Spatial agglomeration and productivity in Italy: a panel smooth transition regression approach

Giulio Cainelli

Department of Economics and Management, University of Padova, via del Santo 33, 35122 Padua, Italy

Andrea Fracasso

Department of Economics and Management, University of Trento, via Inama 5, 38122 Trento, Italy

Giuseppe Vittucci Marzetti*

Department of Sociology and Social Research, University of Milano-Bicocca, Bicocca degli Arcimboldi 8, 20126 Milan, Italy

Abstract

The paper analyzes the nonlinearities in the impact of localization, diversity, urbanization and competition on firm-level TFP, using a large sample of Italian firms from 1999 to 2007. We adopt a Panel Smooth Transition Regression model, so that the TFP elasticities are free to vary smoothly across two or more extreme values. Results show that localization economies and Jacobian externalities materialize only for values of, respectively, intra-industry agglomeration and extra-sectoral diversity above a certain threshold. Local competition exerts a positive effect on productivity, even though the marginal impact shrinks at high levels of competition. We find instead no evidence of diseconomies of agglomeration.

Keywords: Agglomeration economies, Spatial concentration, Nonlinear panel, Smooth Transition Regression, TFP.

JEL Classification: C23, C24, D24, R12, R15.

*Corresponding author: Department of Sociology and Social Research, University of Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milan, Italy. Phone: +39 02 64487457. Fax: +39 02 64487561

Email addresses: giulio.cainelli@unipd.it (Giulio Cainelli), andrea.fracasso@unitn.it (Andrea Fracasso), giuseppe.vittucci@unimib.it (Giuseppe Vittucci Marzetti)

1. Introduction

The effects of spatial agglomeration on productivity have been the object of intense theoretical and empirical research since the seminal work of Marshall (1920).[1] In addition to the classic Marshallian economies (denominated localization economies) that refer to intra-industry externalities (Marshall, 1920; Henderson, 2003a), several contributions in the literature have addressed also urbanization economies (Rosenthal and Strange, 2001; Melo et al., 2009; Combes et al., 2011), inter-industry agglomeration economies connected with variety (Jacobs, 1969; Glaeser et al., 1992; Cainelli and Iacobucci, 2012) and the economies associated with local competition (Porter, 1990).[2]

The actual impact of spatial agglomeration on productivity ultimately remains an empirical issue. While spatial agglomeration may be conducive to positive externalities through the above mentioned channels, congestion effects and rent-seeking behavior tend to have a negative influence on productivity. The impact of the local competition associated with spatial agglomeration is also ambiguous from a theoretical viewpoint. On the one hand, in the face of local competition, firms may try to preserve their market shares and margins through innovation (as suggested by Porter, 1990); on the other hand, competition may depress the returns from innovation, thereby reducing the firms' efforts and results (as in Arrow, 1962; Romer, 1986). Thus, since the seminal work by Glaeser et al. (1992), several contributions have attempted to gauge the impact of the various forms of agglomeration externalities on employment, innovation and productivity.[3]

Notably, while the problems of endogeneity have been addressed in several empirical works investigating the relationship between spatial agglomeration and productivity (see for instance Ciccone and Hall, 1996; Puga, 2010; Combes et al., 2011; Martin et al., 2011), little attention has been devoted so far to the possible nonlinear features of such relationship. This is indeed quite surprising as one would reasonably expect the presence of certain nonlinearities, such as threshold effects or marginally decreasing (or increasing) externalities. In particular, as agglomeration leads to both economies and diseconomies, the estimated impact of agglomeration on productivity is usually an average (weighted across the industry or firms in the sample under investigation) of the *net* effects of economies and diseconomies. There is no reason to assume that these net effects are linearly

[1] Another important strand of the literature on spatial agglomeration has focused on the identification of the direct (localized firm-related spillovers) and indirect (e.g., natural advantage) sources of spatial agglomeration (see, for instance, Ellison and Glaeser, 1999; Rosenthal and Strange, 2001, 2004; Puga, 2010). We shall not address these issues here.

[2] Some authors distinguish static externalities and dynamic externalities (e.g. Glaeser et al., 1992; Henderson, 2003b). The former are one-time efficiency gains generated by spatial concentration. They can account for spatial agglomeration in a homogeneous space, but not for sustained growth differences across regions. The latter are within-industry (Marshall-Arrow-Romer externalities, or dynamic localization economies) and across-industry knowledge spillovers (Jacobs externalities, or dynamic urbanization economies), which can account for the sustained growth differences.

[3] See, among the others, Beaudry and Schiffauerova (2009) and Puga (2010) for a review of the results and Melo et al. (2009) and De Groot et al. (2009) for meta-analyses.

dependent on the level of agglomeration: if economies and diseconomies increase with agglomeration at different rates, the marginal elasticity of productivity to the level of agglomeration is unlikely to be constant. By addressing the possible nonlinearities associated with the net impact of various forms of spatial agglomeration on productivity, this work aims at filling an important gap in the literature.

One possible explanation for the scant attention paid to nonlinear effects can be found in the weaknesses of the traditional methods adopted to deal with nonlinearities. In particular, linear regressions augmented with quadratic and cubic terms (as in Martin et al., 2011) suffer from well-known limitations: i) the linear, quadratic and cubic terms are highly correlated; ii) few outliers can severely affect the results; iii) implausible in- and out-of-sample predicted values tend to be associated with very high/low values of agglomeration. In practice, as we shall show, a specification with quadratic and cubic terms fails to capture some nonlinear effects, produces implausible negative effects for low levels of agglomeration, and prevents from checking for the presence of further nonlinear effects.

Accordingly, this work addresses the nonlinear impact of various forms of spatial agglomeration through the adoption of a Panel Smooth Transition Regression (PSTR) model (González et al., 2005). The latter does not share the weaknesses discussed above: it relaxes both parameter homogeneity and parameter constancy (to account for firm-specific and time-varying elasticities), and allows the impact of spatial agglomeration on productivity to vary smoothly across extreme values. Although this method requires some non-trivial testing and estimating techniques, which shall be duly presented in the methodological section of this work, the results remain easily interpretable and provide interesting insights.

Following the promising approach of Martin et al. (2011), we simultaneously investigate the productivity-enhancing properties of localization, urbanization, diversity and competition, while taking into account their possible nonlinear effects.[4] Starting from an unbalanced panel of more than 12,000 Italian manufacturing firms over the period 1999-2007 (from the AIDA databank by Bureau Van Dijk), we calculate the Total Factor Productivity (TFP) for each firm, following Levinsohn and Petrin's (2003) semi-parametric approach. By choosing the Italian provinces (NUTS 3) as our geographical units, we then estimate the indexes that operationalize the four agglomeration-related variables discussed above. As required by the nonlinear estimation technique, we obtain a balanced panel including more than 7,000 firms on which we estimate the linear specifications, test for the presence of nonlinear effects, and, when it is the case, estimate the nonlinear (PSTR) functional forms. The adoption of yearly data and fixed effects allows to capture the short-run impact of agglomeration on productivity: for this

[4] According to Beaudry and Schiffauerova (2009), only a small number of studies tackled specialization, diversity and competition at the same time. Thus this represents an additional contribution of this work.

reason the work complements those cross-sectional studies that, analyzing time intervals over long periods of time, address the long-run effects of agglomeration accumulated over the years before the observation (Combes et al., 2010). In fact, nonlinear effects related to static externalities, such as labor pooling, input sharing and congestion, are more likely to be detected by exploiting variable changes over short time spans, since the long-term localization patterns are affected by (re)localization choices of individual firms and policy measures.[5]

It is worth noting that the adoption of firm-level data helps to address many of the potential endogeneity problems encountered in the works using industry-level data, as we shall discuss in Section 3. But using disaggregated data is also useful all in itself. As argued by van Oort et al. (2012), showing that regional economic productivity is higher in areas characterized by high spatial agglomeration does not entail that a similar relationship holds at the level of the individual firms.[6] Accordingly, using firm-level data rather than sectoral TFP data (as done, for instance, by Marrocu et al., 2013) allows to address the risk of ecological fallacy and to get closer to the level of the analysis we are more interested in.

To anticipate some of the main findings, localization and variety are shown to have a positive and significant effect on firm productivity only at rather high levels of the variables. Moreover, the rejection of any further nonlinearity associated with localization and variety suggests that no serious congestion effects emerge in the actual concentration patterns of firms: the marginal impact of agglomeration on productivity flattens out after a certain level, but it does not start diminishing. Notably, variety appears significant only once nonlinear effects are allowed for and large values of diversification are reached: by averaging out individual firms conditions, the linear estimates fail to detect such interesting threshold effects. Competition is shown to have a positive effect on TFP, but it turns out that the positive marginal impact is reduced when competition exceeds

---

[5] As suggested by Martin et al. (2011), short-run agglomeration economies are basically related to labor and input market externalities. We therefore exclude a role for spatial knowledge spillovers, which can be mainly captured in a long-run analysis. On the contrary, labor and input market externalities exert their effects on firms TFP in the short-run through two different specific channels. The first one is through labor market pooling. The larger the number of workers with specialized skills in a given geographical area, the larger the expected quality of the matches between firms needs and workers' skills (Helsley and Strange, 1990). The second channel is through local supplier sharing, whereby firms located in agglomerated areas can have access to specialized suppliers and save on transport costs thank to the spatial proximity. In addition, these firms benefit from the better quality of the products of their local suppliers. It is worth noting that our analysis has two main limitations. First, we are not able to discriminate empirically between the two channels. Second, like Martin et al. (2011), we use the term "short-run" as a catch-all term to refer to all the "simultaneous" effects of agglomeration on firm TFP.

[6] Henderson (2003a), van Oort (2007) and Martin et al. (2011), among others, exploit plant-level data in their estimations. This would indeed be preferable to a firm-level analysis for the existence of multi-plant firms. However, no such disaggregated data are available for the Italian firms for a reasonable time span, and, moreover, using plant-level data for employment and firm-level data for value-added and capital, as done by Martin et al. (2011), raises the complex issue of how to allocate the latter across different plants of the same firm.

a critical level. Finally, as often occurs when localization and urbanization are simultaneously included in the estimation, no significant effect of urbanization is found, even allowing for nonlinear effects.

The paper unfolds as follows. In Section 2, we briefly review the literature. In Section 3, we present the baseline linear specifications and discuss the dataset, the variables and the preliminary econometric issues. The PSTR model, together with the methodological issues regarding the tests of (no remaining non) linearity and its estimation, is discussed in Section 4. Section 5 presents the main findings. Section 6 concludes.

2. Background literature

Firms interact with the local environment and positive and negative externalities may emerge from that. The most well known form of such externalities (localization economies) is related to the presence and scale of other firms in the same industry.[7] Marshall (1920) identified three main benefits emerging from the concentration of a given industry in a region: the promotion of knowledge spillovers between firms (as also claimed by Arrow, 1962; Romer, 1986); the emergence of labor market pooling; input-output linkages.[8]

Similar benefits from agglomeration, however, may stem from the features of the entire local production system whose "size" and "structure" may be conducive to demand externalities and cross-fertilization of ideas, information and knowledge. Following the literature, we call the former urbanization externalities and the latter Jacobian externalities.

Clearly, localization and urbanization diseconomies may also arise, in particular out of pollution, congestion, high land prices and the like (Tabuchi, 1998; Higano and Shibusawa, 1999; Zheng, 2001).[9] Accordingly, the sign and the

[7] As an alternative indicator of localization economies, the location quotient (i.e., the regional share of industry employment relative to the national share) has been widely used in the literature since the seminal studies by Glaeser et al. (1992) and Henderson et al. (1995). However, this indicator has no clear theoretical justification. As explicitly acknowledged by Rosenthal and Strange (2004), "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... [While models] 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 cites" (p.2135). The widespread use of the location quotient in papers investigating the impact of agglomeration economies on employment can be rather explained by the fact that the level of employment of an industry in a local area (our measure of localization) is already incorporated in their empirical specifications to account for the mean reversion processes of employment dynamics, and cannot be used twice also to capture localization economies. As our dependent variable is the TFP and not employment, this problem does not apply to our work and there is no reason to use the location quotient.

[8] Albeit standard, this classification is not the only possible one. In their analysis of the micro-foundations of agglomeration economies, Duranton and Puga (2004) suggest a taxonomy based on the various mechanisms at play: sharing, matching and learning.

[9] As stressed by Fujita et al. (1999), the tension between the external economies from spatial concentration and the diseconomies associated with large cities, such as commuting costs, was at the heart of the early models for the analysis of the size distribution of urban areas (e.g.

size of the impact of these economies on growth and productivity (at either the industry or the firm level) are indeterminate at the theoretical level.

Equally debated in the literature is the ultimate effect of local competition on productivity: on the one hand, Jacobs (1969) and Puga (2010) hypothesize that competition spurs efficiency improvements and innovation efforts; on the other hand, Marshall (1920) and Glaeser et al. (1992) contend that local monopoly allows the externalities to be internalized by the innovators and it is thus conducive to higher growth.

The empirical literature devoted to analyze the impact of these economies on employment, productivity and output (levels and growth rates) is huge. In a much cited survey, Rosenthal and Strange (2004) discuss the alternative approaches adopted to address the sources, the scope and the effects of agglomeration economies.[10] With a meta-analysis, Melo et al. (2009) show that urbanization economies are rather small or insignificant in most studies, especially when: localization variables are included in the estimation; panel data, rather than cross-section, analyses are carried out; fixed effects are included. Beaudry and Schiffauerova (2009) cover the literature on Marshallian, Jacobian and competition externalities and attempt at distinguishing the relevance of a Marshallian interpretation of the forces at play (where localization externalities are positive and competition effects negative), and a Jacobian interpretation (where both diversity and local competition have positive effects). The authors are careful in distinguishing the various kinds of indicators (share, size, diversity) and the underlying variables (employment, output, value added, number of firms, and the like) used as proxies in the empirical investigation, and thus show that the results are very sensitive to the choice of the underlying variables and indicators as well as the level of industrial and geographical aggregation, the performance measures considered, the country-time sample, and the empirical method adopted.[11]

Their findings, as well as the conclusions of De Groot et al. (2009), suggest that great care should be used in running the estimations and interpreting the results. Using as a reference another recent work sharing the same baseline specification, thus, appears a reasonable approach to adopt. Accordingly, in the linear specification, we shall follow Martin et al. (2011), who make a promising attempt at tackling the nonlinear effects of localization economies, although only by means of a linear specification augmented of quadratic and cubic terms.

Mills, 1967; Henderson, 1974, 1988).

[10] Very recently, van Oort et al. (2012) review the literature and advocate a multilevel modeling strategy, so as to model the micro and macro levels simultaneously.

[11] Cingano and Schivardi (2004) are among the first to show that focusing on TFP rather than employment growth (as in Combes, 2000) has major consequences on the estimated impact of agglomeration economies. In fact, the employment growth generated by local productivity gains positively depends on the price elasticity of the final demand for the goods, of the local labor supply and of the supply of the other complementary production factors (Combes et al., 2004). So, for instance, when the labor supply in a region depends on the local conditions and these are correlated with the level and structure of agglomeration, employment growth-based estimators might be seriously biased (Cingano and Schivardi, 2004).

3. Baseline specifications and data

3.1. Baseline specification

To analyze the impact of agglomeration and competition on productivity, we adopt the two-step estimation strategy discussed in Martin et al. (2011). In particular, we first compute the series of (value added-based) TFP at the firm level by using data on employment, capital stock and value added. TFP are worked out assuming a Cobb-Douglas production function and using the output elasticities estimated at the industry level by means of Levinsohn and Petrin's (2003) semi-parametric approach.

As in Martin et al. (2011), we then regress these TFP on proxies aimed at capturing the extent of localization externalities, urbanization and Jacobian externalities, and competition. In particular, the baseline linear specification is:

$$a_{it}^{sp} = \beta^l loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \phi_i + \zeta_t + \xi_{it} \tag{1}$$

where the log TFP at time t of firm i located in territory p and active in sector s (a_{it}^{sp}) is regressed on: a localization variable (loc_{it}^{sp}), to capture localization externalities; an urbanization variable (urb_t^{sp}), to capture urbanization economies/diseconomies; an extra-sectoral variety index (div_t^{sp}), to capture Jacobian externalities; an index of local competitive pressure ($comp_t^{sp}$), to capture Porter externalities; ϕ_i for any firm-specific time-constant factor; a time-related component (ζ_t) and an error term (ξ_{it}). More precisely, the time-related component (ζ_t) will be modeled via a set of region-time dummies rather than simple time dummies, whose joint significance is in fact rejected by the data. To eliminate the time-constant unobserved heterogeneity ϕ_i, we use the fixed effects (or within) transformation of the data. We shall come back to the specification of the deterministic components in Section 3.3.

Like in Martin et al. (2011), the localization variable is computed as follows:

$$loc_{it}^{sp} = \ln(emp_t^{sp} - emp_{it}^{sp} + 1)$$

where emp_{it}^{sp} is the number of employees of firm i belonging to sector s and located in territory p at time t, while emp_t^{sp} is the total number of employees of all the firms belonging to s and located in p at t. The exclusion of the employees of firm i from the sum of the employees operating in the same sector and in the same province makes loc_{it}^{sp} vary across firms within the same province-sector pair sp.

The urbanization variable is computed as:

$$urb_t^{sp} = \ln(emp_t^{p} - emp_t^{sp} + 1)$$

where emp_t^p is the sum of the employees of all the firms located in territory p at time t.

The diversity, or variety, variable is calculated as the entropy index of local employment shares outside the sector:

$$div_t^{sp} = -\sum_{s' \in S \setminus s} \frac{emp_t^{s'p}}{\sum_{s' \in S \setminus s} emp_t^{s'p}} \log_2 \left(\frac{emp_t^{s'p}}{\sum_{s' \in S \setminus s} emp_t^{s'p}} \right)$$

where $emp_t^{s'p}$ is the number of employees of firms belonging to sector s' ($\neq s$) and located in p at time t.[12]

Finally, the strength of local competitive pressure is measured by the log of the inverse Herfindahl-Hirschman index of market concentration:

$$comp_t^{sp} = -\ln\left(\sum_i \left(\frac{va_{it}^{sp}}{va_t^{sp}}\right)^2\right)$$

where va_{it}^{sp} is the value added of firm i in sector s located in territory p at time t and va_t^{sp} is the total value added of all the firms in sector s and territory p at t.[13]

Differently from loc_{it}^{sp}, the regressors urb_t^{sp}, div_t^{sp} and $comp_t^{sp}$ do not vary across firms within the same province-sector pair sp. Thus, to avoid the so-called Moulton's (1990) problem associated with explanatory variables observed at a more aggregate level than the dependent variable, in all the estimations we compute cluster-robust standard errors, where a cluster is any province-sector pair, and report the number of province-sector pairs in the sample, i.e. the number of observations available to estimate the parameters associated with urbanization, variety and competition.

Finally, we acknowledge that focusing only on manufacturing firms may be too restrictive since non-manufacturing sectors add to extra-sectoral variety and non-manufacturing firms can actually benefit from agglomeration economies. In particular, some services have become highly integrated with manufacturing and their share in the economy steadily increased over time.[14] In spite of this, we decided not to include services in our analysis for two main reasons. First, the joint stock companies in our sample (see next section) are assigned to the Italian provinces according to the location of their headquarter. This might constitute a problem for service sectors, in particular for trade, transport, banking and financial sectors, where firms are usually large, and so, looking at their headquarter location, might be misleading about the localization patterns. This does not represent a problem for manufacturing since these firms are generally

[12] In Martin et al. (2011), div_t^{sp} is defined as the log of the inverse Herfindahl index of local employment shares of the sectors different from s. We use the entropy index as it does not require any further transformation (it is already a weighted average of logs) and, more importantly, it is a more standard measure of variety given its decomposability property (see, for instance, Frenken et al., 2007). We have nonetheless estimated all the models also with Martin et al.'s (2011) alternative index. The results do not significantly change.

[13] Although Martin et al. (2011) compute the same index using employment, the use of value added seems more appropriate. Admittedly, the Herfindahl-Hirschman index has a number of shortcomings when used as measure of competition. For instance, foreign competition is not taken into account.

[14] See, for instance, Montresor and Vittucci Marzetti (2011) for an analysis of the integration of services in manufacturing in terms of value added and employment, and Marrocu et al. (2013) for a recent analysis including services at the region-sector level about the role played by agglomeration externalities (specialization and diversity) in the productivity differentials across Europe. This integration process is particularly relevant for knowledge-intensive service sectors, where the nature, effects and channels of agglomeration forces can be more easily identified and analyzed (see, for instance, Antonietti et al., 2013).

small and single-plant. Second, for services there is also a problem of quality and data coverage, in particular as far as the number of employees is concerned. This explains why most of the Italian studies on the relationship between firm-level productivity and agglomeration use data on manufacturing firms (e.g. Cingano and Schivardi, 2004).[15]

3.2. Data

We start from an unbalanced panel of Italian manufacturing firms over the period 1999-2007. Data are taken from the AIDA databank (Bureau Van Dijk).

For each firm we have balance-sheet data on value added, employment, equipment, investment and total cost of labor, together with information about the firm's location and its industry classification. The industrial classification used is ATECO 2007, i.e. the Italian version of the European nomenclature NACE Rev. 2. We compute fixed capital using the perpetual inventory method.[16] After a first round of cleaning up and having dropped the first year as we use lagged values in some specifications, we end up with an unbalanced panel of 12,513 firms (104,725 obs).[17]

These data are used to estimate output elasticities at the two-digit industry level following Levinsohn and Petrin's (2003) semi-parametric approach.[18] Estimates of output elasticities range from 0.44 to 0.82 for labor and from 0.01 to 0.46 for capital. This exercise requires to choose the level of industry disaggregation by striking a balance between ensuring the homogeneity of the output elasticities within each sector and preserving a sufficient number of observations by sector. Following the approach in Martin et al. (2011), we estimate the output elasticities at the industry level (2-digit), as often done in other studies focusing on Italian manufacturing data.[19]

[15]The papers extending the analysis to service sectors usually employ census data (e.g. Paci and Usai, 2008).

[16]The stock of capital of firm i at time t (K_{it}) is computed as follows:

$$K_{it} = (1-\delta)\frac{p_t^I}{p_{t-1}^I}K_{i,t-1} + I_{it}$$

where δ is the depreciation rate (assumed equal to 0.085), I_{it} the investment of i at t, and p_t^I the investment good prices at t (drawn from the Italian National Accounts).

[17]In particular, we have: removed the firms with no positive values of turnover and value added for seven consecutive years over the period 1998-2007; made a mild trimming (0.1% on both tails of the distribution) on sales growth (13,282 valid obs); removed firms with incomplete balance sheet data for total cost of labor or fixed capital, no establishment year or province location (NUTS 3), value added-turnover ratio negative or greater than one.

[18]Nominal values on fixed capital and value added have been deflated by using deflators at the same two-digit level. After the estimation, two sectors—manufacture of coke and refined petroleum products (Ateco 2007, 12) and manufacture of tobacco products (Ateco 2007, 19)—have been dropped because of too few observations and inconsistent TFP-related results.

[19]Using 3-digit industry level would force us to drop a number of sectors and, due to limited data availability, we would risk running a not fully reliable TFP estimation also on the remaining sectors.

Table 1: Data description

Variable		Mean	Std. Dev.	Skewness	Kurtosis	Min	Max	Obs
a_{it}^{sp}	overall	1.954	2.328	-0.343	3.182	-9.853	9.372	N = 58,784
	between		1.714	0.117	2.805	-3.916	7.653	n = 7,348
	within		1.576	-0.873	3.748	-9.318	6.609	T = 8
loc_{it}^{sp}	overall	5.625	2.421	-1.039	3.557	0	12.490	N = 58,784
	between		2.383	-1.059	3.524	0	9.716	n = 7,348
	within		0.431	-3.320	44.069	-1.238	9.000	T = 8
urb_{t}^{sp}	overall	9.848	1.307	-0.063	3.897	2.944	12.931	N = 58,784
	between		1.287	-0.094	3.662	3.655	12.375	n = 7,348
	within		0.227	-0.491	9.877	6.565	11.326	T = 8
div_{t}^{sp}	overall	4.529	0.788	-1.619	6.084	0	5.504	N = 58,784
	between		0.711	-1.977	7.835	0	5.354	n = 7,348
	within		0.341	-3.803	24.563	2.264	5.514	T = 8
$comp_{t}^{sp}$	overall	1.235	0.900	0.405	2.301	0	3.553	N = 58,784
	between		0.836	0.403	2.443	0	3.370	n = 7,348
	within		0.333	-0.544	6.263	-0.750	2.545	T = 8

From the estimated output elasticities, we compute the log of TFP of each firm in each given year (a_{it}^{sp}). Then, we compute the other variables of Eq. (1) (loc_{it}^{sp}, urb_{t}^{sp}, div_{t}^{sp}, $comp_{t}^{sp}$) at the NUTS 3 level —i.e., the territory p is the Italian province—and the 3-digit industry level (Ateco 2007). [20]

Finally, for the requirements of the nonlinear estimation, which we discuss in Section 4, we balance the panel to preserve comparability of linear and nonlinear estimations. The resulting balanced panel has got 58,784 observations: 7,348 firms; 8 periods; 88 three-digit sectors (21 two-digit sectors); 97 Italian provinces.[21] Table 1 reports the main summary statistics for all the variables.

3.3. Preliminary linear results and econometric issues

The estimates of Eq. (1) for three alternative specifications of the time-related effects are reported in Table 2: simple time dummies (column I); region-time dummies (column II), where the regions correspond to the NUTS 2 level; industry-time dummies (column III), where the industries are identified according to the

[20] While in the estimation of the firm-level TFP we adopt a 2-digit industry classification so as to ensure the sectoral homogeneity of the output elasticities and to preserve a sufficient number of sectors, to calculate the agglomeration variables we need to choose a level of sectoral aggregation that allows to distinguish localization and urbanization economies. Intuitively, agglomeration indexes at 2-digit sectoral level risk attributing urbanization economies to localization economies as confirmed by Beaudry and Schiffauerova (2009), who find that the 3-digit sectoral classification is in general more appropriate to measure agglomeration and urbanization economies. Finally, we intend to preserve direct comparability with Martin et al. (2011) by following their empirical approach as close as possible, in particular in their sections devoted to the nonlinear results. This entails a two-fold level of analysis with the TFP estimated at the 2-digit level and agglomeration variables calculated at the 3-digit level.

[21] As for the distribution of firms among provinces, the provinces with fewer firms are: Campobasso, Enna, Nuoro (2), and Aosta, Catanzaro, Potenza, Siracusa, Vibo-Valentia (3); whereas the provinces with more firms are: Milano (1074); Bergamo (405); Varese (329); Brescia (322); Torino (289); Treviso (264); Vicenza (231); Modena (221); Bologna (220); Verona (216).

Table 2: Fixed-Effects estimates of the alternative linear specifications

	I	II	III
loc_{it}^{sp}	0.0525	0.0535	0.0534
	(0.0445)	(0.0435)	(0.0415)
urb_t^{sp}	0.0117	-0.0002	0.0001
	(0.0592)	(0.0744)	(0.0589)
div_t^{sp}	0.0143	0.0163	0.0067
	(0.0308)	(0.0389)	(0.03162)
$comp_t^{sp}$	0.4178***	0.4183***	0.4307***
	(0.0526)	(0.0520)	(0.0493)
Estimator	FE	FE	FE
Hausman test statistic (H$_0$: RE and FE consistent; RE efficient)	14.767	10.087	149.264
Hausman test p-value	0.0391	0.0390	0.000
Time-related dummies	Time	Region-time	Industry-time
Wald test statistic (F-form) (H$_0$: no time-related dummies)	1.49	61.99	1.64
Wald test p-value	0.1652	0.0000	0.0000
Observations	58,784	58,784	58,784
Years	8	8	8
Firms	7,348	7,348	7,348
Sectors (3-digit)	88	88	88
Provinces (NUTS 3)	97	97	97
Sector-province pairs	2,234	2,234	2,234
Sector-province observations	17,872	17,872	17,872

Dependent variable: a_{it}^{sp}. Years: 2000-2007. Region: NUTS 3 level. Province: NUTS 2 level. Industry: 2-digit ATECO 2007. Sector: 3-digit ATECO 2007. Cluster-robust standard errors in parentheses (cluster: sector-province pair). Significance at: 1% ***; 5% **; 10% *.

2-digit ATECO 2007 classification. In all the cases, we use a Fixed-Effects (FE) estimator, where the unobserved firm-specific time-constant effect is eliminated via the within transformation of the data. Moreover, to avoid the Moulton's (1990) problem connected with the fact that some of the regressors do not vary within each province-sector pair, we compute the cluster-robust standard errors and adopt them in all the tests.

Table 2 also reports the Hausman test statistics, according to which we reject the null of Random-Effects (RE) at the 5% level in all the specifications. This implies that the fixed effects are jointly statistically significant and correlated with the regressors.

In all the estimations, the TFP elasticity to localization (*loc*) is positive and in line with the literature: Rosenthal and Strange (2004) report that the elasticity of productivity to the size of the industry generally lies between 0.03 and 0.08. However, we fail to reject the hypothesis that the coefficient is not statistically different from zero when we adopt cluster-robust standard errors.

The TFP elasticities to urbanization (*urb*) and diversity (*div*) are close to zero and not statistically significant (even using the usual OLS standard errors). This is again in line with the literature: according to the meta-analysis by Melo et al. (2009), the estimated impact of urbanization is lower when, as in the present case, localization economies are included, panel data rather than cross-section data are used, and FE rather than RE or pooled OLS estimators are used. Moreover, Beaudry and Schiffauerova (2009) show that both the elasticities are mostly insignificant in cases similar to ours in terms of: administrative geographical unit (NUTS 2); medium sectoral disaggregation (3-digit); dependent variable (TFP rather than value added per worker). As observed by Martin et al. (2011), urbanization economies are typically thought as long-term externalities which the short-term kind of analysis conducted here, using yearly data and fixed effects, can hardly capture. Finally, the impact of competition on productivity (*comp*) is positive, large and highly statistically significant, supporting Porter's (1990) intuition and in line with Glaeser et al. (1992).

However, these estimates may suffer from a number of problems and limitations, which we briefly discuss in what follows. In particular, although FE estimators can cope with time-constant unobserved heterogeneity, they cannot address issues related with: omitted time-variant industry- and/or location-specific factors; simultaneity bias; firm demography; geographical unit; nonlinearities in the effects of agglomeration on productivity (see also Martin et al., 2011, where these problems are discussed in details).

As for the presence of time-variant industry- and/or location-specific omitted factors,[22] a simple, although admittedly rather rough, way to cope with the problem is to add region-time and/or industry-time dummies to the baseline specification, as done by Henderson (2003a).[23] We therefore estimate the baseline model including either region-time dummies (NUTS 2) or industry-time (2-digit) dummies. As shown in columns II-III of Table 2, none of the estimates significantly change across the three specifications. It seems also worth noting that no firm in the sample changes location or sector during the period considered, and this is something that likely reduces the role of time-variant location/industry specific factors. Given our interest in the geographical dimension of TFP externalities, we take a conservative stance and choose specification II as our baseline linear specification in the remaining of the paper; accordingly, we shall encompass a set of region-time dummies to capture the deterministic time-related components in all the nonlinear specifications investigated hereafter.

As for the possible simultaneity bias, although it is discussed at some length by Martin et al. (2011) and actually presented as one of the main reasons they adopt a GMM approach, we do not think it to be very relevant in the present context. In fact, when addressing the possible simultaneity bias, Martin et al. (2011, p.184) actually combine the endogeneity bias coming from omitted variables,

[22]For example, a negative (positive) economic shock in a region/industry can change the TFP of a given firm and also lead the other firms in the region/industry to close (open) or lay off (hire) employees: this would give rise to spurious regressions.

[23]A more systematic way to deal with location-specific unobserved factors is to estimate a model with spatial lags and/or spatially autocorrelated errors (LeSage and Pace, 2009). Given that the main focus of this paper is the analysis of the nonlinear effects of agglomeration, while the linear estimates in this section are rather preliminary, we do not apply such techniques here.

and the proper simultaneity bias, occurring when the right-hand variable (e.g. *loc* or *urb*) is simultaneously determined along with the left-hand variable (a). Omitted variables, in the form of (time-variant) industry- or location-specific factors, are important in principle, but we control for them by adding region-time and industry-time dummies. The proper simultaneity bias, in turn, is not a major concern in the present application. Indeed, since the dependent variable is the TFP at the firm level and *loc*, *urb* and *div* are all defined without counting the workers of the firm whose TFP is explained,[24] it is hard to imagine how such variables might be significantly affected by idiosyncratic productivity shocks at the firm level, especially when the firm itself accounts for just a small portion of the overall production in the sector-area, like in Italy where nearly all firms are small and medium-sized (see the Appendix for a formal analysis of reverse causality issues).

Firm demography and spatial sorting are two additional issues that we do not directly address in the paper because of the adoption of a balanced panel, needed to carry out the nonlinear tests and estimations. For instance, either if the survival of firms is correlated with agglomeration or if firms choose their optimal location by taking agglomeration economies into account, the estimators can be biased. Notwithstanding the adoption of a balanced panel, we argue that these issues should not be overemphasized in the present analysis. First, the FE estimator already purges the time-constant effects, which are clearly much more important than time-varying ones given the short time span. Second, all the variables (*loc*, *urb*, *div* and *comp*) are calculated on the entire unbalanced sample. Even at the theoretical level, optimal locational choices are guided more strongly by long-run agglomeration effects on productivity than by the short-run effects addressed here: this is confirmed by the absence of spatial mobility in our sample, in line with Martin et al. (2011) and Combes et al. (2012).

We compute our measures of agglomeration at the province level (Ciccone, 2002; Cainelli and Iacobucci, 2012): the Italian territory is split into 103 provinces (NUTS 3 level of the EU geographical classification). An alternative geographical unit would have been the Italian Local Labor System (LLS), i.e. a set of contiguous municipalities characterized by a self-contained labor market (e.g. Cingano and Schivardi, 2004). In fact, using the 2001 population census, the Italian Statistical Institute (ISTAT) identifies 686 LLS. Most of them are composed of two (69/686) or three municipalities (74/686) with very few firms (including services). They are classified as LLS because within their territory people commute for working. Only when these systems satisfy particular conditions they are also classified as industrial districts. In our opinion, this can have two main implications. First, these systems do not provide a sufficient level of "critical mass" in terms of number of firms and employees. As suggested by Brunello

[24] This argument does not hold for *comp*, where the value-added of the firm enters the calculation of the index. In fact, a productivity shock at the firm level could actually "cause" a decrease of *comp*. This is something that possibly leads to a (likely slight) downward bias of the estimator of the impact of competition on TFP.

and De Paola (2008), "travel to work" areas are too small to take into account all relevant spillovers. Second, as our sample is composed of 7,348 firms, most local labor systems include very few firms and several local systems would thus be excluded from the analysis, thereby producing a sort of "spatial" sample selection. For these reasons, we choose to use the province as the geographic unit of our investigation.

3.4. Preliminary nonlinear results

A final issue to consider, on which the remaining of this work shall focus, is related to the possible presence of nonlinearities in the impact of the variables on productivity: the linear specification cannot capture the nonlinearities of agglomeration economies, such as the possible presence of critical masses or congestion effects. In this case, any effect estimated by the linear model (1) turns out to be just the average net effect of agglomeration economies and diseconomies.

A simple way to account for the existence of such nonlinearities is to introduce in the baseline regression both quadratic and cubic terms, as done by Martin et al. (2011). The results of these specifications are reported in Table 3. In columns IV to VII we introduce the quadratic and cubic terms for each of the variables at a time. The Wald tests for the joint significance of the nonlinear terms reveal that only for *loc* and *comp* there is evidence of nonlinearities. In column VIII we report the results when all the quadratic and cubic terms are jointly estimated and the same results hold. Finally, column IX reports the estimates of a specification where only the quadratic and cubic terms for *loc* and *comp* are included, again with no major change in the results. In sum, the quadratic and cubic terms of localization and competition variables are significant in the various specifications. This points to possible nonlinear patterns in the productivity impact of agglomeration, at least for what concerns these two variables. The estimates suggest that the marginal effect of localization starts low and then grows with the level of localization, while the marginal impact of competition has a U-shaped pattern.

However, the simple inclusion of quadratic and cubic terms in the linear model might not address properly the possible nonlinearities. The results of these augmented regressions should be taken at most as suggestive because of their several limitations: linear, quadratic and cubic terms tend to be highly correlated; outliers might drive the results; the estimates imply implausible in- and out-of-sample predicted values for very high/low values of agglomeration. So, for instance, in our sample, when the initial level of localization is low the marginal total impact of localization on TFP implied by the point estimates of the coefficients is negative and decreases by increasing localization further. Such impact becomes positive only when localization becomes high. This result, in fact similar to what found by Martin et al. (2011, Figure 1 and 2, p. 192), looks odd and suggests that the linear specification with quadratic and cubic terms is not fully appropriate.

The possible existence of critical masses and/or congestion effects could be better captured by Hansen's (1999b) Threshold Regression (TR) model, as this

Table 3: Fixed-Effects estimates of the augmented linear specifications

	IV	V	VI	VII	VIII	IX
loc_{it}^{sp}	0.0910	0.05325	0.0538	0.0163	-0.0481	-0.0444
	(0.0638)	(0.0436)	(0.0434)	(0.0528)	(0.0666)	(0.0662)
$(loc_{it}^{sp})^2$	-0.0358**				-0.0206	-0.0210
	(0.0139)				(0.0143)	(0.0144)
$(loc_{it}^{sp})^3$	0.0032***				0.0028***	0.0028***
	(0.0007)				(0.0008)	(0.0008)
urb_t^{sp}	0.0038	-0.6639	-0.0058	0.0055	-0.9810	.0051
	(0.0727)	(0.6518)	(0.0759)	(0.0752)	(0.6869)	(0.0723)
$(urb_t^{sp})^2$		0.0789			0.12225	
		(0.0763)			(0.0807)	
$(urb_t^{sp})^3$		-0.029			-0.0048	
		(0.0028)			(0.030)	
div_t^{sp}	-0.0000	-0.0025	0.0751	0.0191	0.1900	-0.0016
	(0.0388)	(0.0642)	(0.468)	(0.0398)	(0.4793)	(0.0392)
$(div_t^{sp})^2$			-0.0011		-0.278	
			(0.1405)		(0.1404)	
$(div_t^{sp})^3$			-0.0010		-0.0005	
			(0.0136)		(0.0132)	
$comp_t^{sp}$	0.4709***	0.4183***	0.4183***	1.4947***	1.6945***	1.6920***
	(0.0426)	(0.0521)	(0.0522)	(0.1657)	(0.1267)	(0.1270)
$(comp_t^{sp})^2$				-0.6528***	-0.7469***	-0.7454***
				(0.1108)	(0.1047)	(0.1051)
$(comp_t^{sp})^3$				0.0996***	0.1172***	0.1167***
				(0.0248)	(0.0246)	(0.0247)
Estimator	FE	FE	FE	FE	FE	FE
Region-time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wald test statistic (F-form) (H_0: no region-time dummies)	65,324	10,879	4,961	440,000	750,000	320,000
Wald test *p*-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Wald test statistic (F-form) (H_0: no quadratic and cubic terms)	41.34	0.53	0.10	44.41	35.67	65.28
Wald test *p*-value	0.0000	0.5858	0.9087	0.0000	0.0000	0.0000
Observations	58,784	58,784	58,784	58,784	58,784	58,784
Years	8	8	8	8	8	8
Firms	7,348	7,348	7,348	7,348	7,348	7,348
Sector-province pairs	2,234	2,234	2,234	2,234	2,234	2,234
Sector-province observations	17,872	17,872	17,872	17,872	17,872	17,872

Dependent variable: a_{it}^{sp}. Years: 2000-2007. Region: NUTS 3 level. Province: NUTS 2 level. Industry: 2-digit ATECO 2007. Sector: 3-digit ATECO 2007. Cluster-robust standard errors in parentheses (cluster: sector-province pair). Significance at: 1% ***; 5% **; 10% *.

allows the productivity impact of the variables (i.e., the coefficients) to take a discrete number of different values (called "regimes"), and the switch of a cross-sectional unit from one regime to another is driven by a chosen "transition variable" (in this case measuring the level of agglomeration) being above or below a given identified "threshold".

Though more flexible than the linear model, the TR approach still suffers from an important limitation: it imposes abrupt transitions across the regimes, i.e. the estimated impact of agglomeration jumps when the transition variable crosses any identified threshold. This rules out any continuous variation of the productivity impact as well as any smooth transition of the cross-sectional units from one regime to another. To address this limitation, González et al. (2005) and Fok et al. (2005) put forward the Panel Smooth Transition Regression (PSTR) model, building on the work of Granger and Teräsvirta (1993) on Smooth

Transition Autoregressive (STAR) models.

The PSTR approach we adopt in this paper allows to relax the hypotheses of homogeneity and time-stability of the parameters in a convenient and flexible way, since they are allowed to change smoothly between the values associated with two (or more) extreme regimes. As in the TR model, the change is still driven by a transition variable (in this application capturing the level of agglomeration), but in the PSTR model the transition can be smooth rather than abrupt. Both the threshold (called "location parameter" in the PSTR literature) and the smoothness of the transition are estimated from the data, thereby making the linear and the TR models nested in the PSTR one. Since the transition variable is time-varying and unit-specific, the coefficients (measuring the impact of agglomeration on productivity) for each of the cross-sectional units in the panel gradually change over time. Thus, the coefficients are not forced to assume at each point in time and for each unit the values associated with either of the extreme regimes, but are let free to vary within them. The adoption of a bounded and continuous (typically, a logistic) function of the transition variable to guide the fluctuations of the coefficients across the extreme regimes guarantees their gradual variation, in contrast with the dichotomous switches superimposed by the TR model.

It is worth stressing that, since the PSTR model nests both the TR model and the linear one, using a PSTR model neither excludes the possibility of estimating a TR-like model, nor it imposes a nonlinear shape to what is in fact a linear relation.

In the next section we will illustrate in some details the PSTR model along with the main methodological issues related with the preliminary tests of nonlinearity and the subsequent model estimation.

4. Nonlinear specification

To illustrate the rationale and the features of the PSTR model relevant in the present application, let us start from the baseline linear specification (1) and assume that, instead of being constant, the TFP elasticity to localization (the coefficient β^l) can vary across units and time, i.e.:

$$a_{it}^{sp} = \beta_{it}^{l}\, loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \phi_i + \zeta_t^p + \xi_{it} \tag{2}$$

where, on the basis of the results discussed in the previous section, ζ_t^p are region-time effects.

In a way consistent with the PSTR model, we assume that:

$$\beta_{it}^l = \beta_0^l + \sum_{j=1}^{r} \beta_j^l\, g(q_{it}; \gamma_j, c_j) \tag{3}$$

where:

$$g(q_{it}; \gamma_j, c_j) = \frac{1}{1 + \exp\Big(- \gamma_j (q_{it} - c_j)\Big)}. \tag{4}$$

According to Eq. (3), the TFP elasticity to localization is a weighted average of the coefficients associated with the $r+1$ regimes, with weights given by the r Eqs. (4). These weights are logistic functions of a transition variable q_{it}—the variable that guides the transition across regimes—, where c_j is a location parameter—the critical level separating two contiguous regimes—and the parameter γ_j (>0) determines the smoothness of the transition across the regimes.[25]

When $\gamma_j \to \infty$, for all j, the r transition functions $g(.)$ become indicator functions and the model reduces to a simple panel TR model. On the contrary, when $\gamma_j \to 0$, the r functions $g(.)$ become constant and the model collapses to a simple panel linear regression model with fixed effects.

Finally, it is worth stressing that the procedure, to be discussed in what follows, provides for the estimation of all the parameters of interest, including the number of nonlinear regimes (r) and the associated location (c_j) and smoothness (γ_j) parameters, so that no *a priori* identification of the number and values of the (extreme) regimes, or the shape of the transitions between them, is required.

4.1. Nonlinearity testing

The first preliminary step (see González et al., 2005, for a thorough discussion) is to test the linearity of Eq. (1) against a PSTR model with two regimes (Eqs. (2)-(4) with $r=1$) and a given transition variable (q_{it}), that is:

$$\begin{aligned} a_{it}^{sp} = & \left(\beta_0^l + \beta_1^l\, g(q_{it};\gamma_1,c_1)\right) loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta^d div_t^{sp} \\ & + \beta^c comp_t^{sp} + \phi_i + \zeta_t^p + \xi_{it} \end{aligned} \tag{5}$$

Testing linearity is a non-standard problem since under the null of linearity there are unidentified nuisance parameters.[26] To solve the identification problem, Hansen (1999a,b, 2000) proposes to compute the supremum LR test statistic to test the null $\beta_1^l = 0$. Andrews and Ploberger (1994) suggest instead to use alternative statistics: AveLM, ExpLM or wLM, i.e. weighted averages of (heteroskedasticity-robust) LM-test statistics computed for several combinations of γ_1 and c_1 spanning the parameter space.[27] Since these statistics have (asymp-

[25] The PSTR model is actually more general: it can accommodate different transition variables, one for each nonlinear regime, and/or more complex transition patterns across pairs of identified extreme regimes (González et al., 2005).

[26] Linearity follows by imposing either $\beta_1^l = 0$ or $\gamma_1 = 0$. When the null is $\beta_1^l = 0$ $(\gamma_1 = 0)$, c_1 and γ_1 (β_1^l) are unidentified nuisance parameters.

[27] AveLM, ExpLM and wLM are, respectively, the average test statistic, the exponential average test statistic and the weighted average test statistic. They are defined as follows:

$$\begin{aligned} \text{AveLM} &= \int_{\gamma_1}\int_{c_1} \text{LM}(\gamma_1,c_1)\ \mathrm{d}W(\gamma_1,c_1) \\ \text{ExpLM} &= \ln\left(\int_{\gamma_1}\int_{c_1} \exp\left(\frac{1}{2}\,\text{LM}(\gamma_1,c_1)\right) \mathrm{d}W(\gamma_1,c_1)\right) \\ \text{wLM} &= \int_{\gamma_1}\int_{c_1} \omega(\gamma_1,c_1)\ \text{LM}(\gamma_1,c_1)\ \mathrm{d}W(\gamma_1,c_1) \end{aligned}$$

where $W(\gamma_1,c_1)$ is the weight function that allocates weights on the pairs (γ_1,c_1), while totically pivotal but) non-standard distributions, which depend on the moments of the distribution of the nonlinear parameters and whose critical values cannot therefore be tabulated, to obtain the critical values one has to bootstrap the tests.[28]

In this literature, Hurn and Becker (2009) and Becker and Osborn (2012) have recently dealt with the problem of heteroskedasticity and the distortions the latter causes to the size of the test in small samples. Dealing with heteroskedasticity in nonlinearity tests can be problematic: on the one hand, neglecting heteroskedasticity may lead to reject the null of linearity when it is not the case; on the other hand, robustification can remove most of the test power (Lundbergh and Teräsvirta, 1998). To cope with this, Hurn and Becker (2009) suggest to compute heteroskedasticity-robust test statistics and calculate the critical values of the tests using fixed-design wild bootstrap (Gonçalves and Kilian, 2004). They show via simulation that this leads to a significant reduction in the distortions of the test. Accordingly, we follow their approach and use heteroskedasticity-robust test statistics.[29]

It is worth pointing out that the PSTR procedure requires to test for the presence of nonlinearities associated with each variable at a time. For this reason we shall first test and estimate the nonlinearities associated with each of the various agglomeration variables and, only subsequently, we shall estimate the model with all the nonlinear effects at once.

4.2. Estimation of the PSTR specification

If the null of linearity is rejected, a two-regime PSTR model is estimated (Eq. (3) with $r = 1$). The estimation is carried out by minimizing a concentrated Sum of quadratic Residuals (SSR). The SSR is concentrated with respect to the fixed effects ϕ_i and the linear coefficients β's by applying a standard FE estimator for panel data conditional on a given combination of the nonlinear parameters (c_1 and γ_1) characterizing an iteration. Spanning a meaningful set of combinations of such nonlinear parameters, the conditional FE estimates are recomputed at each iteration in the nonlinear optimization and the concentrated SSR is calculated.

$\omega(\gamma_1, c_1)$ is the weight function on LM, with weights proportional to the magnitude of the values of the LM statistic, for the test not to be too heavily influenced by redundant values of γ_1 and c_1, that may have a negative effect on its power.

[28] An alternative way to deal with the identification problem is by testing the null $\gamma_1 = 0$ via a m-order Taylor expansion of the nonlinear model around this point (Luukkonen et al., 1988). González and Teräsvirta (2006) have studied the finite sample properties of Andrews (1993) and Andrews and Ploberger's (1994) test statistics (SupLM, AveLM, ExpLM or wLM) and compared them with the Taylor expansion-based test, thus showing that AveLM, ExpLM or wLM are always more powerful than SupLM and Taylor expansion-based tests. This is why we decided not to report such tests, that we have nonetheless computed. The results (available at request) do not alter the conclusions.

[29] The alternative heteroskedasticity-robust bootstrap procedure, discussed in Hansen (1999a) for TR models, is able to preserve the observed heteroskedasticity, but it does not exactly reproduce the heteroskedastic pattern of the data.

To explore the set of combinations of the nonlinear parameters and then find the minimum of the concentrated SSR and the corresponding minimizers, i.e. $(\hat{c}_1, \hat{\gamma}_1)$, we follow González et al. (2005) and González and Teräsvirta (2006) and implement the Simulated Annealing (SA) algorithm proposed by Corana et al. (1987) (see also Goffe et al., 1994, for an application to M-estimation problems).

SA—so named as it resembles the process undergone by the atoms in a heated metal when it cools slowly—denotes a large class of probabilistic algorithms used to locate global minima/maxima of functions in large search spaces, when the problem is unmanageable using combinatorial or analytical methods. SA improves more standard iterative optimization algorithms by introducing the "Metropolis criterion": some steps are taken in the "wrong direction" with a certain probability, as they serve to better explore the space of possible solutions. The probability of taking the wrong direction decreases when no significant improvements in the result is reached after many consecutive iterations (this is regulated by the decrease in the "temperature", again by analogy with the annealing of a metal).[30]

Finally, since it is not desirable to identify a regime containing only very few observations, we also check that the estimated location parameters are within the 5–95th percentiles of the sample values of the transition variables, so that each regime can be estimated using at least 5% of all the observations.[31]

4.3. Testing for no remaining nonlinearity

After the estimation of a two-regime PSTR model, it is necessary to test the hypothesis that it adequately captures the nonlinearities in the panel and no additional nonlinearities are present. We follow González et al. (2005) and perform a test of (no remaining) nonlinearity on the following specification:

$$\begin{aligned} a_{it}^{sp} &= \left(\beta_0^l + \beta_1^l\, g(q_{it}; \hat{\gamma}_1, \hat{c}_1) + \beta_2^l\, g(q_{it}; \gamma_2, c_2)\right) loc_{it}^{sp} + \beta^u urb_t^{sp} \\ &\quad + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it} \end{aligned} \tag{6}$$

where $\hat{\gamma}_1$ and $\hat{c}_1$ are the estimates of the nonlinear parameters in the two-regime PSTR model.

The rejection of the null at this stage implies that the variation of the parameters of the model is not fully captured by a two-regime PSTR model. This suggests that the parameters should be let change across time and units in a more complex way. In this case, the unit and time-varying parameters become

[30] Corana et al.'s (1987) algorithm is just one of the many proposed in the literature. For a systematic treatment of SA see Otten and van Ginneken (1989).

[31] This restriction on the values that the locations can take has the side-effect of preventing, especially in the case of very smooth transitions, the detection of locations in the tails of the distribution of the transition variables. Having said that, given the need of a sufficiently large number of observations in each regime for the estimation of its parameters, in the literature the 5–95th percentiles of the transition variable are typically used to determine the boundaries of the location parameters. Although the large size of our panel would allow for percentiles lower (higher) than the 5th (95th), we nevertheless follow the standard approach.

weighted averages of the three β_n^l parameters ($n \in \{0,1,2\}$) characterizing the three extreme regimes.

Proceeding as before, this test of (no remaining) nonlinearity is performed testing $\beta_2^l = 0$ by computing AveLM, ExpLM or wLM test statistics.

Following a sequential procedure, as in González et al. (2005), the test can be generalized to a generic number of regimes to determine the number of transitions in the model. In particular, after the estimation of a model with $r+1$ regimes, one can perform a nonlinearity test on:

$$\begin{aligned} a_{it}^{sp} = & \left(\beta_0^l + \sum_{j=1}^{r} \beta_j^l \, g(q_{it}; \hat{\gamma}_j, \hat{c}_j) + \beta_{r+1}^l \, g(q_{it}; \gamma_{r+1}, c_{r+1}) \right) loc_{it}^{sp} \\ & + \beta^u urb_t^{sp} + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it} \end{aligned} \tag{7}$$

where the null is $\beta_{r+1}^l = 0$. If it is rejected, a $(r+2)$-regime PSTR model can be estimated and one can continue adding regimes until the first acceptance of the null of no remaining nonlinearity.

4.4. Nonlinear specifications

To analyze the nonlinearities in the effects of agglomeration variables on firm productivity, we assume that each one of the lagged variables is the transition variable of the nonlinear regimes associated with the variable itself, that is:[32]

$$\beta_{it}^l = \beta_0^l + \textstyle\sum_{j=1}^{r_l} \beta_j^l \, g(loc_{i,t-1}^{sp}; \gamma_j, c_j) \tag{8}$$

$$\beta_{it}^u = \beta_0^u + \textstyle\sum_{j=1}^{r_u} \beta_j^u \, g(urb_{t-1}^{sp}; \gamma_j, c_j) \tag{9}$$

$$\beta_{it}^d = \beta_0^d + \textstyle\sum_{j=1}^{r_d} \beta_j^d \, g(div_{t-1}^{sp}; \gamma_j, c_j) \tag{10}$$

$$\beta_{it}^c = \beta_0^c + \textstyle\sum_{j=1}^{r_c} \beta_j^c \, g(comp_{t-1}^{sp}; \gamma_j, c_j) \tag{11}$$

and estimate four different PSTR models:

A. nonlinear effects of localization:

$$a_{it}^{sp} = \beta_{it}^l loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it} \tag{12}$$

B. nonlinear effects of urbanization:

$$a_{it}^{sp} = \beta^l loc_{it}^{sp} + \beta_{it}^u urb_t^{sp} + \beta^d div_t^{sp} + \beta^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it} \tag{13}$$

C. nonlinear effects of diversity:

$$a_{it}^{sp} = \beta^l loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta_{it}^d div_t^{sp} + \beta^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it}. \tag{14}$$

[32]We use the lagged values of the indicators of interest (i.e., $loc_{i,t-1}$, $urb_{i,t-1}$, $div_{i,t-1}$, $comp_{i,t-1}$) as candidate transitions, instead of their actual values, to avoid problems of collinearity in the tests of nonlinearity.

D. nonlinear effects of competition:

$$a_{it}^{sp} = \beta^l loc_{it}^{sp} + \beta^u urb_t^{sp} + \beta^d div_t^{sp} + \beta_{it}^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it} \qquad (15)$$

As a robustness check, we also estimate a model where the nonlinearities not rejected by the tests appear simultaneously in the same equation.

E. nonlinear effects of localization, urbanization, diversity and competition:

$$a_{it}^{sp} = \beta_{it}^l loc_{it}^{sp} + \beta_{it}^u urb_t^{sp} + \beta_{it}^d div_t^{sp} + \beta_{it}^c comp_t^{sp} + \zeta_t^p + \phi_i + \xi_{it}. \qquad (16)$$

Before proceeding, it is worth discussing the reasons why we analyze only the cases in which the transition variable of the nonlinear regime is associated with the variable itself. Inspired by the theory illustrated in Sections 1 and 2, we aim at identifying the presence and size of possible nonlinear marginal effects of each of the agglomeration variables whereas we do not search for possible nonlinear interactions between the variables (as it would instead have been the case, had we interacted an agglomeration variable in the linear component with another variable in the nonlinear one). This is not different from and directly comparable with what done when using a linear specification with quadratic and cubic terms, as also in this case an agglomeration variable is interacted with itself (in the quadratic term) and with its own squared value (in the cubic term). We do not exclude that nonlinear interaction effects may be at play, yet their detection falls beyond the scope of this paper.

5. Results

Taking stock of the robust linear estimates reported in Table 2 and building on the preliminary findings on the augmented linear specifications (Table 3), we move to address the main object of this work, i.e., the existence of nonlinear effects of spatial agglomeration on TFP productivity at the firm-level.[33]

5.1. Nonlinearity tests

Following the procedure outlined in Section 4.1, we start by testing the null of linearity against the alternative of a two-regime PSTR model for each of the four models detailed in Section 4.4. In particular, for each model A-D we compute the three LM test statistics put forward by Andrews and Ploberger (1994), i.e. the average LM (AveLM), the exponential average LM (ExpLM), and the weighted average LM (wLM). All the test statistics are computed from heteroskedasticity-robust LM statistics and their p-values calculated via fixed-design wild bootstrap, as suggested by Hurn and Becker (2009).[34]

[33] All the calculations were done using gretl 1.9.14 (`gretl.sourceforge.net`). Code available at request.

[34] To calculate AveLM, ExpLM and wLM, we first compute a heteroskedasticity-robust LM test statistic for each of 100 pairs (γ_1, c_1): $LM(\gamma_1^{(j)}, c_1^{(j)})$. Each pair is built as follows: γ_1 is

Table 4: Heteroskedasticity-robust nonlinearity tests

	Transition variable	Test statistic		p-value
A	$loc^{sp}_{i,t-1}$	AveLM	4.3521	0.0000
		ExpLM	4.1418	0.0000
		wLM	0.1525	0.0000
B	urb^{sp}_{t-1}	AveLM	1.1071	0.3030
		ExpLM	1.1485	0.2121
		wLM	0.0561	0.2727
C	div^{sp}_{t-1}	AveLM	1.7953	0.1212
		ExpLM	2.7635	0.0303
		wLM	0.1409	0.0101
D	$comp^{sp}_{t-1}$	AveLM	6.1258	0.0000
		ExpLM	4.9033	0.0000
		wLM	0.1776	0.0000

The test statistics, with the associated p-values, are reported in Table 4. They strongly reject the null of linearity in all the models but model B, i.e. the one searching for possible nonlinearities in the externalities of urbanization. Accordingly, we shall proceed in what follows with the estimation of the models A, C and D. Before moving on, however, we would like to notice that the failure to reject the null of linearity for urbanization provides an important insight on its own: the insignificant impact of urbanization economies on the TFP in the linear estimate in Table 2 is not due to the failure to account for threshold effects and non-constant elasticity.

5.2. Nonlinear estimates

As the tests of nonlinearity suggest that adopting a PSTR model is warranted for all the indicators but urbanization, we proceed with estimating models A, C

drawn from a uniform distribution 0-100; c_1 is drawn uniformly at random from the set of observed values of the transition variables within the 5-95th percentile in the sample. Then we apply the following formulas:

$$\text{AveLM} = \frac{1}{100}\sum_{j=1}^{100} LM(\gamma_1^{(j)}, c_1^{(j)})$$

$$\text{ExpLM} = \ln\left(\sum_{j=1}^{100} \frac{\exp\left(0.5LM(\gamma_1^{(j)}, c_1^{(j)})\right)}{1000}\right)$$

$$\text{wLM} = \frac{1}{100}\sum_{j=1}^{100} \omega_j LM(\gamma_1^{(j)}, c_1^{(j)})$$

where $\omega_j = LM(\gamma_1^{(j)}, c_1^{(j)})/\sum_{j=1}^{100} LM(\gamma_1^{(j)}, c_1^{(j)})$. To calculate bootstrap p-values via fixed-design wild bootstrap, we compute AveLM, ExpLM and wLM for 99 bootstrap replications, where, in each replication, we randomize the sign of the residuals of the estimated linear model. The bootstrap p-value is equal to the fraction of bootstrap test statistics greater than the original one.

Table 5: Panel Smooth Transition Regression estimates

	A	C	D	E
β_0^l	0.0311	0.0542	0.0505	0.0287
	(0.0353)	(0.0436)	(0.0425)	(0.0349)
β_0^u	-0.0031	0.0207	0.0035	0.0226
	(0.0741)	(0.0750)	(0.0746)	(0.0749)
β_0^d	0.0158	0.0001	0.0180	0.0004
	(0.0390)	(0.0392)	(0.0389)	(0.0392)
β_0^c	0.4270***	0.4209***	0.4433***	0.4551***
	(0.0481)	(0.0520)	(0.0526)	(0.0494)
Region-time dummies	Yes	Yes	Yes	Yes
Wald test statistic (F-form) (H_0: no region-time dummies)	137.18	83.70	72.22	144.01
Wald test p-value	0.0000	0.0000	0.0000	0.0000
β_1^l	0.0641***			0.0646***
	(0.0212)			(0.0200)
β_1^d		0.0406**		0.0423**
		(0.0123)		(0.0123)
β_1^c			-0.0634***	-0.0637***
			(0.0194)	(0.0196)
γ_l	3.7830			4.0470
c_l	6.9101			6.9144
γ_d		737.14		905.66
c_d		3.6793		3.6787
γ_c			877.96	877.96
c_c			1.3904	1.3904
Observations	58,784	58,784	58,784	58,784
Years	8	8	8	8
Firms	7,348	7,348	7,348	7,348
Sector-province pairs	2,234	2,234	2,234	2,234

Dependent variable: a_{it}^{sp}. Years: 2000-2007. Region: NUTS 3 level. Province: NUTS 2 level. Cluster-robust standard errors in parentheses (cluster: sector-province pair). Significance at: 1% ***; 5% **; 10% *. NB: although we report the s.e. also for the slope parameters associated with the nonlinear regimes $(\beta_l^1, \beta_d^1, \beta_c^1)$, because of the nuisance parameter problem their significance is evaluated by using the tests reported in Table 4.

and D assuming two-regimes, i.e. $r_l = r_d = r_c = 1$ in Eqs. (8), (10) and (11). The results are summarized in Table 5.[35]

As detailed in Section 4.3, the estimation procedure stops only when the test of no remaining nonlinearity does not reject the null, for there is no evidence of additional nonlinear regimes not yet included in the specification. None of the tests reported in Table 6 reject the null at the 5% significance level and, therefore, the estimates of the two-regime PSTR models in Table 5 turn out to

[35] As outlined in Section 4.2, the minimum of the concentrated SSR to estimate the nonlinear parameters is find by using Corana et al.'s (1987) SA algorithm. In particular, in implementing the algorithm we set the initial temperature at 50, well above the average difference in SSR. The temperature reduction factor is 0.85. The algorithm adjusts the step-size vector every 20 parameter changes and the loop is repeated 20 times before each temperature reduction. The tolerance is set to 1e-05. In all the cases, the procedure converges after on average 350,000 function evaluations.

Table 6: Heteroskedasticity-robust tests of no remaining nonlinearity

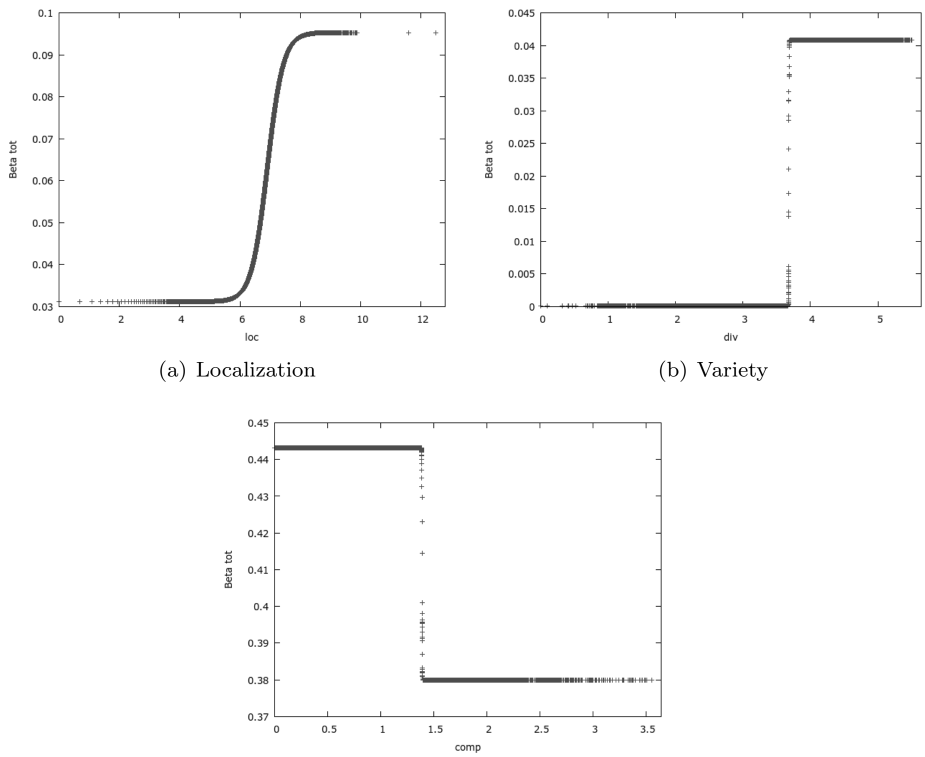
	Transition variable	Test statistic		p-value
A	$loc^{sp}_{i,t-1}$	AveLM	0.8875	0.5252
		ExpLM	0.6448	0.5354
		wLM	0.0449	0.5051
C	div^{sp}_{t-1}	AveLM	2.7119	0.0404
		ExpLM	1.8903	0.0505
		wLM	0.0811	0.0505
D	$comp^{sp}_{t-1}$	AveLM	0.9422	0.4747
		ExpLM	0.6246	0.5051
		wLM	0.04214	0.5354

be the final ones.[36]

In particular, the first column of Table 5 reports the estimates of model A, where the TFP elasticity to localization varies nonlinearly according to its lagged level ($loc^{sp}_{i,t-1}$). The slope parameter associated with the lower extreme regime (β^l_0) is positive but low (0.031) and not statistically different from zero, whereas the one associated with the upper extreme regime (β^l_1) is above 0.095 and highly significant. This points out that localization externalities do not materialize at low levels of spatial agglomeration, but only when the latter reaches a certain threshold. This is clearly showed by Figure 1(a), that plots the total elasticity of TFP to localization, β^l, against the lagged level of localization, $loc^{sp}_{i,t-1}$ (each firm at each moment in time is associated with one of the points along the plotted line). The graph makes apparent that, for levels of localization close to the sample mean (median) of $loc^{sp}_{i,t-1}$, i.e. 5.65 (6.17), which corresponds to 285 (475) workers employed outside the firm in the same sector and province, the estimated elasticity is about 0.032 (0.035); for values of *loc* equal to $\hat{c}_l$ (6.91), corresponding to roughly 1000 workers employed outside the firm in the same sector and province, localization economies materialize: the estimated elasticity is two times larger and statistically significant (0.063). Finally, when *loc* is equal to 8 (about 3000 workers) the elasticity is 0.094, almost three times larger than before.[37]

[36]As for diversity, one of the tests does not reject the null at the 5% level and the other two are very close to the boundary of the rejection region. We nonetheless decided not to include in the specification the second nonlinear regime for diversity for a number of reasons. First, since the tests of remaining nonlinearity are not robust to intra-cluster correlation and *div* is constant at the sector-province pair, these tests tend to over-reject the null. Second, González et al. (2005) suggest to reduce the significance level by a constant factor when testing for the presence of additional nonlinear regimes in order to avoid excessively large models. Third, and more importantly, once estimated: i) the location parameter associated with the additional regime happens to be quite close to the location parameter of the previous regime, with a sudden transition; ii) the sum of the point estimates of the coefficients β^d_1 and β^d_2 in the specification with the second nonlinear regime is very similar to that of β^d_1 in the specification with only one nonlinear regime. Therefore, this additional regime does not change any important conclusion while it makes harder illustrate and discuss the results.

[37]We warn the reader that these are just point estimates and should be taken with care, for they strongly depend on $\hat{\gamma}_l$ and $\hat{c}_l$.



(a) Localization

(b) Variety

(c) Competition

Figure 1: Total TFP elasticity as a function of the initial level of the variable

Looking at their time average, only 35% of the firms are above the location parameter and thus appear to receive consistent benefits from localization. 11.4% of the firms in the sample "change regime" in the period, crossing the threshold in either way in the period.[38] This finding clearly improves upon the estimated marginal impact of localization on TFP implied by the point estimates of the coefficients in the linear specification augmented by quadratic and cubic terms adopted by Martin et al. (2011) and in Table 3. Indeed, the estimated marginal impact from the linear augmented specification is negative for low-medium levels of localization and becomes positive only when localization is high, and continues to increase ever afterwards; more meaningfully, our findings indicate that the marginal impact is not statistically different from zero at low levels of agglomeration and gradually increases until it reaches a constant value, for further economies and diseconomies of localization offset each other.

[38] Let us note that in this case, since γ is not high, the change is smooth and therefore no clear regime switching can be identified.

Moreover, the fact that the tests of no remaining nonlinearity do not reject the null (Table 6) shows that there is no evidence of further regimes associated with localization. In particular, there is no strong evidence in the data in favor of regimes allegedly arising from the negative externalities of localization associated with agglomeration diseconomies (i.e., congestion effects, commuting costs, and the like). Clearly, this does not mean that these effects do not exist, but rather that the actual spatial distribution of firms is such that these negative effects do not materialize in practice.

Similar qualitative findings apply for the specification encompassing the nonlinear effect of diversity (div), namely Model C. According to the estimates (second column of Table 5), the TFP elasticity to variety (β^d) grows as the level of variety overcomes a certain threshold. The elasticity associated with the low extreme regime ($div < 3.68$) is not statistically different from zero, whereas the one associated with the high regime is again positive, larger and statistically significant. The transition across the regimes is less smooth than in the case of loc, because the smoothness parameter (γ_d) is relatively high: this implies that most of the firms in the sample fall in either of the two extreme regimes. Also in this case, we cannot reject the hypothesis of no remaining nonlinearity (Table 6) and the model is not extended to include further regimes. Figure 1(b) plots the total TFP elasticity to diversity (β^d) against its lagged level. These findings suggest that firms located in provinces characterized by a high variety of other industries benefit from this only once variety passes a critical level. Looking at their time average, in our sample 85% of the firms are above this threshold and roughly 19% crosses the threshold in either way at least one time. It is worth noticing that this result would not be detected by using a specification with quadratic and cubic terms, such as that adopted by Martin et al. (2011) and in Table 3. This strengthens the case of using a flexible approach such as the PSTR to detect properly nonlinear effects.

The case of competition (model D) is somehow symmetric to the previous ones. The TFP elasticity to competition (β^c), plotted in Figure 1(c), is reduced once competition exceeds a critical value. This finding is in line with the idea that competition is beneficial to innovation and productivity (as suggested by Porter, 1990), but its contribution decreases when competition becomes too fierce, probably because too high a level of local competition jeopardizes the ability of the firms to internalize the benefits of their innovation (Glaeser et al., 1992). The implied TFP elasticities in the two regimes are always positive and statistically significant. That said, it is worth stressing that, for values of $comp$ close to the location parameter, c_c, the regime switching produced by the increased (decreased) competition likely produces a decrease (an increase) of the overall TFP. Hence, competition and TFP might exhibit a negative correlation for values close to the threshold. In particular, since the average value of $comp$ (1.23) is smaller than but quite close to the estimated location parameter ($\hat{c}_c = 1.39$), this implies that, according to our estimates, economic policies aimed at increasing the average level of competition could have a negative impact on firm productivity in Italy, at least in the short-run.

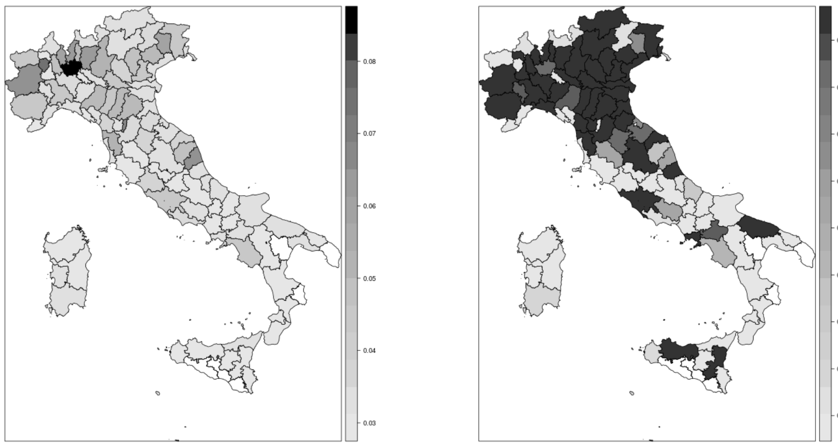
Table 5 reports also the result of model E, where all the nonlinear interaction terms are simultaneously considered. To estimate the model we started by introducing the nonlinear regime for *comp*, as this variable is associated with the highest statistics in the nonlinearity tests (Table 4). Then we tested for the presence of other nonlinear regimes associated with *loc*, *urb*, *div* or *comp*. On the basis of the results of these tests, we decided to include a nonlinear interaction associated with *loc*. We further test for any remaining nonlinearity associated with any of the agglomeration variables. On the basis of the test statistics we included the nonlinear regime associated with *div*. Then we stopped as the tests do not detect any additional nonlinearity. This specification shows that it makes practically no difference, both in terms of point estimates and of confidence intervals, whether all the nonlinearities are estimated simultaneously or individually as in model A, C and D. On the contrary, dealing with one nonlinearity at the time provides some evidence on the robustness of the results across the various specifications, since the estimated linear coefficients remain almost identical.

Having illustrated the results of the empirical analysis, we need to address the implications of such findings for the Italian provinces and sectors. In this respect, one advantage of the PSTR model with respect to the linear augmented one is that the former is less prone than the latter to suffer from the presence of outliers and to produce unrealistic in-sample projections for high and low values of the nonlinear variable. Hence, the PSTR model seems appropriate to produce a visual representation of the marginal impact of agglomeration economies.

In Figure 2, we map the marginal impact on TFP of both localization and diversity across the various Italian provinces, calculated by averaging firm values over years and sectors. The two maps in Figure 2 show that the marginal impact of localization and diversity is greater in the North and in a few provinces of the Center than in the South and in the islands, in line with the literature on the Italian industrial districts/local production systems. As the estimated threshold for the localization variable is relatively high in the sample and the regime switching quite definite, it is not surprising that there is little differentiation across the provinces in the South of Italy, with the exception of those in Campania, where one can find a concentration of food, clothing, textile, and leather manufacturing firms.

Figure 3 instead shows the total TFP impact of the four agglomeration economies (localization, variety, urbanization and local competition) across the Italian provinces. This map clearly suggests that agglomeration economies contribute to reinforce the historical North-South divide (measured in terms of GDP per capita, as well as productivity, efficiency and R&D spillovers, as shown among the others by Foddi and Usai, 2013; Dettori et al., 2012; Aiello and Cardamone, 2012).[39]

[39] The average total TFP impact of agglomeration economies at the province level is positively correlated with other indicators of productivity and agglomeration. In particular, it exhibits significant positive correlations with: i) the "district degree" of the province, computed from the 2001 "Italian Industrial census" of the Italian National Statistical Institute (ISTAT) as the number of workers employed in an industrial district as a fraction of the total manufacturing



(a) Localization (b) Variety

Figure 2: TFP elasticity to localization and diversity across Italian provinces

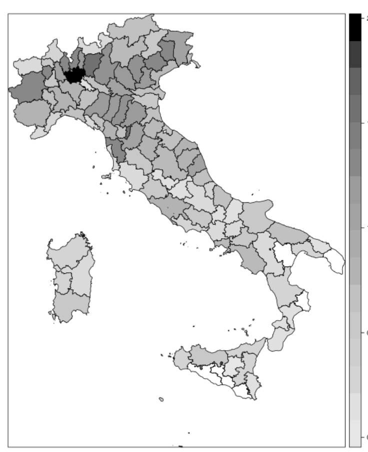


Figure 3: Total TFP accounted by agglomeration factors (localization, variety, urbanization and local competition) across Italian provinces

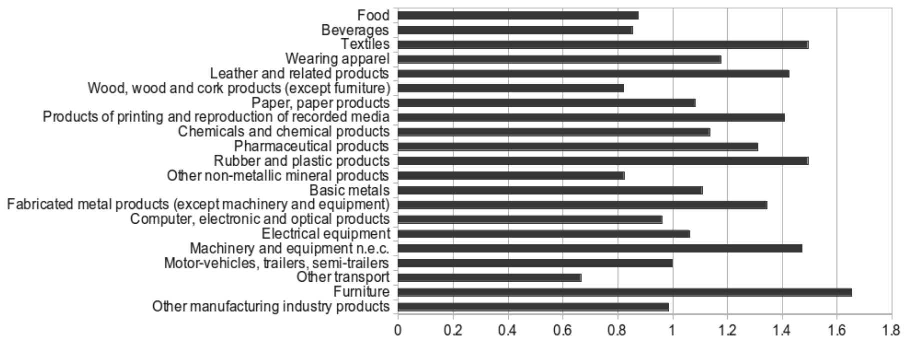


Figure 4: Total TFP accounted by agglomeration economies across industries (2-digit)

Finally, in Figure 4 we plot the total TFP impact of agglomeration economies (localization, diversity, urbanization and local competition) across industries, aggregated at the 2-digit level and averaged over time and provinces. Unsurprisingly, agglomeration economies have the highest impact in the industries representing the bulk of the Italian manufacturing sectors (textile, leather and wearing apparel, machinery, and furniture), as well as metal and pharmaceutical products.

6. Conclusions

This work revisits the vexed question regarding the impact of spatial agglomeration on productivity and, by extending the comprehensive approach by Martin et al. (2011) to analyze four kinds of agglomeration externalities, it addresses the overlooked issue of the plausible nonlinear impact of spatial economies on firm-level productivity.

By investigating the issue on a large sample of Italian manufacturing firms located in 97 Italian provinces (NUTS 3) over the period 1999-2007, it is shown that significant localization economies materialize only for values of intra-industry agglomeration above a certain threshold. While the significance of localization economies is in line with the linear estimates and with the findings in the literature, the nonlinear results suggest some interesting insights: localization economies are more influential when agglomeration is relatively large. Moreover, once a certain level of localization is reached, their marginal impact stops growing. There is however no sign of diseconomies (e.g., congestion effects) able to reduce the impact of agglomeration on the firm-level TFP, as the test statistics exclude any further nonlinearity. As we emphasize in the paper, this does not mean that

employment in the province (corr = 0.57); ii) the manufacturing export propensity in the province averaged over the period 1999-2007, computed from the yearly series of "Provincial Accounts" (Conti Provinciali) of ISTAT as the value share of exports in the total manufacturing production (corr = 0.63).

these negative effects do not exist or cannot prevail, but rather that the actual spatial distribution of firms is such that they do not materialize in our sample.[40]

Variety-related (Jacobian) externalities have also a positive impact on productivity, but this occurs only when diversity passes a certain threshold. Thus, neither a traditional linear analysis, nor a linear specification augmented by quadratic and cubic terms allow this result to emerge. To the best of our knowledge, there is no other study addressing the nonlinear impact of Jacobian externalities.

In line with previous studies, the impact of local competition on productivity is found to be positive, even though the nonlinear estimates show that when competition is above a certain threshold (close to the median value) its impact is reduced, although still largely positive and significant.

Finally, urbanization economies do not appear statistically significant and no nonlinearity is detected either.

We have explored the implications of the identified nonlinearities for the provinces and sectors in the Italian economy, by analyzing the average estimated marginal impacts of localization and variety across the Italian provinces, and the average estimated total impact on TFP of all agglomeration economies across regions and industries. Our findings are in line with the regional divide that characterizes Italy and with the sectoral specialization of its industrial system.

As pointed out by Greenstone et al. (2010), agglomeration externalities can have a dramatic practical importance and their features should inform economic policy. In this respect, our results actually show that, given the existence of critical masses and thresholds for localization economies and Jacobian externalities to materialize, tax and other localization incentives should manage to attract a non-negligible number of new firms to be successful. Moreover, given that the marginal spillovers for firm-level TFP flatten out after a certain level of agglomeration/diversity is reached, our results suggest the existence of an interesting trade-off at the policy level. On the one hand, the greatest increase in productivity at the firm level could be achieved by clustering policy actions on those geographical areas where the agglomeration/variety level can be pushed from just below to just above the critical threshold. This is particularly relevant for local policy makers. On the other hand, the aggregate effect on productivity depends on the number of firms which may benefit from such interventions. Accordingly, national policy makers should discriminate across industrial clusters by undertaking differentiated real and/or financial actions with a view to focusing on the areas where the marginal impact is higher.[41]

[40] Let us note that the fact that we do not observe a decreasing marginal net effect of agglomeration on TFP is not the result of optimal location choices of firms, as it might be expected in a longer time span where firms can revise their location choice, because no firm in our sample relocates.

[41] As explained, our sample is fully balanced due to the features of the estimation procedure. This implies that our results do not take into account an important phenomenon that affects aggregate productivity, that is firm demography. While this does not impinge upon the results, it should be taken into account in designing policy measures to boost productivity.

Table A.7: Panel unit root tests

	Harris-Tzavalis	Im-Pesaran-Shin ($Z_{\tilde{t}-bar}$)
a_{it}^{sp}	-0.1250 (0.0000)	-69.9664 (0.0000)
loc_{it}^{sp}	0.2876 (0.0000)	-610.000 (0.0000)
urb_{t}^{sp}	0.1243 (0.0000)	-25.4291 (0.0000)
div_{t}^{sp}	-0.1001 (0.0000)	-99.2528 (0.0000)
$comp_{t}^{sp}$	-0.1044 (0.0000)	-37.6448 (0.0000)

Cross-sectional averages subtracted from the series to mitigate the impact of cross-sectional dependence, as suggested by Levin et al. (2002). H_0: panels contain unit roots. H_a: Panels are stationary (HT); Some panels are stationary (IPS). p-values in parentheses.

This analysis could be extended in other directions. First, one could explore nonlinear interactions between different forms of agglomeration. It is possible, for instance, that urbanization economies become relevant once a certain level of localization is achieved; or that the impact of local competition externalities may depend on the overall size of the industry to which a firm belongs. Second, one could properly address the possible presence of time-variant location-specific unobserved factors.[42] Third, the analysis might be extended by considering the service sectors, in addition to the manufacturing ones. This is indeed one of the main limits of the paper as the share of services is steadily growing over time and some services are becoming increasingly knowledge intensive and, in general, highly integrated with manufacturing. These issues represent venues of future research.

Acknowledgements

We are grateful to the participants to the AISRe Conference in Rome (13-15 September 2012) and, in particular, Giuseppe Espa and Roberto Basile for useful suggestions. We also acknowledge Allin Cottrell and Riccardo (Jack) Lucchetti for having developed gretl and for all the tips and tricks they gave us. Usual disclaimers apply.

AppendixA. Granger tests

To get some hints on possible reverse causality/simultaneity issues, we have carried out a Granger test for each explanatory variable. In particular, after having checked the absence of unit roots (Table A.7 reports Harris and Tzavalis (1999) and Im et al. (2003) tests for panel-data unit-root; both the tests lead to

[42] One possibility is to use a spatial lag instead of or along with the temporal lag, as transition variable. For example, the transition variable q_{it} might be defined as $\sum_j w_{ij} q_{j,t-\tau}$, where w_{ij} is a spatial weight reflecting the geographical proximity between firm i and firm j and $\tau \geq 0$ is the temporal lag. We are indebted to one of the referees for this suggestion, which we do not follow as proceeding in this direction would make the paper strive far from its original goal and its main term of comparison, i.e. Martin et al. (2011).

Table A.8: Dynamic panel-data estimation, two-step difference GMM

	Dependent variable (x_t)			
	loc_{it}^{sp}	urb_t^{sp}	div_t^{sp}	$comp_t^{sp}$
$a_{i,t}^{sp}$	-1.4117*	-1.3181	0.8777	-1.1027
	(0.6971)	(1.1905)	(0.5992)	(1.1560)
$a_{i,t-1}^{sp}$	0.0007	0.0010	-0.0023	0.0058
	(0.0118)	(0.0110)	(0.0074)	(0.0097)
x_{t-1}	0.8228***	0.0353	-0.0021	0.0454
	(0.2662)	(0.2815)	(0.0694)	(0.1110)
x_{t-2}	0.2931***	0.1416	0.2712***	0.0752
	(0.1049)	(0.1193)	(0.0328)	(0.0584)
Observations	36,740	36,740	36,740	36,740
Years	5	5	5	5
Firms	7,348	7,348	7,348	7,348
Sector-province pairs	2,234	2,234	2,234	2,234
N. instruments	11	11	11	11
AR(1) [p-value]	-2.04 [0.041]	-1.10 [0.270]	-1.47 [0.142]	-1.01 [0.313]
AR(2) [p-value]	-0.87 [0.386]	0.06 [0.955]	0.09 [0.926]	-0.35 [0.725]
Sargan test [p-value]	0.92 [0.631]	1.54 [0.464]	0.33 [0.847]	1.41 [0.494]
Hansen test [p-value]	0.90 [0.639]	1.56 [0.459]	0.33 [0.848]	1.40 [0.497]
Wald test statistic [p-value] H$_0$: $a_{i,t}^{sp}$ and $a_{i,t-1}^{sp}$ jointly not significant	4.13 [0.127]	1.25 [0.537]	2.15 [0.341]	1.04 [0.596]

Unreported constant and time dummies. HAC s.e. in round brackets. Significance at: *** 1%; ** 5%; * 10%.

reject the null of panels containing unit roots at the 1% level for all the series),[43] we have estimated the following dynamic panel models by two-step difference GMM (Bond, 2002):

$$x_{it} = \sum_{m=1}^{2} \alpha_m x_{i,t-m} + \sum_{l=0}^{1} \gamma_l a_{i,t-l}^{sp} + \zeta_t + \phi_i + \xi_{it}$$

where x_{it} is equal, in turn, to $loc_{it}^{sp}, urb_t^{sp}, div_t^{sp}, comp_t^{sp}$, and we have computed Wald statistics to test the null $\gamma_0 = \gamma_1 = 0$. Table A.8 summarizes the main results. As shown in the Table, the Wald tests always fail to reject the null hypothesis at the 10% level, i.e. that the TFP does not Granger-cause the explanatory variable ($loc_{it}^{sp}, urb_t^{sp}, div_t^{sp}, comp_t^{sp}$).

References

Aiello, F., Cardamone, P., 2012. Regional economic divide and the role of technological spillovers in Italy. evidence from microdata. Structural Change and Economic Dynamics 23 (3), 205–220.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–56.

[43]Both the test statistics have well-defined asymptotic distributions with a fixed T and $N \to \infty$.

Andrews, D. W. K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62 (6), 1383–1414.

Antonietti, R., Cainelli, G., Lupi, C., 2013. Vertical disintegration and spatial colocalization: The case of KIBS in the metropolitan region of Milan. Economics Letters 118 (2), 360–363.

Arrow, K. J., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R. (Ed.), The Rate and Direction of Inventive Activity. Princeton University Press, Princeton, NJ, pp. 609–626.

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

Becker, R., Osborn, D. R., 2012. Weighted smooth transition regressions. Journal of Applied Econometrics 27 (5), 795–811.

Bond, S., 2002. Dynamic panel data models: a guide to micro data methods and practice. Portuguese Economic Journal 1 (2), 141–162.

Brunello, G., De Paola, M., 2008. Training and density. some evidence from Italian provinces. Labour Economics 15 (1), 118–140.

Cainelli, G., Iacobucci, D., 2012. Agglomeration, related variety, and vertical integration. Economic Geography 88 (3), 255–277.

Ciccone, A., 2002. Agglomeration effects in europe. European Economic Review 46 (2), 213–227.

Ciccone, A., Hall, R. E., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Cingano, F., Schivardi, F., 06 2004. Identifying the sources of local productivity growth. Journal of the European Economic Association 2 (4), 720–742.

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., Gobillon, L., 2011. The identification of agglomeration economies. Journal of Economic Geography 11 (2), 253–266.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica 80 (6), 2543–2594.

Combes, P. P., Duranton, G., Gobillon, L., Roux, S., 2010. Estimating Agglomeration Economies with History, Geology, and Worker Effects. NBER books. University of Chicago Press, Chicago, pp. 15–66.

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

Corana, A., Marchesi, M., Martini, C., Ridella, S., 1987. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. ACM Trans. Math. Softw. 13, 262–280.

De Groot, H., Poot, J., Smit, M., 2009. Agglomeration externalities, innovation and regional growth: Theoretical perspectives and meta-analysis. In: Capello, R., Nijkamp, P. (Eds.), Handbook of Regional Growth and Development Theories. Edward Elgar, Cheltenham, pp. 256–281.

Dettori, B., Marrocu, E., Paci, R., 2012. Total factor productivity, intangible assets and spatial dependence in the European regions. Regional Studies: The Journal of the Regional Studies Association 46 (10), 1401–1416.

Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4 of Handbooks in Economics. Elsevier, Ch. 48, pp. 2063–2117.

Ellison, G., Glaeser, E. L., 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review 89 (2), 311–316.

Foddi, M., Usai, S., 2013. Regional knowledge performance in Europe. Growth and Change 44 (2), 258–286.

Fok, D., van Dijk, D., Franses, P. H., 2005. A multi-level panel STAR model for US manufacturing sectors. Journal of Applied Econometrics 20 (6), 811–827.

Frenken, K., Van Oort, F., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. Regional Studies 41 (5), 685–697.

Fujita, M., Krugman, P., Venables, A., 1999. The Spatial Economy. Cities, Regions and International Trade. The MIT Press, Cambridge.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., Shleifer, A., 1992. Growth in cities. Journal of Political Economy 100 (6), 1126–1152.

Goffe, W., Ferrier, G., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60 (1-2), 65–99.

Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.

González, A., Teräsvirta, T., 2006. Simulation-based finite sample linearity test against smooth transition models. Oxford Bulletin of Economics and Statistics 68, 797–812.

González, A., Teräsvirta, T., van Dijk, D., 2005. Panel smooth transition regression models. Research Paper 165, Quantitative Finance Research Centre.

Granger, C., Teräsvirta, T., 1993. Modelling nonlinear economic relationships. Oxford University Press, New York.

Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. Journal of Political Economy 118 (3), 536–598.

Hansen, B. E., 1999a. Testing for linearity. Journal of Economic Surveys 13 (5), 551–76.

Hansen, B. E., 1999b. Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics 93 (2), 345–368.

Hansen, B. E., 2000. Sample splitting and threshold estimation. Econometrica 68 (3), 575–603.

Harris, R. D. F., Tzavalis, E., 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. Journal of Econometrics 91, 201–226.

Helsley, R. W., Strange, W. C., 1990. Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics 20 (2), 189–212.

Henderson, J. V., 1974. The sizes and types of cities. American Economic Review 64, 640–656.

Henderson, J. V., 1988. Urban development: Theory, Fact, and Illusion. Oxford University Press, Oxford.

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

Henderson, V., 2003b. Marshall's scale economies. Journal of Urban Economics 53 (1), 1–28.

Henderson, V., Kuncoro, A., Turner, M., 1995. Industrial development in cities. Journal of Political Economy 103 (5), 1067–1090.

Higano, Y., Shibusawa, H., 1999. Agglomeration diseconomies of traffic congestion and agglomeration economies of interaction in the information-oriented city. Journal of Regional Science 39 (1), 21–49.

Hurn, S., Becker, R., 2009. Testing for nonlinearity in mean in the presence of heteroskedasticity. Economic Analysis and Policy 39 (2), 311–326.

Im, K. S., Pesaran, M. H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53–74.

Jacobs, J., 1969. The Economy of Cities. Vintage, New York, NY.

LeSage, J., Pace, R. K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC.

Levin, A., Lin, C. F., Chu, C. S. J., 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics 108, 1–24.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70 (2), 317–341.

Lundbergh, S., Teräsvirta, T., 1998. Modelling high-frequency time series with STAR-GARCH models. SSE/EFI working paper 291, Stockholm School of Economics.

Luukkonen, R., Saikkonen, P., Terasvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.

Marrocu, E., Paci, R., Usai, S., 2013. Productivity growth in the Old and New Europe: The role of agglomeration externalities. Journal of Regional Science 53 (3), 418–442.

Marshall, A., 1920. Principles of Economics, 8th Edition. Macmillan, London.

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

Melo, P. C., Graham, D. J., Noland, R. B., 2009. A meta-analysis of estimates of urban agglomeration economies. Regional Science and Urban Economics 39 (3), 332–342.

Mills, E. S., 1967. An aggregative model of resource allocation in a metropolitan area. American Economic Review 57, 197–210.

Montresor, S., Vittucci Marzetti, G., 2011. The deindustrialisation/tertiarisation hypothesis reconsidered: a subsystem application to the OECD7. Cambridge Journal of Economics 35 (2), 401–421.

Moulton, B. R., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. The Review of Economics and Statistics 72 (2), 334–38.

Otten, R. H. J. M., van Ginneken, L. P. P. P., 1989. The Annealing Algorithm. Kluwer, Boston, MA.

Paci, R., Usai, S., 2008. Agglomeration economies, spatial dependence and local industry growth. Revue économie industrielle 123, 87–109.

Porter, M., 1990. The Competitive Advantage of Nations. Free Press, New York.

Puga, D., 2010. The magnitude and causes of agglomeration economies. Journal of Regional Science 50 (1), 203–219.

Romer, P., 1986. Increasing returns and long run growth. Journal of Political Economy 84, 1002–1037.

Rosenthal, S. S., Strange, W. C., 2001. The determinants of agglomeration. Journal of Urban Economics 50 (2), 191–229.

Rosenthal, S. S., Strange, W. C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4 of Handbooks in Economics. Elsevier, Ch. 49, pp. 2119–2171.

Tabuchi, T., 1998. Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. Journal of Urban Economics 44 (3), 333–351.

van Oort, F. G., 2007. Spatial and sectoral composition effects of agglomeration economies in the Netherlands. Papers in Regional Science 86 (1), 5–30.

van Oort, F. G., Burger, M. J., Knoben, J., Raspe, O., 2012. Multilevel approaches and the firm-agglomeration ambiguity in economic growth studies. Journal of Economic Surveys 26 (3), 468–491.

Zheng, X.-P., 2001. Determinants of agglomeration economies and diseconomies: Empirical evidence from Tokyo. Socio-Economic Planning Sciences 35, 131–144.