently used to distinguish benchmarked and non-benchmarked measures of specialization, as already done in the literature (e.g. WREN and TAYLOR, 1999; COMBES *et al.*, 2008).

The location quotient (LQ) is a famous index of relative specialization (according to the definition we adopt here) and it is widely used in the literature on agglomeration economies. The LQ can be read as an index of relative specialization because it benchmarks the degree

of industrial specialization at the local level against the degree of industrial specialization of the entire national economy. Following our previous notation, the LQ can be written as:

$$LQ_{rst} = \frac{\sigma_{rst}}{\sigma_{st}} = \frac{n_{rst}/n_{st}}{n_{rt}/n_t} \quad (5)$$

where the index of specialization of r in s (σ_{rst}) is benchmarked against the same index computed at the country level (σ_{st}).

Despite its wide diffusion, the LQ remains in fact ‘an indicator in search of a theory’, to quote the effective image recently proposed by KEMENY and STORPER (2015). To address this issue and account for the correct and incorrect attempts to motivate its inclusion in empirical specifications like (1), in what follows we shall critically review the main justifications offered in the literature.

LOCATION QUOTIENT AND LOCALIZATION ECONOMIES

In their seminal article re-initiating the literature on agglomeration externalities, GLAESER *et al.* (1992) investigate employment growth rates in two-digit manufacturing industries in different cities and relate these rates to indexes capturing the structure and the size of local employment. With a view to estimating dynamic localization externalities, Glaeser and co-authors regress city-industry employment growth in the span 1956-87 on the LQ computed at the starting period t_0 , and then include, among the other explanatory variables, the initial level of city-industry employment (n_{rst_0}) to correct for ‘potential measurement problems’ (1992, p.1142),⁶ and other controls (\mathbf{z}'_{rst_0}). Accordingly, their specification reads as follows:

$$\ln \left(\frac{n_{rst_1}}{n_{rst_0}} \right) = \alpha + \beta LQ_{rst_0} + \zeta n_{rst_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (6)$$

where the regressand is the relative change in the employment level in region r and industry s from t_0 to t_1 .

GLAESER *et al.* (1992) are among the first to express the idea that the LQ of industry employment can capture the potential for MAR externalities. By referring to MARSHALL (1890), ARROW (1962) and ROMER (1986), Glaeser and co-authors maintain that MAR externalities are associated with ‘the concentration of an industry in a city’ (1992, p. 1127)

and that this can be captured by the LQ. This latter ‘measures how specialized a city is in an industry relative to what one would expect if employment in that industry was scattered randomly across the United States. The variable corrects for situations in which a city-industry is large only because the city is large’ (1992, p.1141).

Given the distinction between specialization and concentration we discussed in the previous section, this sentence seems to mix two different concepts up. First, the idea of local specialization (‘how specialized a city is in an industry’) is mentioned and, subsequently, the concept of relative concentration (‘relative to what one would expect if employment in that industry was scattered randomly across the United States’) is referred to. Notwithstanding such ambiguity, many subsequent studies (e.g. COMBES, 2000; SÜDEKUM, 2008; BALDWIN *et al.*, 2008) also use the LQ in their empirical specifications without providing better motivations.

KEMENY and STORPER (2015) have recently addressed the issue and concluded that ‘the strongest theory one can adduce in its support is the notion that there is a fixed external (national or international) demand for the output of a sector, so that if a region is specialized in a sector with external demand that increases faster than the regional demand, then the specialization will be favorable to regional growth’ (2015, p.5). This theoretical justification requires to interpret the LQ as an index of specialization; however, as clarified above, there is no theoretical relationship between the degree of industrial specialization and the emergence of localization economies.

In what follows, therefore, we shall explore some theoretical and empirical motivations that may justify the adoption of the LQ in certain specifications aimed at capturing MAR externalities.

Justifying the location quotient: a theoretical approach

If the inclusion of the LQ alone cannot be justified from a theoretical viewpoint when the LQ is interpreted as an index of industrial specialization, the question is whether there exists any theoretical motivation justifying its employment in an empirical specification

such as:

$$\ln \left(\frac{n_{rst_1}}{n_{rst_0}} \right) = \alpha + \lambda \ln LQ_{rst_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (7)$$

A first way to approach this problem is to notice that the interpretation of the LQ as an index of relative *specialization* is at variance with the fact that the LQ was originally introduced by FLORENCE (1939) as a measure of relative *concentration*. Some simple algebra on (5) allows to derive the definition of the LQ of sector s in region r proposed by FLORENCE (1939):

$$LQ_{rst} = \frac{n_{rst}/n_{st}}{n_{rt}/n_t}$$

where the index of absolute concentration of sector s in region r at time t (n_{rst}/n_{st}) is benchmarked against the same index computed aggregating all economic activities (n_{rt}/n_t). Notably, the LQ can be interpreted both as an index of relative concentration and of relative specialization because: i) the whole economy is chosen as a benchmark; ii) the benchmarking method resolves itself in a ratio of ratios. Were the underlying benchmark or the benchmarking method different, such equivalence would not hold.⁷

In the previous section, we discussed why localization economies have more to do with the concept of concentration than with that of specialization. It is for this very reason that the LQ has become so popular in the literature focusing on the spatial distribution of economic activities (see GUIMARÃES *et al.*, 2009; FIGUEIREDO *et al.*, 2009; BILLINGS and JOHNSON, 2012, for recent developments).

The idea of interpreting the LQ as a measure of relative concentration able to capture localization economies has been recently explored by GUIMARÃES *et al.* (2009). Building upon the dartboard approach put forward by ELLISON and GLAESER (1997), GUIMARÃES *et al.* (2009) show that the LQ can be derived as the maximum likelihood estimator of the strength of static localization economies and/or natural advantages at the region-industry level (the two factors are observationally equivalent in the model). The absence of localization economies and/or natural advantages implies an expected value of the LQ equal to one; a value greater than one, instead, provides evidence about the occurrence of

industrial clustering and localization economies (and/or region-industry specific natural advantages). In a nutshell, the reasoning justifying the adoption of the LQ as an indicator of localization economies is the following: when firms relocate to take advantage of static Marshallian economies, industry-region clusters emerge and the underlying phenomenon of concentration at the region-industry level is captured by a relatively large LQ.

It is not clear whether [GUIMARÃES *et al.*'s \(2009\)](#) contribution is sufficient to rescue, from a theoretical perspective, the use of LQ alone in specifications like equation (7). While the authors' considerations hold for static localization externalities, they do not necessarily extend to dynamic localization externalities, nor they necessarily hold if urbanization externalities are also at work. In the last main section of the paper, we shall illustrate through a numerical simulation what happens to the estimates of the parameters capturing the dynamic externalities when these tenets do not hold.

Justifying the location quotient: an empirical approach

As argued before, several studies adopted the LQ as an indicator of localization economies by means of vague references to [GLAESER *et al.* \(1992\)](#). In practice, several authors have mechanically used the (log of) LQ in place of n_{rst_0} or σ_{rst_0} to estimate localization economies in specifications like equations (2) and (3). In such cases, the estimated function has taken the following form:

$$\ln \left(\frac{n_{rst_1}}{n_{rst_0}} \right) = \alpha + \lambda \ln LQ_{rst_0} + \theta \ln n_{rt_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs} \quad (8)$$

In fact, that the LQ can be used in place of either n_{rst_0} or σ_{rst_0} in multiple regressions depends on whether it allows to distinguish localization and urbanization economies. As previously explained, to identify and distinguish the two channels through which agglomeration externalities impact on the local economy, one has to include in the specification a measure of the size of the entire local economy (n_{rt_0}) to capture dynamic urbanization externalities. It is the very same inclusion of a term capturing the total size of the local economy that justifies, from an econometric perspective, the substitution of a measure of the local size of the industry (such as n_{rst_0}) with absolute and relative indexes of specialization (e.g., σ_{rst_0} or LQ_{rst_0}).⁸

It follows that the LQ can be included in a specification like (8) not because it is, from a theoretical perspective, a better proxy of localization externalities than the level of economic activity. In fact, specialization indexes have no direct relationship with localization economies as there is no monotonic relation between local own industry scale and specialization. Actually, one can think of many situations in which local own industry scale and specialization might move in opposite directions. For instance, when total local employment increases more than local own industry employment in a region via a relocation of workers from outside the region (in a way that keeps constant the ratio of industry employment to total employment at the country level), the local own industry size raises and the local own industry share falls. This is not just a theoretical possibility: as shown by KEMENY and STORPER (2015), the rankings of US metropolitan areas based on sizes and shares are very different.

Rather, the use of the LQ is justified by the fact that the simultaneous inclusion of the LQ and of a term capturing the total size of the local economy (n_{rt_0}) allows to distinguish the impact of both localization and urbanization economies.

In the next section, we will elaborate a simulation exercise to show that our considerations about the conditions justifying the empirical use of the LQ are correct and practically relevant.

EMPIRICAL RESULTS AND RELEVANCE

In order to show that the use of the specialization index and the LQ in the empirical estimation of dynamic localization externalities is correct (and almost equivalent to the specification in levels) when a term capturing the size of the local economy is also included in the functional form, we shall follow a Monte Carlo approach to investigate under what conditions the estimated parameters capturing the dynamic agglomeration externalities can be biased.⁹

We analyze various scenarios and the procedure characterizing each run in the simulation in each scenario consists of three steps. In the first step, we generate an allocation of establishments across regions and industries allowing establishments to relocate into regions

on the basis of a probability that depends on static agglomeration externalities and/or natural advantages, along the lines of the stochastic localization models put forward by ELLISON and GLAESER (1997), GUIMARÃES *et al.* (2009) and BOTTAZZI *et al.* (2007). In the second step, given the allocation of establishments resulting at the end of the first step, we generate a number of values of the dependent variable, $\ln\left(\frac{y_{rst1}}{y_{rst0}}\right)$, for each and every region-industry pair on the basis of a theory-consistent random model of dynamic agglomeration externalities. The last step consists in the estimation of the parameters of seven alternative functional forms (corresponding to the specifications discussed in the previous sections) on the data generated at the end of the second step and in the calculation of the means of the estimated parameters. We shall describe in more detail these three steps in what follows. This procedure allows to study the presence of an estimation bias (or lack thereof) with alternative specifications and different underlying hypotheses regarding the forces driving the spatial distribution of firms.

To anticipate the main results, we will show that not all the alternative specifications lead to unbiased estimations in all scenarios. Interestingly, the presence and the size of the bias depend both on the nature of the natural advantages and the static agglomeration externalities driving the (re)location of establishments across regions and on the specification adopted to capture the dynamic externalities. These results confirm the intuitions discussed in the previous sections. On this basis we shall provide some operational recommendations.

Natural advantages, static agglomeration externalities and spatial distribution of firms

The first step of each replication is aimed at generating micro-founded data on n_{rs} with 10 industries and 50 regions. To generate such data, we assume that the allocation of establishments across regions is driven by static agglomeration externalities and/or region natural advantages in a stochastic localization model. In each simulation run, we start by sampling from a Poisson distribution whereby the expected number of establishments in each industry-region pair is 20, so that the expected number of establishments in each sample is 10,000. Then, we make five millions iterations. In each iteration, one

establishment is selected uniformly at random to revise its location choice across regions on the basis of static (localization and/or urbanization) externalities and/or regions-specific natural advantages. In particular, we assume that the probability that the selected establishment in industry s chooses to (re)locate in region r is:

$$p_{rs} \propto \beta_{0r} + \beta_l n_{rs} + \beta_u \sum_{k \neq s} n_{rk} \quad (9)$$

where β_{0r} is a (possibly region-specific) strictly positive parameter measuring the region attractiveness (i.e., natural advantage), and β_l and β_u are non negative parameters measuring, respectively, the strength of static localization and urbanization economies.¹⁰

This stochastic localization model, similar in spirit to BOTTAZZI *et al.* (2007) (see also BOTTAZZI and SECCHI, 2007; BOTTAZZI and GRAGNOLATI, 2015; BOTTAZZI *et al.*, 2015), gives raise to a finite Markov Chain (MC). Since in each period every establishment in every industry and region has a positive probability to revise its locational choice and every region has a strictly positive probability of being chosen as new location ($p_{rs} > 0 \forall r, s$), if the state space is defined so to include all and only the possible allocations of establishments that do not alter the sectoral distribution (the number of establishments in each industry is given and constant), the finite MC is irreducible (for all states communicate with each other) and aperiodic (for the diagonal elements of the transition matrix are not zeros), it has therefore a unique stationary and limiting distribution, and the MC is ergodic (KARLIN and TAYLOR, 1975).

Since we let the process run and take a realization of the Markov process after many iterations, we actually sample from the stationary distribution.¹¹ It follows that the final allocation of establishments across regions does not depend on the initial regional allocation in the simulation. On the contrary, the model does not change the initial allocation of establishments across industries.¹²

We map different scenarios where natural advantages, static localization economies and static urbanization economies are more or less important in driving firms' (re)location choices by modifying the values of β_{0r} (setting them either equal or diversified across regions) and of the coefficients β_l and β_u . The nature of the forces driving firms' allocation

across regions is an important aspect to investigate because the theoretical justification of using the LQ alone in equation (7) is based on two assumptions: i) strong positive static localization externalities; ii) negligible urbanization externalities.

We discuss eight scenarios. In the first four scenarios (I-IV), we assume that no region enjoys a natural advantage and thus set $\beta_{01} = \beta_{02} = \dots = \beta_{0R} = 1$ in equation (9). In the other four scenarios (V-VIII), we relax the assumption of ‘space homogeneity’ and sample the β_{0r} ’s from a uniform distribution: $\beta_{0r} \sim \mathcal{U}(0, 2)$.

Regardless of whether we allow for the presence of natural advantages, we analyze four alternative combinations of static externalities: i) no static externalities, $\beta_l = \beta_u = 0$ (scenarios I and V); ii) only static localization economies, $\beta_l = 1$ and $\beta_u = 0$ (II, VI); iii) only static urbanization economies, $\beta_l = 0$ and $\beta_u = 1$ (III, VII); iv) both static localization and urbanization economies, $\beta_l = \beta_u = 1$ (IV, VIII).

Dynamic agglomeration externalities

In the second step of each simulation run, we calculate the variables of interest ($\ln n_{rst_0}$, $\ln n_{rt_0}$, $\ln \sigma_{rst_0}$, $\ln LQ_{rst_0}$) by using the allocation of plants across industries and regions obtained at the end of the first step (n_{rst_0} in 10 industries and 50 regions, for a total of 500 observations).

We then generate the dependent variable, $\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right)$, according to a specification of dynamic agglomeration externalities that is closely related to the theory focusing on the size of economic activity, both at the industry-region level (localization) and at the regional level (urbanization):

$$\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right) = \alpha + \lambda \ln n_{rst_0} + \theta \ln n_{rt_0} + \epsilon_{rs} \quad (10)$$

where ϵ_{rs} is sampled from a normal distribution with zero expected value and standard deviation equal to 0.005, and the coefficients are set equal to $\alpha = 0.02$, $\lambda = 0.03$ and $\theta = 0.06$.¹³

Estimated parameters and biases

In the last step of each simulation run, we run OLS regressions on the sample of 500 observations using seven functional forms found in the literature. More specifically, we estimate the following specifications:

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln n_{rst0} + \hat{\theta} \ln n_{rt0} \quad (\text{a})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln \sigma_{rst0} \quad (\text{b})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln \sigma_{rst0} + \hat{\gamma} \ln n_{rt0} \quad (\text{c})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln \sigma_{rst0} + \hat{\phi} \ln n_{rst0} \quad (\text{d})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln LQ_{rst0} \quad (\text{e})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln LQ_{rst0} + \hat{\gamma} \ln n_{rt0} \quad (\text{f})$$

$$\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right) = \hat{\alpha} + \hat{\lambda} \ln LQ_{rst0} + \hat{\phi} \ln n_{rst0} \quad (\text{g})$$

where $\hat{\alpha}$, $\hat{\lambda}$, $\hat{\theta}$, $\hat{\gamma}$ and $\hat{\phi}$ are the estimated coefficients and $\ln \left(\widehat{\frac{y_{rst1}}{y_{rst0}}} \right)$ is the predicted value.

Equation (a) matches the functional form of the data generating process. Equations (b), (c), and (d) encompass the index of specialization, respectively: alone, with the control for the size of the local economy n_{rt0} (as in equation (3) and suggested by COMBES and GOBILLON, 2015), and with the bad control for the size of the local industry n_{rst0} (as in equation (4) and in HENDERSON *et al.*, 1995). Equations (e), (f), and (g) encompass the LQ, respectively: alone, as in equation (7); with the control for the size of the local economy n_{rt0} , as in equation (8); and with the bad control for the size of the local industry n_{rst0} , as in GLAESER *et al.* (1992). We recall that the parameter γ should be equal to the sum of the parameters λ and θ in equation (10).

In each of the seven specifications for each of the eight scenarios, we made 50 simulation runs and obtain 50 different estimations of each parameter. Their sample means are computed to obtain unbiased and consistent estimators of the (finite-sample) bias of

the OLS estimators of λ (dynamic localization economies) and θ (dynamic urbanization economies): for each coefficient, the difference between the mean of the estimates and the real value of the coefficient.

Results

Table 1 reports the means (and the sample standard deviations) of the OLS estimates for the seven specifications in the eight scenarios. These latter differ in the combinations deriving from the presence/absence of natural advantages, static localization economies and static urbanization economies, that is the factors affecting establishments' distribution across regions.

In all cases, when the specification (a) is estimated, the means of the OLS estimates $\hat{\lambda}$ and $\hat{\theta}$ are equal to the values adopted to generate the simulation. No bias is present when the specification matches the data generating process, as expected.

The adoption of the specialization index alone in equation (b) does not lead to a bias in the estimated parameter λ only when firms' spatial distribution is not affected by either static urbanization economies or regions' natural advantages (I, II). When either of these forces is present, the failure to control for the overall size of the local activity produces an extremely large bias (with estimated coefficients twice as large as those used to generate the data). In fact, the presence of static localization economies may be sufficiently strong to prevent natural advantages from biasing the estimates of λ in specifications where a variable for dynamic urbanization economies is not included (VI). On the contrary, when static urbanization economies are present and sufficiently strong, the estimator is biased even when static localization economies are also at work (VIII).

The implicit empirical relationship between natural advantages, static and dynamic urbanization externalities, and its impact on the estimated coefficients of the dynamic externalities are far from obvious and have not been highlighted in the literature before. The simulation shows that the source of the upward bias of the estimator of localization economies in specification (b) is amenable to the fact that natural advantages and static urbanization economies drive firms' spatial distribution so as to create a positive correlation

Table 1. Means of OLS estimates

Equation	Parameter	No natural advantage: $\beta_{01} = \beta_{02} = \dots = \beta_{0R} = 1$				Natural advantages: $\beta_{0r} \sim \mathcal{U}(0, 2)$			
		(I) $\beta_l = \beta_u = 0$	(II) $\beta_l = 1, \beta_u = 0$	(III) $\beta_l = 0, \beta_u = 1$	(IV) $\beta_l = \beta_u = 1$	(V) $\beta_l = \beta_u = 0$	(VI) $\beta_l = 1, \beta_u = 0$	(VII) $\beta_l = 0, \beta_u = 1$	(VIII) $\beta_l = \beta_u = 1$
(a)	$\hat{\lambda}$.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)	.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)
	$\hat{\theta}$.060 (.003)	.060 (.001)	.060 (.001)	.060 (.001)	.060 (.001)	.060 (.000)	.060 (.001)	.060 (.001)
	$\hat{\lambda}$.030 (.001)	.030 (.001)	.061 (.0045)	.060 (.004)	.055 (.003)	.032 (.002)	.056 (.006)	.053 (.006)
(c)	$\hat{\lambda}$.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)	.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)
	$\hat{\gamma}$.090 (.003)	.090 (.001)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)
	$\hat{\lambda}$	-.060 (.003)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.000)	-.060 (.001)	-.060 (.001)
(d)	$\hat{\phi}$.090 (.003)	.090 (.001)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)
	$\hat{\lambda}$.030 (.001)	.030 (.001)	.061 (.0045)	.060 (.004)	.0555 (.003)	.032 (.002)	.056 (.006)	.053 (.006)
	$\hat{\lambda}$.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)	.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)
(e)	$\hat{\lambda}$.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)	.030 (.001)	.030 (.000)	.030 (.001)	.030 (.001)
	$\hat{\gamma}$.090 (.003)	.090 (.001)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)
	$\hat{\lambda}$	-.050 (.005)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.000)	-.060 (.001)	-.060 (.001)
(g)	$\hat{\phi}$.080 (.005)	.090 (.001)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)
	$\hat{\lambda}$	-.050 (.005)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.001)	-.060 (.000)	-.060 (.001)	-.060 (.001)
	$\hat{\phi}$.080 (.005)	.090 (.001)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)	.090 (.000)

Note: $\gamma = .03, \theta = .06, \gamma = \lambda + \theta = .09$. Observations: 50. Sample standard deviations in brackets.

between $\ln \sigma_{rs}$ and $\ln n_r$. When there are dynamic urbanization economies ($\theta > 0$), this necessarily upwardly biases the estimator of λ .¹⁴

The use of the specialization index to proxy for dynamic localization economies turns out to be appropriate again when the specification includes a control for dynamic urbanization economies (equation c).¹⁵ In all the eight scenarios, specification (c) produces no bias in the estimator of λ : $\bar{\hat{\lambda}} = 0.03$. Notably, as γ is equal to $\lambda + \theta$ and the mean of $\hat{\gamma}$ in specification (c) is 0.09, $\hat{\gamma} - \hat{\lambda}$ is an unbiased estimator of the strength of dynamic urbanization economies.

As anticipated, the specification chosen by HENDERSON *et al.* (1995), namely (d), leads to a biased estimator of λ for the inclusion of the bad control $\ln n_{rs}$: in all scenarios, for specification (d), $\bar{\hat{\lambda}} = -0.06$.

The same conclusions across the eight scenarios can be derived for the specifications where the LQ is used to proxy for the dynamic localization economies. Introducing the LQ alone (equation e) leads to an unbiased estimator of λ only when the firms' distribution does not depend either on natural advantages or static urbanization economies (I, II). With no static urbanization economies ($\beta_u = 0$), the presence of regions' natural advantages also leads to a bias – $\bar{\hat{\lambda}} = 0.0555$ in scenario IV and specification (e) –, unless static localization externalities are strong enough to overcome the natural advantages among the forces driving firms' location (VI). On the contrary, when one controls for the size of the local economy ($\ln n_r$) as in specification (f), no bias arises in any scenario. Finally, in specification (g) with the bad control for $\ln n_{rs}$, the estimator of λ is always seriously biased.¹⁶

Operational recommendations

Our results suggest that, to the extent that dynamics externalities depend on the size of the economic activities as the theory maintains, the applied researcher can choose freely among the size of the local industry, the index of specialization and the LQ to proxy for dynamic localization economies as far as she also encompasses in the specification a correct proxy for the size of the local economy. Failure to do so may lead to biased estimations,

especially when natural advantages and static urbanization externalities impact on firms' spatial distribution. As the bias grows with the size of θ and with the covariance between $\ln n_{rs}$ and $\ln n_r$, a negligible correlation in the sample between the sizes of local industries and local total economic activity may justify the use of the specialization index or the LQ alone in the specification.

The attempt to control for dynamic urbanization externalities with an incorrect proxy, such as $\ln n_{rs}$ instead of $\ln n_r$, is conducive to biased estimations as well, independently from the process driving the (re)location of firms.

CONCLUSIONS

In this paper we focused on the indicators and on the empirical specifications used in the literature to estimate dynamic localization economies. We clarified the notions of scale, concentration and specialization, and analyzed what indicators can be correctly used to capture dynamic externalities in multiple regressions. We provided some empirical and theoretical motivations justifying the adoption of the location quotient (LQ) to identify MAR externalities, thereby revealing common misuses and misinterpretations of such indicator. Finally, by means of a numerical simulation, we showed that the adoption of certain specifications may lead to biased estimators of the strength of dynamic localization externalities and that the bias depends on the nature of the natural advantages and the static agglomeration externalities driving the distribution of firms across regions.

Besides providing more solid arguments for the adoption of the LQ in multiple regressions, our analysis more generally informs the literature on the risks of using inappropriately and of interpreting incorrectly certain share-based localization indexes. Although the majority of the scholars typically adopted only few indicators and functional forms to study the impact of dynamic agglomeration externalities on local economic activity, their interpretation has often been imprecise. Circular cross-references and vague citations of the seminal papers in the literature have so far prevented a throughout analysis of the appropriateness of indicators and functional forms receiving widespread support. In this work, we attempted to shed light on the possible inconsistencies between the features

of the agglomeration indexes, the theoretical concepts that these measures are meant to operationalize, their names and definitions, and their interpretation.

Analyzing each and every contribution produced in this rich field of the literature and assessing whether it made a proper use of agglomeration indexes falls beyond the scope of the paper. For instance, one could revisit the recent meta-analyses proposed by BEAUDRY and SCHIFFAUEROVA (2009); MELO *et al.* (2009); DE GROOT *et al.* (2009, 2015) and verify whether a potential source of variation in the empirical findings in the literature is due to either an imprecise inclusion of some indexes in certain estimated functional forms, or an incorrect interpretation of the individual coefficients, or a too loose relationship of the adopted indexes with the underlying theory. This represents a venue for further research.

NOTES

1. We would like to thank an anonymous reviewer and the associate editor for the suggestion.
2. In fact, both concentration and specialization can be calculated with reference to different aspects of production (e.g. output, employment, establishments).
3. In fact, as we shall discuss in the following sections, this confusion is already present in the seminal papers by GLAESER *et al.* (1992) and HENDERSON *et al.* (1995), that contain some incorrect references to specialization and concentration. Note in passing that, although COMBES *et al.* (2008) properly distinguish the two concepts, they maintain that sorting the regions by their shares in total employment of a certain industry is sorting them by ‘their degree of specialization’ (2008, p.260).
4. Considering that the outcome variable (y_{rst}) in HENDERSON *et al.* (1995) is the local industry employment (n_{rst}), equation (4) can be re-expressed as:

$$\ln \left(\frac{y_{rst_1}}{y_{rst_0}} \right) = \alpha + \psi \sigma_{rst_0} + (\phi - 1) \ln n_{rst_0} + \mathbf{z}'_{rst_0} \boldsymbol{\delta} + \epsilon_{rs}$$

where the parameter ψ is expected to capture MAR externalities. Apart from the fact that the specialization index does not enter in log, this equation differs from equation (3)

because: i) there is a control for the local industry employment ($\ln n_{rst_0}$); ii) there is no control for the total local employment ($\ln n_{rt_0}$).

5. This is incidentally noted also by GLAESER *et al.* (1992, p.1142): ‘The MAR view is somewhat incompatible with the presence of real (as opposed to measurement-induced) mean reversion.’ Nonetheless, when local industry employment is concerned, a different kind of mean reversion due to the presence of congestion effects may be at work. And when mean reversion is present, as explicitly acknowledged by HENDERSON (2003, p.4), ‘it is difficult to disentangle dynamic externalities from mean reversion processes – both typically involve the same quantity, measures of past own industry employment’. In fact, employment growth is a very indirect measure of agglomeration economies: as in theory these latter directly affect only productivity, a more direct approach would require to focus on the effects of agglomeration on productivity. Notably, employment growth is positively associated with agglomeration economies only when several additional assumptions hold. First, as argued by COMBES *et al.* (2004), a positive relationship requires an elastic demand for the product and an imperfectly inelastic supply of local labor. Second, one should be able to exclude other sources of local advantages, such as fixed local factors (e.g., land), which affect the marginal productivity of labor and employment responses to changes in productivity. Moreover, one should exclude that labor supply depends on local conditions correlated with the level and structure of agglomeration (see CINGANO and SCHIVARDI, 2004).
6. In fact, as noted by COMBES (1999, 2000), the inclusion of this term does not solve the endogeneity problem resulting from the measurement errors and, as previously discussed, resolves itself in a serious ‘bad control problem’.
7. Given the ambivalent function of the LQ as a measure of specialization and concentration, it could be argued that we somehow unfairly assumed that GLAESER *et al.* (1992) treated the LQ as a specialization index while what they had in mind was a concentration index. In fact, our interpretation follows what done in most subsequent works and the common interpretation of GLAESER *et al.*’s (1992) use of LQ as an index of specialization. In their meta-analysis of the literature, BEAUDRY and SCHIFFAUEROVA (2009) maintain that

‘GLAESER *et al.* (1992) first expressed the idea that the degree of specialization may better represent the potential for Marshall externalities than current size of an industry’ and that ‘following GLAESER *et al.* (1992) the majority (of the studies reviewed by DE GROOT *et al.*, 2009) proxied for specialization using location quotients, and secondarily other forms of relative specialization.’

8. It could be argued that even the inclusion of n_{rt_0} in the specification does not allow to account for the effects of changes of n_{st_0}/n_{t_0} on LQ_{rst_0} . Accordingly, the estimated coefficient of LQ_{rst_0} reflects the variation in both the scale of local industry and in n_{st_0} . While theoretically correct, this does not raise any issue in practice as we shall show in the following section.
9. The simulations were done using R. Code available upon request.
10. This stochastic choice model can be derived assuming an unobservable term which enters the profit function in the form of a stochastic component à la MCFADDEN (1974), or assuming heterogeneous preferences (e.g. ANDERSON *et al.*, 1989; JÄIBI and TEN RAA, 1998). See BOTTAZZI and SECCHI (2007) for a discussion within the context of localization models.
11. The rate of convergence to the invariant distribution in the Ehrenfest-Brillouin model, framing the model in BOTTAZZI *et al.* (2007) and BOTTAZZI and GRAGNOLATI (2015), depends on the number of establishments (n) and the total sum of weights of the different regions (A). With one establishment moving in each period, the rate of convergence is equal to: $A/(n(A + n - 1))$ (see GARIBALDI and SCALAS, 2010, Ch.7).
12. The number of plants in each industry, being the sum of 50 independent Poisson random variables with expected value 20, is Poisson distributed with expected value 1,000.
13. In the simulation, we use the number of plants in the various regions and industries to measure the size of the economic activity, instead of the number of employees. Under the assumption that all plants have an identical size, the two exercises are clearly equivalent.

14. The bias of the OLS estimator of λ when we run the simple regression of $\ln\left(\frac{y_{rst1}}{y_{rst0}}\right)$ on σ_{rs} excluding n_r is given by (WOOLDRIDGE, 2002, pp.61-63):

$$\text{Bias}(\hat{\lambda}) = \theta \frac{\text{Cov}(\ln \sigma_{rs}, \ln n_r)}{\text{Var}(\ln \sigma_{rs})}$$

where:

$$\begin{aligned} \frac{\text{Cov}(\ln \sigma_{rs}, \ln n_r)}{\text{Var}(\ln \sigma_{rs})} &= \frac{\text{Cov}(\ln n_{rs} - \ln n_r, \ln n_r)}{\text{Var}(\ln n_{rs} - \ln n_r)} = \\ &= \frac{\text{Cov}(\ln n_{rs}, \ln n_r) - \text{Var}(\ln n_r)}{\text{Var}(\ln n_{rs}) + \text{Var}(\ln n_r) - 2 \text{Cov}(\ln n_{rs}, \ln n_r)} \end{aligned}$$

This ratio increases with the covariance between $\ln n_{rs}$ and $\ln n_r$.

Notably, when the sample includes all the regions and industries in a certain country, the sample covariance between σ_{rs} and n_r is *always* zero. Indeed, by denoting with R the number of regions, with S the number of industries, and with $s_{\sigma n_r}$ the sample covariance between σ_{rs} and n_r , the following equality holds:

$$\begin{aligned} (RS - 1) s_{\sigma n_r} &= \sum_{r,s} (\sigma_{rs} - \bar{\sigma})(n_r - \bar{n}_r) = \sum_{r,s} \sigma_{rs} n_r - RS \bar{\sigma} \bar{n}_r \\ &= \sum_{r,s} \frac{n_{rs}}{n_r} n_r - RS \left(\frac{\sum_{r,s} \frac{n_{rs}}{n_r}}{RS} \frac{S \sum_r n_r}{RS} \right) = \sum_{r,s} n_{rs} - RS \left(\frac{1}{S} \frac{n}{R} \right) = n - n = 0 \end{aligned}$$

where \bar{x} is the sample mean of x . However, this equality does not hold anymore when either σ_{rs} or n_r is in log.

15. Alternatively, one can add region dummies to the specification (b). The inclusion of regional fixed effects leads to an unbiased (although less efficient) estimator of λ , but prevents the estimation of θ , the strength of urbanization economies.
16. It is worth noting that the estimates delivered by the specifications with the LQ – (e) and (f) – are practically equal to the specifications with the specialization index – (b) and (c), notwithstanding the LQ can change also for changes in n_s . This is because: i) although in each simulation run the number of establishments in each industry changes, it is constant in expected value across industries; ii) the covariance between $\ln LQ_{rs}$ and $\ln n_r$ is roughly

equal to the covariance between $\ln \sigma_{rs}$ and $\ln n_r$:

$$\begin{aligned}\text{Cov}(\ln LQ_{rs}, \ln n_r) &= \text{Cov}(\ln \sigma_{rs} - \ln \sigma_s, \ln n_r) = \text{Cov}(\ln \sigma_{rs}, \ln n_r) - \text{Cov}(\ln \sigma_s, \ln n_r) \\ &= \text{Cov}(\ln \sigma_{rs}, \ln n_r) - \text{Cov}(\ln n_s, \ln n_r) \approx \text{Cov}(\ln \sigma_{rs}, \ln n_r).\end{aligned}$$

REFERENCES

- ANDERSON S., DE PALMA A. and THISSE J.-F. (1989) Demand for differentiated products, discrete choice models, and the characteristics approach, *Review of Economic Studies* **56**(1), 21–35.
- ARBIA G., ESPA G., GIULIANI D. and MAZZITELLI A. (2010) Detecting the existence of space-time clustering of firms, *Regional Science and Urban Economics* **40**(5), 311–323.
- ARROW K. J. (1962) Economic welfare and the allocation of resources for invention, in NELSON R. (Ed.) *The Rate and Direction of Inventive Activity*, pp. 609–626, Princeton University Press, Princeton, NJ.
- BALDWIN J. R., BECKSTEAD D., MARK BROWN W. and RIGBY D. L. (2008) Agglomeration and the geography of localization economies in Canada, *Regional Studies* **42**(1), 117–132.
- BEAUDRY C. and SCHIFFAUEROVA A. (2009) Who’s right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy* **38**(2), 318–337.
- BILLINGS S. B. and JOHNSON E. B. (2012) The location quotient as an estimator of industrial concentration, *Regional Science and Urban Economics* **42**(4), 642–647.
- BOTTAZZI G., DOSI G., FAGIOLO G. and SECCHI A. (2007) Modeling industrial evolution in geographical space, *Journal of Economic Geography* **7**(5), 651–672.
- BOTTAZZI G. and GRAGNOLATI U. (2015) Cities and clusters: Economy-wide and sector-specific effects in corporate location, *Regional Studies* **49**(1), 113–129.
- BOTTAZZI G., GRAGNOLATI U. and VANNI F. (2015) Non-linear externalities in firm localization, LEM Working Paper 28, Scuola Superiore Sant’Anna.
- BOTTAZZI G. and SECCHI A. (2007) Repeated choices under dynamic externalities, LEM Working Paper 08, Scuola Superiore Sant’Anna.

- BRAKMAN S., GARRETSEN H. and VAN MARREWIKJ C. (2002) *An Introduction to Geographical Economics. Trade, Location and Growth*, Cambridge University Press, Cambridge.
- BRÜLHART M. (1998) Economic geography, industry location and trade: the evidence, *The World Economy* **21**(6), 775–801.
- BRÜLHART M. and TRAEGER R. (2005) An account of geographic concentration patterns in Europe, *Regional Science and Urban Economics* **35**(6), 597–624.
- CAINELLI G., FRACASSO A. and VITTUCCI MARZETTI G. (2015) Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach, *Papers in Regional Science* **94**(S1), S39–S67.
- CICCONE A. and HALL R. E. (1996) Productivity and the density of economic activity, *The American Economic Review* **86**(1), 54–70.
- CINGANO F. and SCHIVARDI F. (2004) Identifying the sources of local productivity growth, *Journal of the European Economic Association* **2**(4), 720–742.
- COMBES P. P. (1999) Marshall-Arrow-Romer externalities and city growth: A methodological note, Working paper 99–06, CERAS.
- COMBES P. P. (2000) Economic structure and local growth: France, 1984-1993, *Journal of Urban Economics* **47**(3), 329–355.
- COMBES P. P., DURANTON G. and GOBILLON L. (2011) The identification of agglomeration economies, *Journal of Economic Geography* **11**(2), 253–266.
- COMBES P. P. and GOBILLON L. (2015) The empirics of agglomeration economies, in DURANTON G., HENDERSON J. V. and STRANGE W. (Eds.) *Handbook of Urban and Regional Economics*, vol. 5, Elsevier, Amsterdam.
- COMBES P. P., MAGNAC T. and ROBIN J.-M. (2004) The dynamics of local employment in France, *Journal of Urban Economics* **56**(2), 217–243.
- COMBES P. P., MAYER T. and THISSE J.-F. (2008) *Economic geography: The integration of regions and nations*, Princeton University Press, Princeton, NJ.
- DE GROOT H. L., POOT J. and SMIT M. J. (2009) Agglomeration externalities, innovation and regional growth: Theoretical perspectives and meta-analysis, in CAPELLO

- R. and NIJKAMP P. (Eds.) *Handbook of Regional Growth and Development Theories*, chap. 14, pp. 256–281, Edward Elgar, Cheltenham.
- DE GROOT H. L., POOT J. and SMIT M. J. (2015) Which agglomeration externalities matter most and why?, *Journal of Economic Surveys* **30**(4), 756–782.
- DEKLE R. (2002) Industrial concentration and regional growth: evidence from the prefectures, *Review of Economics and Statistics* **84**(2), 310–315.
- DEKLE R. and EATON J. (1999) Agglomeration and land rents: evidence from the prefectures, *Journal of Urban Economics* **46**(2), 200–214.
- DURANTON G. and PUGA D. (2000) Diversity and specialisation in cities: Why, where and when does it matter?, *Urban Studies* **37**(3), 533–555.
- DURANTON G. and PUGA D. (2001) Nursery cities: Urban diversity, process innovation, and the life cycle of products, *American Economic Review* **91**(5), 1454–1477.
- ELLISON G. and GLAESER E. (1997) Geographic concentration in US manufacturing industries: A dartboard approach, *Journal of Political Economy* **105**(5), 889–927.
- FIGUEIREDO O., GUIMARÃES P. and WOODWARD D. (2009) Localization economies and establishment size: was Marshall right after all?, *Journal of Economic Geography* **9**(6), 853–868.
- FLORENCE P. S. (1939) *Report of the location of industry*, Political and Economic Planning, London.
- FUJITA M., KRUGMAN P. and VENABLES A. (1999) *The Spatial Economy. Cities, Regions and International Trade*, The MIT Press, Cambridge.
- GARIBALDI U. and SCALAS E. (2010) *Finitary probabilistic methods in econophysics*, Cambridge University Press, Cambridge.
- GLAESER E. L., KALLAL H. D., SCHEINKMAN J. A. and SHLEIFER A. (1992) Growth in cities, *Journal of Political Economy* **100**(6), 1126–1152.
- GUIMARÃES P., FIGUEIREDO O. and WOODWARD D. (2009) Dartboard tests for the location quotient, *Regional Science and Urban Economics* **39**(3), 360–364.
- HENDERSON J. V. (1974) The sizes and types of cities, *American Economic Review* **64**, 640–656.

- HENDERSON J. V. (1986) Efficiency of resource usage and city size, *Journal of Urban Economics* **19**(1), 47–70.
- HENDERSON J. V. (2003) Marshall’s scale economies, *Journal of Urban Economics* **53**(1), 1–28.
- HENDERSON J. V., KUNCORO A. and TURNER M. (1995) Industrial development in cities, *Journal of Political Economy* **103**(5), 1067–1090.
- JACOBS J. (1969) *The Economy of Cities*, Vintage, New York, NY.
- JÄIBI M. and TEN RAA T. (1998) An asymptotic foundation for logit models, *Regional Science and Urban Economics* **28**, 75–90.
- KARLIN S. and TAYLOR H. (1975) *A first course in stochastic processes*, Academic Press, New York.
- KEMENY T. and STORPER M. (2015) Is specialization good for regional economic development?, *Regional Studies* **49**(6), 1003–1018.
- KRUGMAN P. (1991) *Geography and Trade*, MIT Press, Cambridge, Mass.
- MARCON E. and PUECH F. (2010) Measures of the geographic concentration of industries: improving distance-based methods, *Journal of Economic Geography* **10**(5), 745–762.
- MARSHALL A. (1890) *Principles of Economics*, Macmillan, London.
- MARTIN P., MAYER T. and MAYNERIS F. (2011) Spatial concentration and plant-level productivity in France, *Journal of Urban Economics* **69**(2), 182–195.
- McFADDEN D. (1974) Conditional logit analysis of qualitative choice behavior, in ZAREMBKA P. (Ed.) *Frontiers of Econometrics*, Academic Press.
- MELO P. C., GRAHAM D. J. and NOLAND R. B. (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics* **39**(3), 332–342.
- MIDELFART-KNARVIK K. H., OVERMAN H. G., REDDING S. J. and VENABLES A. J. (2000) The location of European industry, Economic Papers 142, European Commission.
- PALAN N. and SCHMIEDEBERG C. (2010) Structural convergence of European countries, *Structural Change and Economic Dynamics* **21**, 85–100.
- QUIGLEY J. M. (1998) Urban diversity and economic growth, *Journal of Economic Perspectives* **12**(2), 127–138.

- ROMER P. (1986) Increasing returns and long run growth, *Journal of Political Economy* **84**, 1002–1037.
- ROSENTHAL S. S. and STRANGE W. C. (2003) Geography, industrial organization, and agglomeration, *Review of Economics and Statistics* **85**(2), 377–393.
- ROSENTHAL S. S. and STRANGE W. C. (2004) Evidence on the nature and sources of agglomeration economies, in HENDERSON J. V. and THISSE J. F. (Eds.) *Handbook of Regional and Urban Economics, Handbooks in Economics*, vol. 4, chap. 49, pp. 2119–2171, Elsevier, Amsterdam.
- SÜDEKUM J. (2008) Convergence of the skill composition across German regions, *Regional Science and Urban Economics* **38**(2), 148–159.
- WOOLDRIDGE J. M. (2002) *Econometric Analysis of Cross-section and Panel-data*, MIT Press, Cambridge, MA.
- WREN C. and TAYLOR J. (1999) Industrial restructuring and regional policy, *Oxford Economic Papers* **51**(3), 487–516.