Agglomeration and the Italian North-South divide

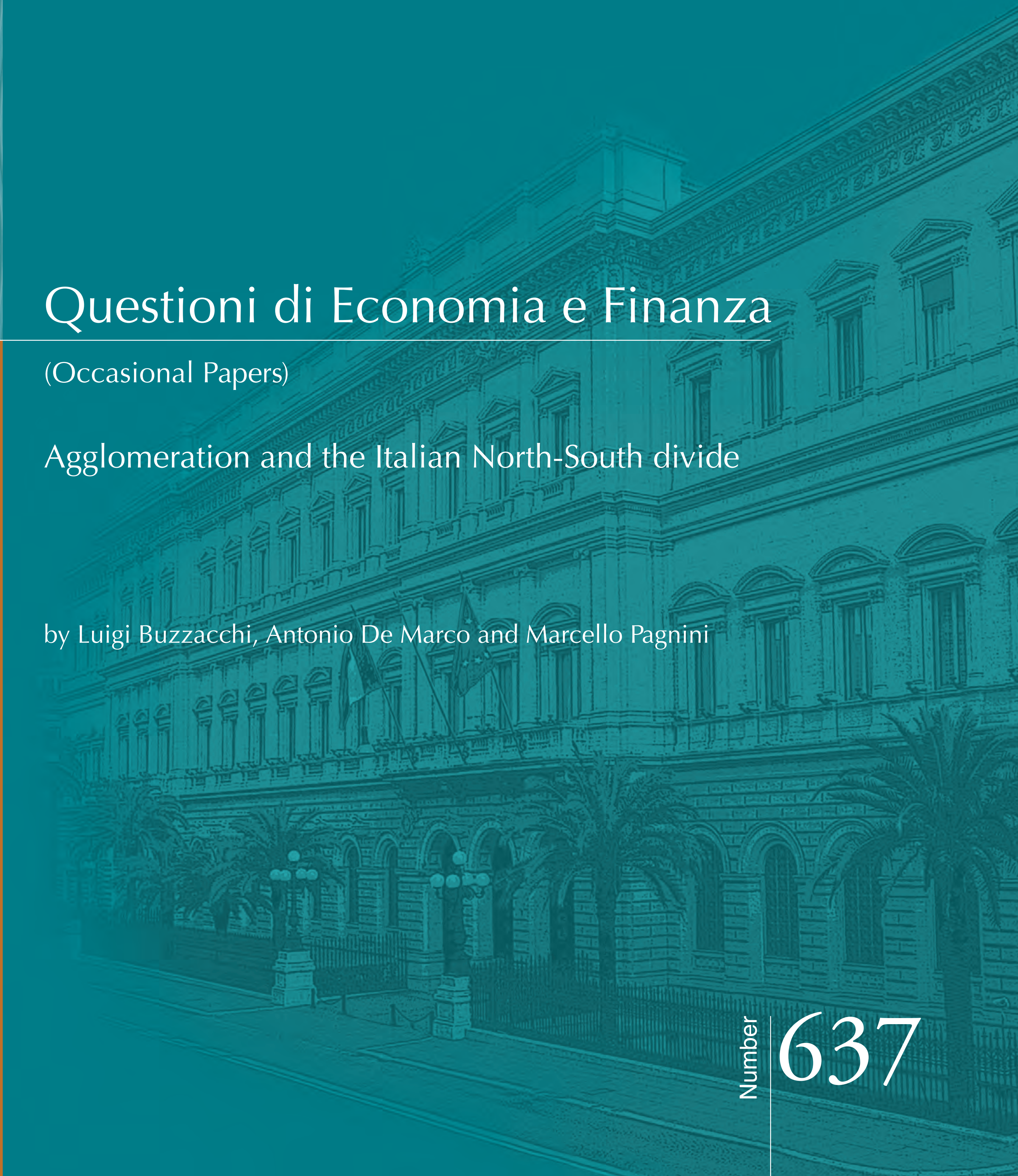
BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Agglomeration and the Italian North-South divide

by Luigi Buzzacchi, Antonio De Marco and Marcello Pagnini



October 2021

Number 637

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Agglomeration and the Italian North-South divide

by Luigi Buzzacchi, Antonio De Marco and Marcello Pagnini

Number 637 – October 2021

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

AGGLOMERATION AND THE ITALIAN NORTH-SOUTH DIVIDE

by Luigi Buzzacchi[a,b], Antonio De Marco[a,b] and Marcello Pagnini[c]

Abstract

This paper offers novel evidence on agglomeration economies by examining the link between total factor productivity (TFP) and employment density in Italy. TFP is estimated for a large sample of manufacturing firms and then aggregated at the level of Local Labor Market Areas (LLMAs). We tackle the endogeneity issues stemming from the presence of omitted covariates and reverse causation with an instrumental variable (IV) approach that relies on historical and geological data. Our estimate of the TFP elasticity with respect to the spatial concentration of economic activities is about 6%, a magnitude comparable to that measured for other developed countries. We find that the TFP-density nexus contributes to explaining a large share of the substantial productivity gap between the northern and southern regions of Italy. We also show that no significant heterogeneity emerges in the intensity of agglomeration economies across the country and that the positive TFP difference in favor of the firms located in the North is not due to the tougher competition taking place in those areas.

JEL codes: R12, R23.
Keywords: agglomeration economies, density, total factor productivity, regional disparities, selection effects.
DOI: 10.32057/0.QEF.2021/0637

Contents

[a] Politecnico di Torino, Interuniversity Department of Regional and Urban Studies and Planning
[b] Politecnico di Torino, Future Urban Legacy Lab
[c] Banca d'Italia, Local Economic Research and Analysis Division, Bologna Branch

1. Introduction

The concentration of workers, firms, or institutions in specific areas might generate productivity advantages for the firms located within those borders. An extensive theoretical and empirical literature has investigated this nexus, showing, for instance, that Total Factor Productivity (TFP) of firms increases with the density of economic activities in the local markets. In this paper, we contribute to the existing literature in two ways. First, we provide a measure of the elasticity of TFP with respect to an indicator of local economic density for the Italian private sector. Although previous research has dealt with similar topics, ours is the first study to provide such estimate to the best of our knowledge. Notably, this makes it possible to compare the Italian case with related exercises carried out for other countries.

Armed with this evidence, we investigate whether agglomeration economies can contribute to the explanation of the traditional productivity gap between the firms located in the northern and southern regions of Italy. In other words are *ceteris paribus* firms located in Southern Italy less productive than those located in Northern Italy because local markets in the South display a lower density of economic activity? Could the positive North-South productivity gap stem from a lower TFP elasticity with respect to such density in the South as compared with the North? As regards the second question, despite the fact that several works have mentioned agglomeration as one of the potential explanations for the inefficiency of southern firms in Italy, we are not aware of any work directly addressing a similar topic in the way we do in this paper. Moreover, since variations in employment density do not fully explain the observed heterogeneity in TFP, we implement a non-parametric methodology to discriminate between classes of determinants (other than agglomeration effects) that are believed to reasonably affect the distribution of firm efficiency. Such approach allows us to test whether part of the North-South disparities in terms of log-productivity has been determined by sorting or localized selection mechanisms, and not by different levels of density.

To address these issues, we measure TFP at the firm level by resorting to a rich dataset that includes a large sample of Italian manufacturing corporations observed for the years between 1995 and 2015. Such TFP data has been aggregated for the Local Labor Market Areas (LLMAs) as defined by ISTAT in 2011. Subsequently, we carry out an estimation of agglomeration economies at this level of aggregation by regressing (the logarithm of) the TFP for each LLMA on (the logarithm of) the number of employees per square kilometer. Within this framework, we also analyze the questions mentioned above concerning the role of spatial concentration in explaining the North-South productivity gap. We are fully aware that the estimation of the TFP elasticity is plagued with endogeneity problems,

The views expressed in this paper are those of the authors and do not necessarily represent the positions of the Bank of Italy. We thank Antonio Accetturo, Giuseppe Albanese, Filippo Scoccianti, Luigi Federico Signorini, Roberto Torrini, Elena Gentili, and Francesco Biancalani for their valuable comments that greatly improved our paper. Constructive suggestions and useful critiques were also provided by the participants in the workshop on the Progetto Mezzogiorno, we are grateful to all of them.

which are addressed by resorting to an instrumental variable (IV) regression. We use a rich set of instruments encompassing historical and geological variables. Finally, we discuss some subtle identification questions that might affect our estimations and that are seldom addressed in other contributions.

Our main results are that *i)* the estimate of the TFP elasticity with respect to the spatial concentration of economic activities is about 6%, a magnitude comparable to that computed for other developed countries by scholars using methodologies similar to the ones employed in our work; *ii)* the TFP-density nexus contributes to explaining a large share of the substantial productivity gap between the northern and southern regions of Italy; *iii)* no significant heterogeneity is detected in the intensity of agglomeration economies between the northern and southern regions; in other words, we do not find evidence that the returns to density are lower in the South; and *iv)* the positive TFP difference in favor of the firms located in the North is not due to the tougher competition effects taking place in those areas.

The remainder of the paper is organized as follows. Section 2 gives a brief review of the existing literature dealing with theoretical and empirical issues of agglomeration economies. Section 3 is a short preview of the main results. Additional details on the measurement of the average local productivity and the density of economic activities are provided in Section 4. Section 5 outlines the econometric strategy as well as the choice of instruments and controls. A discussion on model specification, findings, and some robustness tests is presented in Section 6. Section 7 examines the contribution of agglomeration economies to the differences in productivity between the northern and southern regions of Italy. The paper ends with some concluding remarks in Section 8.

2. Literature review on agglomeration economies

Population and economic activities are not evenly distributed in space. The most evident reasons for explaining such evidence since the early stages of economic development are the physical endowment and the morphology of territories, the so-called *first-nature* characteristics: climate, raw resources, and accessibility. Though agglomeration (i.e., density spikes) can be a by-product of multitudes of location choices aimed at capturing the benefits of first-nature factors, natural advantages account for just a fraction of the observed spatial differences in levels of density[1]. Various strands of the literature argue that agglomeration (i.e., proximity among firms, workers, and people) allows economic agents to "*economize on local trade costs, spread information and ideas more easily, diversify the range of products produced, and access larger pools of workers and jobs*" (Duranton and

[1] Ellison and Glaeser (1999) attribute roughly one-fifth of the observed industry spatial concentration to a very small set of natural advantages. Henderson et al. (2018) show that the effects of specific first-nature characteristics on the density of economic activities explain about half of the worldwide variation of density and more than one-third of within-country variations. For the Italian case, Accetturo and Mocetti (2019) analyze the role of geography and history for explaining the distribution of the population in space and its evolution over time.

Puga, 2004, p. 2065). These benefits are available, to a large extent, independently from the geographic features of the territories where they are generated.

The share of agglomeration that cannot be explained by the exogenous space heterogeneity is the focus of two different streams of research: the *urban economics* (UE), and the *new economic geography* (NEG) approaches (Combes et al., 2005). In both literatures, possible mechanisms for endogenous emergence of agglomeration are modeled.

In the tradition of UE, starting from the seminal work of Henderson (1974), the observed agglomeration at the equilibrium is associated with the advantages that density brings forth directly, determined by pure positive externalities. The Marshallian idea that denser local markets produce positive externalities and make incumbent firms more efficient can be derived from several models[2]. Duranton and Puga (2004) proposed the now-standard classification of agglomeration economies consisting of the triad of *matching*, *sharing*, and *learning* mechanisms[3]. In this literature, consequently, agglomeration (and density) is just a channel through which economic activities generate and catch *localized* externalities, the true sources of economic advantage. In that sense, agglomeration is then an *intermediate* determinant of productivity, and its effectiveness could variate with the intensity of the available externalities.

In NEG models (e.g., Fujita et al., 1999), increasing returns at the firm level, imperfect competition, and trade costs might drive to concentration at the equilibrium. In agglomerated areas, some pecuniary advantage (e.g., higher wages, land values, and rents) can emerge, but the agglomeration, *per se*, does not grant any productivity advantage.

The empirical economic research aimed at measuring the advantages of agglomeration is flourishing. This literature attempts to detect increasing returns in a local (i.e., relative to a well-defined geographic area, but external to the boundaries of single firms) production function, where the density of firms, workers, or individuals is (sort of) an input. The first wave of empirical studies on agglomeration economies is surveyed in Rosenthal and Strange (2004). A major challenge in the more recent literature (starting from Ciccone and Hall, 1996) is to sort out the direct causal relationship from agglomeration onto the productivity of input factors, from the relations where agglomeration is the effect (and not the cause), or the by-product, of productivity.

Three main classes of mechanisms might determine an emerging spurious correlation between agglomeration and economic advantages (i.e., productivity)[4]. *First-nature* advan-

[2] Agglomeration externalities arise because of the indivisibility in the provision of certain goods or facilities, the specialization of labor forces and the production, diffusion (thanks to face-to-face communications), and accumulation of ideas.

[3] Of course, the dark side of agglomeration is the emergence of negative externalities, i.e., congestion effects, that explain the observed upper bounds for density. As Duranton and Puga (2004, p. 2065) put it, "*we can then regard cities as the outcome of a trade-off between agglomeration economies or localized aggregate increasing returns and the costs of urban congestion*".

[4] While pure agglomeration economies can arise even in a homogeneous space and among homogeneous individuals, the three mechanisms described above need some form of heterogeneity (non-homogeneity of

tages turn into better local outcomes that attract firms and workers in specific locations, affecting their local performances. Starting from the assumption that firms are *ex-ante* heterogeneous in terms of productivity, a positive relationship between productivity and density will also be observed if agglomerated markets develop stronger selection effects. Recent NEG models (starting from Melitz and Ottaviano, 2008), assume that tougher competition is associated with the dimension of the market, so that denser areas show higher levels of productivity. Sorting mechanisms, rooted in the idea that firms and workers that are intrinsically more productive may prefer agglomerated areas, either because they benefit more from agglomeration effects or because agglomerated areas turn out to have better institutions, higher amounts of amenities, *et cetera*. In this sense, more productive individuals are over-represented in denser areas, even if agglomeration is not a determinant of productivity (Combes et al., 2012; De La Roca and Puga, 2017; Gaubert, 2018). The methodological solutions for detecting true agglomeration economies into the emerging correlations between density and productivity, netting from spurious effects, will be discussed in Section 5.

The empirical evidence on *proper* agglomeration economies is fairly established, and several survey papers (in particular, Rosenthal and Strange, 2004; Melo et al., 2009; Combes and Gobillon, 2015; Ahlfeldt and Pietrostefani, 2019) already illustrate the methods and the motivations for detecting and measuring the underlying phenomena.

A first element common to all this research is that it often does not discriminate between the different channels behind agglomeration economies so that the mechanism is usually a black box whose content is only conceptually known. Secondly, the empirical evidence proves to be strongly dependent on the industry, time, country, and the spatial structure assumed. Lastly, whatever the empirical strategy, the evidence provided in the literature is always an assessment of the *net* agglomeration effects, what scholars can observe is the part of positive effects that are not offset by the negative ones (i.e., congestion).

In general, the available empirical evidence confirms that the elasticity of productivity with respect to the spatial concentration of economic activities is significantly positive, even if estimates fluctuate greatly in magnitude. Moreover, as witnessed by various surveys and meta-analyses, estimation strategies are rather differentiated, and, in particular, they usually do not simultaneously control for endogeneity, selection, and sorting, resulting in somehow generally upward biased estimates. All that said, the usual ranges for the elasticity are in the range of 2% to 9% (Rosenthal and Strange, 2004; Melo et al., 2009; Combes, 2011; Ahlfeldt and Pietrostefani, 2019).

As for the way the effects of agglomeration economies are measured, factor productivity, wages, and sometimes employment are usually considered. These variables can then be obtained from regional (or urban) aggregate data or individual firm data. Combes and Gobillon (2015) maintain that (p. 302) "*it is worth studying the effects*" of agglomeration on TFP rather than on wages "*since it is a direct measure of productivity*", and its use (p.

the space; non-homogeneity of firms, workers, or both in the following two cases).

283) "*avoids making any assumption about the relationship between the local monopsony power*" of firms on labor "*and agglomeration economies*". Moreover, individual firm data are needed when dealing with selection issues (see also the discussion in Section 7.2).

As for the available evidence in this type of studies, empirical results focused on city size as determinants of productivity date back to the Seventies (Sveikauskas, 1975; Segal, 1976). However, a seminal paper that first estimates increasing returns to density, taking into consideration endogeneity issues, is Ciccone and Hall (1996). They explain differences in labor productivity across the US states with elasticity to density of about 6%. Henderson (2003) in the US, Cingano and Schivardi (2004) in Italy, as well as Graham (2009) in the UK are the first to introduce a measure of TFP based on individual firm data. The subsequent research in this stream of literature offers more sophisticated estimates of TFP and try to better address possible endogeneity biases. In particular, Combes et al. (2010), regress wages and TFP (estimated with the method proposed by Olley-Pakes at the firm level, and then aggregated at the local scale) on density for the French case, controlling for reverse causation and workers sorting. They report a proper density elasticity of wages at 2% and around 3.5% for the density elasticity of TFP.

3. A preview of the main findings

The issue of the North-South disparities in Italy has been explored under several perspectives. In this section of the paper, we offer a short preview of our results on the topic using the lens of the economic geography and, in particular, of those of the literature on the agglomeration economies. A more in-depth analysis, as well as the motivations lying behind this empirical evidence, will be presented in the following sections. Here we want to summarize some findings to help the reader to have an easy grip on them beyond the long technicalities that will be addressed later on.

As for the spatial scale of our analysis, we partition Italy into the 611 local labor market areas (LLMA) defined by the Italian National Institute of Statistics (ISTAT) for 2011. Starting from the 8,092 Italian municipalities, such territorial units are built by aggregating municipalities based on their spatial contiguity, and the self-containment of daily commuting flows for work reasons. These spatial cells represent an ideal reference for the analysis of agglomeration economies since many of the externalities mentioned by the theory occur at the level of a local labor market that exactly matches the LLMA definition. Notice that this partition is produced every ten years, as the data needed in this respect comes from the census of population and economic activities carried out by ISTAT at the beginning of each decade. Although this mapping had been evolving somewhat (the number of LLMAs had been always decreasing from 1981 onwards), it also exhibits a certain degree of stability, justifying its use for a reference year in a structural analysis of the North-South disparities.

Figure 1a offers a representation of such a territorial reference grid in Italy and depicts the

LLMAs belonging to the North and South according to our definition[5]. The South hosts 281 LLMAs, whereas 330 units are located in the North, respectively 46.0% and 54.0%. In terms of land covered, the South represents 40.9% of the entire national territory[6]. If the surfaces of LMMAs were represented as circles, those located in the southern regions would have, on average, a ray of 11.0 kilometers as compared to 12.4 kilometers for the northern ones. Differences become even more pronounced when we look at other variables. In 2001, the share of population living in the southern LLMAs amounted to 36.0%, the number of employees in all the sectors, including both services and the building industry, to 21.7%, and those in the manufacturing sectors to 16.4% of the total. Summing up these pieces of evidence, it turns out that southern LLMAs are more fragmented and smaller than those in the North. The reasons behind these features could be attributed to first nature disadvantages as well as to other factors, like the lower availability and efficiency of transport infrastructures in the South.

Coming to the gaps in terms of productivity and density of the economic activities, we computed an indicator of TFP in the manufacturing sector for each LLMA that is averaged across all the years between 1995 and 2015 and netted out for the effects of sectoral composition at the local level within the manufacturing sector. Since we use firm-level yearly data, this averaging is needed not to emphasize short-term variations in productivity and instead concentrate on the long-run effects of agglomeration economies. As for the density, we have computed an indicator based on the number of employees working in all the sectors of each LMMA per square kilometer.

The LLMAs in the South display a lower TFP, on average, by 26.7% compared to those in the North; medians indicate a difference of 28.7% (see Table 1). Similar gaps are confirmed across the other percentiles of the productivity distribution[7]. As for the density, LLMAs located in northern and southern regions host, on average, 62.0 and 33.6 (with medians 36.3 and 14.1) employees per square kilometer respectively. Those differences are spatially represented in the two Figures 2a and 2b. Apart from confirming the existence of a North-South gradient in the spatial distributions, the two choropleth maps also display a high degree of overlap for the above-mentioned spatial patterns. In other words, the less productive LLMAs in the South also show a lower density than those in the North.

Finally, we come to the central question investigated in this paper, i.e., the link between productivity and density. Figure 1b is a simple correlation plot between (the logarithm of) our TFP indicator and (the logarithm of) density at the level of LLMAs. The graph points to a positive relationship between the concentration of economic activity and productivity that may be consistent with the existence of the agglomeration economies. The southern LLMAs are mostly concentrated in the third quadrant of the plane, i.e.,

[5] In our definition the North macroarea includes the regions of Piemonte, Valle d'Aosta, Lombardia, Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Liguria, Emilia Romagna, Toscana, Umbria, Marche, and Lazio. The remaining regions, including the Islands, are grouped into the South macroregion.

[6] Note that the largest flat area, the Po valley, is located in the northern Italian regions.

[7] We use the `qreg` command in Stata.

that of the units exhibiting both a low level of productivity and a low density[8]. Such evidence would suggest that southern regions are less productive in part because they feature a low number of workers per square kilometer. The disproportionate presence of blue-colored observations below the regression line in the scatter plot also highlights significant localization effects that should be investigated further. According to the tenets of the urban economic literature surveyed in Section 2, the fact that employees are sparser in the South would weaken the formation of those positive externalities that are associated with agglomerated areas.

As shown in the next sections, these results based on simple correlations will be confirmed even when controlling for many sources of variation between the North and the South as well as for many possible sources of endogeneity.

4. Estimation of TFP

In this section, we detail how we obtained our measures of *i*) average local productivity in terms of TFP and *ii*) density of economic activities in terms of workers density.

TFP is estimated at the firm level using the same data as in Locatelli et al. (2019), thus, hereafter, we illustrate only some relevant aspects of the estimation process and refer to that paper for all details in that respect. Specifically, we have data from the CERVED-CB archive on value-added, value of intermediated goods, labor cost, capital stock, location (i.e., the municipality), and economic activity[9] for a large sample of Italian manufacturing firms observed for the period between 1995 and 2015. Our panel includes 188,124 unique

[8] Two quite well-known examples of successful agglomeration stories in Italy include the location of the major Italian firms in northern cities, in particular within the so-called *industrial triangle* (Turin, Milan, and Genoa) and the case of the industrial districts. The former has been considered as one of the main drivers of the industrial take-off of the Italian economy during the first half of the twentieth century. As for industrial districts, these consist of a spatial concentration of small-sized firms mostly located in non-urbanized areas of the north and the center of Italy displaying a strong specialization into specific industrial activities implying a huge accumulation of local competencies in the production of a particular good. They were also defined as a socio-territorial entity characterized by the active presence of both a community of people and a population of firms in one naturally and historically bounded area (Becattini, 1990). Due to the externalities generated within the local network, incumbent firms were able to obtain substantial productivity benefits that are fully consistent with the agglomeration economies described in the text (see Section 2). For previous contributions measuring the local productivity advantages associated with industrial districts, see Signorini (1994) and Signorini (2000). Di Giacinto et al. (2014) compare the productivity of urban areas and industrial districts in Italy, finding that the advantages of industrial districts have been declining over time, while those of urban areas have remained stable.

[9] The classification of industries is obtained by aggregating 19 sectors of the two-digit ATECO classification into 10 categories (see Table B.13) to obtain an adequate number of observations in each cell. Some manufacturing industries have been excluded from our sample: coke and refined oil product firms were omitted because their performance is closely tied to commodity prices; pharmaceutical firms were excluded because their trends are heavily affected by the budget policies for public health expenditure. We also removed the residual sector *other manufacturing activities* because they are generally not very relevant, and their data cannot be easily interpreted.

firms, which correspond to an average of about 74,975 firms per year[10]. Such a figure is substantially larger with respect to the samples used in previous empirical analyses of productivity at the firm level for the manufacturing sectors in Italy. Sample size might vary across years due to entry, exit, change in the legal form of the firms, or other factors[11]. The number of employees, distinguished in white and blue collars, derive instead from the National Institute for Social Security (INPS) database (we merged the two datasets by firm code). TFP is then estimated from the following production function:

$$Y_{i,t} = A_{i,t} \left(\sum_h s_h \; L_{h,i,t} \right)^{\alpha} K_{i,t}^{\beta} \tag{1}$$

where $Y_{i,t}$ denotes the physical output produced by firm i at time t, and $A_{i,t}$ is the TFP. As for labor input, the production function in Eq. (1) considers the possibility to use different h types of L_h employees, with $(h = 1, \ldots, H)$, each having a different efficiency represented by the parameter s_h. Furthermore, $K_{i,t}$ represents the capital stock owned by firm i at time t. Finally, α and β stand for the production function parameters to be estimated to recover the value of TFP. Since we do not have data on physical variables, the output that we observe is actually the firm value-added. Hence, our estimate is the following value-added based TFP:

$$VA_{i,t} = P_{i,t} \; A_{i,t} \left(\sum_h s_h \; L_{h,i,t} \right)^{\alpha} K_{i,t}^{\beta} \;\rightarrow\; P_{i,t} \; A_{i,t} = \frac{VA_{i,t}}{\left(\sum_h s_h \; L_{h,i,t} \right)^{\alpha} K_{i,t}^{\beta}} \tag{2}$$

where $P_{i,t}$ and $VA_{i,t}$ denotes the price of the output and the value-added of firm i at time t, respectively. Our monetary variables are all deflated[12]. However, deflated variables are still not a proxy of physical outputs because they do not account for differences in prices across locations, due to different demand elasticity or firms concentration. This means that the determinants of our empirical measure of TFP at a local level will also include those factors that might affect the output prices of the local incumbent firms. Notice, however, that our firms belong to the manufacturing sector, and therefore their market is expected to be mainly not local[13].

[10] Note that roughly one-third of all the firms (i.e., 36.2%) are observed for at least 10 years and less than one-tenth (i.e., 8.5%) throughout the entire reference period.

[11] Each firm is observed for 8.4 years, on average. Note that we do not have information on the factors causing the entry (exit) of the firms into (from) our sample. Since we cannot determine whether the exit of a firm is due either to actual bankruptcy or misreported data for that given year, we cannot associate such an event to selection processes.

[12] All the variables from the CERVED-CB archive (i.e., net revenues, value-added, tangible fixed assets, and cost of labor) are deflated with the Eurostat sector-specific deflator of the value-added with a base year equal to 2010.

[13] In a recent paper, Mion and Jacob (2020) demonstrate, using a large sample of manufacturing firms, that differences in prices explain a large fraction of the revenue-productivity advantage of denser areas in France, thus suggesting that less productive regions could be more disadvantaged in terms of their competitiveness than for lower technical efficiency.

As for the measure of capital stock, our proxy is through the book value of the (deflated) data on tangible assets, net of amortization and depreciation. Moreover, to take account of the heterogeneity of the labor input, we follow both Fox and Smeets (2011) and Locatelli et al. (2019) and use the total wage bill paid by firm i at time t, $W_{i,t}$, as a proxy for $\sum_h s_h \; L_{h,i,t}$ in Eq. (1). Locatelli et al. (2019) discuss at length about the consequences deriving from alternative measures for labor inputs on the TFP differential across Italian macro-regions (see more on this later). Using logarithms, we get:

$$\log\left(VA_{i,t}\right) = \alpha \;\log\left(W_{i,t}\right) + \beta \;\log\left(K_{i,t}\right) + \varepsilon_{i,t} \tag{3}$$

where the error term $\varepsilon_{i,t} = \log\left(P_{i,t}\; A_{i,t}\right) + \mu_{i,t}$ includes our empirical TFP measure and the actual error term stemming from measurement errors in the production function inputs. To estimate Eq. (3) and correct for input-output simultaneity, we use the Levinsohn and Petrin (2003) procedure[14], we also compute the residuals and define our empirical measure of TFP as follows:

$$TFP_{i,t} = \log\left(VA_{i,t}\right) - \hat{\alpha}\;\log\left(W_{i,t}\right) - \hat{\beta}\;\log\left(K_{i,t}\right) \tag{4}$$

Remember that estimations are carried out separately for each of the 10 manufacturing sectors defined in Footnote 9 and Table B.13.

The second step of our empirical strategy consists of getting an aggregate measure of TFP at the local level. Knowing the municipality where each firm is located, we are able to place those firms within our mapping system based on the LLMA definition[15]. Based on that, we first compute the following quantity:

$$TFP_{r,s,t} = \sum_{i\in(r,s)} \left(\frac{L_{i,t}}{L_{r,s,t}}\right) TFP_{i,t} \tag{5}$$

where we get a weighted average of individual firm TFP with weights defined by the share of employment in firm i over the total area ($r = 1, \dots, 610$), sector ($s = 1, \dots, 10$), and year ($t = 1995, \dots, 2015$) employment. Through this averaging, large firms are assigned more weight in the TFP computation[16]. Following Combes et al. (2010), as a further step in the aggregation procedure, we run a WLS regression[17]:

[14] See more on this in the section on robustness checks of Locatelli et al. (2019).

[15] Note that our sample will be composed of 610 out of the 611 existing LLMAs. One very small local market (i.e., Ayas) located in a mountain region has been dropped from the analysis since we do not have any firm in our sample belonging to that area.

[16] We computed this variable as a weighted average, alternatively using the employment or the wages or the value-added as weights. Our preferred weighting scheme, based on the employment, can also be interpreted as a measure of the productivity of the aggregated production function of the area r and the sector s, assuming constant returns to scale (see Appendix A). Moreover, it is the same definition adopted by Combes et al. (2010).

[17] Weights are given by the number of firms in each LLMA and sector.

$$TFP_{r,s,t} = \delta_s + \phi_{r,s,t} \tag{6}$$

where δ_s represents industry fixed effects. We then define $TFP_{r,t}$ as the average of estimated residuals of Eq. (6) by area and year. This measure allows us to get rid of the differences in terms of sectoral composition that characterize the local markets and makes TFP comparable. $TFP_{r,t}$ is further averaged across t, thus obtaining the variable TFP_r. We average these data because we are interested in the long term effects of agglomeration economies. The averaging process should also help to reduce the effects of measurement errors. We are now ready to express the model that we propose for estimating the intensity of agglomeration economies in Italy, i.e., the elasticity of productivity to density:

$$TFP_r = \gamma \ \log\left(\frac{L_r}{S_r}\right) + X_r \ \psi + \omega_r \tag{7}$$

L_r/S_r (in logarithms) is the main variable of the model, as it is our measure of agglomeration. In the previous literature, the concentration of economic activity has been variously defined (as the count or spatial density of workers or firms). Accordingly, we used alternative definitions for this variable. Some pieces of evidence associated with the alternative concentration measures are reported in the robustness checks section below. The main variable in our baseline model is (the logarithm of) the number of workers L_r divided by the surface of the area S_r. L_r includes the employees of all the sectors featuring the local economy in 2001 except those working for the public administration[18]. We decided to include the workers of sectors not considered in our TFP analysis (in particular, those operating in the service industries) because we want to capture their possible contribution to the generation and transmission of agglomeration externalities. Provided that the parameter γ is correctly estimated, we will be able to answer questions such as the impact on local productivity derived from shifts in the L_r/S_r ratio (i.e., when density doubles, the impact on productivity is equal to $2^\gamma - 1$). The vector of the explanatory variables X_r contains various controls that are discussed in the next section.

5. The econometric strategy and the choice of instruments and controls

To tackle the endogeneity of $\log(L_r/S_r)$ under the form of both omitted variables and reverse causation, we resort to an instrumental variable (IV) regression. As usual, variables that are suitable to play the role of instruments have to be correlated, conditional on the other exogenous regressors in the model, with the endogenous variables (i.e., the relevance) and uncorrelated with the error term in the main equation (i.e., the exogeneity). The latter orthogonality condition can also be expressed by saying that the instrument has to affect the dependent variable in the main specification only through its impact on the

[18] Data are obtained from the industry and service census of ISTAT.

endogenous variable (i.e., the exclusion restriction), in our case $\log(L_r/S_r)$. It is relatively straightforward to check for the validity of the instruments, while it is cumbersome to assess their exogeneity.

We follow previous literature (among others, see Ciccone and Hall, 1996 and Combes et al., 2010), and resort to data from several waves of Italian population censuses dating back to 1861 as the first set of instruments. In particular, our preferred instrument is defined by the population density at the LLMA level in 1921[19]. We argue that the past population density was high in places featuring high land fertility. Once cities are created, they usually display a strong morphology persistence through time, and this might justify the respect of the relevance condition for our instrument. At the same time, while high land fertility is important to determine the productivity of firms in agriculture, it is unlikely that it might have relevance for the TFP of manufacturing firms nowadays. This orthogonality condition might not hold in the presence of other long-term factors that could have driven both population density and productivity in recent times. For instance, proximity to a coastline and similar local amenities could be one of those factors. To circumvent this potential criticism against the exogeneity of our instrument, we add similar variables as well as a set of spatial fixed effects in our main specification.

Furthermore, aside from the aforementioned long-term factors, we argue that during the last century, the Italian economy underwent many structural transformations to make the orthogonality condition very likely to hold: the transition from an agricultural-based economy to one focused on industry and services (for a description on the Italian industrial take-off, also known as the *miracolo economico*, see Daniele et al., 2018), the evolution of technologies including the advent of the digital economy in the last decades, the mass scholarization that led the Italian economy to achieve levels of human capital that are comparable to those of other European countries. As for political changes, Italy was still ruled by a monarchy in 1921. The fascist regime came then to power between 1922 and 1943, followed by a constitutional republic from the end of WWII onwards.

Our second group of instruments includes a rich dataset of geological data about soil characteristics and (historical) climate variables, such as seasonal rainfalls, and January and July temperature measured for each of the LLMA. The rationale for introducing those additional instruments is that they might add explicative power to the prediction of the endogenous variables beyond what it is already accomplished with the other historical instruments (see Combes and Gobillon, 2015). That goes without saying that these variables should also meet the exogeneity requirement. Therefore, the main problem, in this case, would be that of not matching the conditions to be valid instruments.

Detecting the exact spatial range of the agglomeration economies is a difficult task[20]. As

[19] Although having observations that date back to 1861, we opted to use 1921 as the reference point for our historical instruments since most of the territorial changes that have altered the national borders and the number of municipalities occurred before that year. However, the results do not change substantially if longer time lags are used in the specifications.

[20] Agglomeration effects generate their impact at different spatial scales, but the degree of proximity

explained above, our spatial units based on LLMAs are the ideal candidates to gauge the intensity of the positive spillovers that might be generated within spatially concentrated and self-contained local labor markets. However, although the degree of self-containment of each LLMA is high by construction, it is far from being complete[21]. In other words, a local market interacts with other spatial units in terms of employment relationships, and the strength of these interactions is likely to increase with the proximity between the focal LLMA and its neighbors. Therefore, the positive externalities linked to agglomeration might spill over across different and closer LLMAs. If not properly accounted for, this circumstance could affect the quality of our estimates. To this aim, our baseline specification includes the variable CTG_r, defined as the sum of (the logarithm of) the employment densities of the local markets contiguous to the focal LLMA r:

$$CTG_r = \sum_{l \in B(r)} \log\left(\frac{L_l}{S_l}\right) \tag{8}$$

where $B(r)$ is the set of LLMAs sharing one or multiple borders with the focal local market r. We also compute an instrumental variable for the density of contiguous LLMAs using the population density in 1921 because we consider CTG_r as endogenous.

Our controls include the fraction of LLMA land with direct access to a coastline, (the logarithm of) its average altitude, and different sets of dummy variables describing the five Italian macro-regions (i.e., North-West, North-East, Centre, South, and Islands), 20 regions, and 110 provinces (respectively corresponding to NUTS-1, NUTS-2, and NUTS-3 classifications). Given the sharp and persistent differences across Italian regions that might be correlated with our instruments and the productivity indicator, the spatial controls play an essential role in our context.

The summary descriptives of all instruments and controls are reported in Table 2, whereas the correlation matrix is provided in Table 3.

6. Model specification and the estimation of agglomeration economies in Italy

We are now ready to estimate Eq. (7), where TFP_r is the response variable and $\log(L_r/S_r)$ is the main regressor. Our baseline specification includes the full set of controls (the employment density of the contiguous markets, the variables measuring access to a coastline and average altitude and one of the sets of territorial dummies) and a subset of instruments:

that matters is not usually investigated in the prevailing literature. However, such an issue is obtaining growing attention in light of the continued importance of closeness and the large decrease in interaction costs witnessed in recent years. Rosenthal and Strange (2020) offer robust evidence of the crucial role of very short-scale spillover, at the neighborhood or even block levels.

[21] In 54.3% of the units, which represent 71.3% of the national population, the average self-containment is equal to 81.2%, meaning that about three-quarters of the labor force lives and works inside the borders of the LLMA. More information is available on the ISTAT website (www.istat.it/it/informazioni-territoriali-e-cartografiche/sistemi-locali-del-lavoro/indicatori-di-qualità-sll).

(the logarithm of) the population density in 1921, two variables that represent the local soil characteristics (i.e., ruggedness and depth to rock), the logarithm of rainfall, and the average temperature in 1921.

Table 4 reports the main results for the IV estimation for each group of spatial controls and the corresponding OLS results[22]. Note that standard errors are computed with the default variance-covariance estimator in the former set of models whereas they are clustered at the level of the spatial controls (i.e., macro-areas, regions, and provinces) in the latter ones[23][24]. Starting from columns (1) through (3), the diagnostic tests clearly indicate that we can reject the null hypothesis of exogeneity for our density variable and, hence, that it is the case to move to an IV approach. Both the partial R-squared in the first stage and the F test on the null hypothesis of simultaneous irrelevance of all the instruments point to the fact that our instrumental variables are strongly correlated with the employment density (i.e., they are relevant). As for the exogeneity of the instruments, the over-identification test is passed for two of the three specifications, the one in which the test fails is that including macro-area fixed effects.

The estimated parameter for the employment density is positive and significant through all the three IV models. The magnitude of the elasticity is around 6% and is relatively stable across the different specifications, including the more demanding one based on the 110 provinces fixed effects. These results are extremely important in our context as they show that we are not excluding relevant and persistent local factors that might influence both our instruments and the dependent variable. From here onwards, we will comment results for the specification that includes the 20 regional effects, i.e., column (2). The reason for this preference is that the latter specification has a set of rich spatial controls that should attenuate the omitted variable problem and, at the same time, should not cause problems with the degrees of freedom (remember that our cross-section has 610 observations). However, all our estimations have also been carried out for the other two sets of spatial controls based on macro-regions and provinces. The results are rather stable, but we will mention the exception to this while we proceed through the analysis.

By comparing the IV results to the corresponding OLS, i.e., columns (4) through (6), we observe that the estimated elasticities are larger in the latter case. In particular, for our

[22] See Table B.14 for the estimates of the models without the contiguous employment density.

[23] We tested the models by specifying an error structure that allows for intragroup correlation among the observations as well as using a set of arbitrary correlation regressions estimated in both spatial and network settings with the `acreg` library (Colella et al., 2019). In the first case, standard errors are adjusted by considering the physical distance between LLMAs (Conley, 1999) with several cutoff thresholds (i.e., from 100 to 300 kilometers) whereas in the second one we consider the adjacency matrix reporting the links between neighboring local markets. In other words, we assume that the standard error of each LLMA is correlated with those of the other units that are either located within a given radius from the focal one or contiguous to it. Our results are robust to such alternative specifications.

[24] We do not show clustered standard errors for the set of models computed with the 2SLS estimator because such an option is not compatible with the post-estimation diagnostics on instruments relevance and exogeneity.

preferred gauge, the parameter decreases from 0.078 in the case of OLS to 0.059 for the IV estimation. Correcting for endogeneity seems to have a higher impact in our case as compared to other contributions in the literature (see Melo et al., 2009; Combes et al., 2010; De La Roca and Puga, 2017). Notice also that the parameter in the IV estimation is quite precisely measured even when it is compared to the corresponding OLS regression (Wooldridge, 2010).

All in all, we confirm the existence of positive effects deriving from the local density of economic activities in Italy. An estimated elasticity of 5.9% would involve that doubling employment density would increase the productivity of incumbent firms by slightly more than 4%. Moving from a location at the twenty-fifth percentile of the density distribution to one at the seventy-fifth would increase the local TFP by around 11.0%, given that the ratio between the two densities amounts to a factor of 5.87 (see again Table 1). Although being in the upper tail of the distribution, our estimated elasticity is in the range of those estimated for other developed countries that usually fluctuate between 2% and 9%. However, it is not easy to make a proper comparison due to the sharp differences between methodologies, data, and years examined (Melo et al., 2009). Ciccone and Hall (1996), a seminal paper in this field, estimated an elasticity close to ours for the US. Combes et al. (2010) adopted a very similar methodology to that developed in this paper and found an elasticity between 3.5% and 4% in France, displaying a lower intensity of agglomeration economies for that country. As for Italy, Di Giacinto et al. (2014) estimated a productivity advantage associated with urban areas, i.e., LLMA with more than 200,000 inhabitants, of approximately 10% using comparable data to ours. Cingano and Schivardi (2004), were among the first to measure agglomeration economies by resorting to firm-level TFP data. Although focused on TFP dynamics, they reported in a footnote (p. 735) that the elasticity of TFP with respect to the logarithm of local manufacturing employment was equal to 6.7%, a figure that is very close to our results.

As for the controls, we emphasize the interesting pattern of the results for the contiguity variable, CTG_r. Its coefficient is positive, significant, and highly stable across estimation methods and specifications. The density of the contiguous LLMAs has a positive effect on the local TFP, thereby pointing to the fact that agglomeration economies are not limited within the borders of each local market and instead may travel across different LLMAs (see more on this in the next sections).

6.1. Robustness checks

Before moving to the second part of the paper dedicated to the North-South TFP disparities, we devise a set of robustness checks for our results. First, we introduce a quadratic term of the density in our baseline specification. The idea is that of capturing the variable effects that agglomeration can have on productivity depending on the level of concentration of economic activity (Au and Henderson, 2006; Graham, 2007). Specifically, we resort to the following specification:

$$TFP_r = \gamma \ \log\left(\frac{L_r}{S_r}\right) + \theta \left[\log\left(\frac{L_r}{S_r}\right)\right]^2 + X_r \ \psi + \omega_r \tag{9}$$

To estimate Eq. (9) with IV, we follow the suggestion by Wooldridge (2010) and use (the square of) the linear prediction of $\log(L_r/S_r)$ regressed on all the exogenous variables (both the controls and the set of instrumental variables used in the previous estimation) as an additional instrument. The results are reported in Table 5. It turns out that the estimated parameter for the linear term is still positive and significant. The coefficient of the quadratic term is negative, significant, and precisely estimated. The results jointly indicate that the returns from agglomeration are positive but diminishing. To investigate further this issue, we plot marginal effects in Figure B.3a and the function linking TFP to the density in Figure B.3b by using the estimated parameter from the IV estimation of Eq. (9) and the actual data for $\log(L_r/S_r)$. Although we recognize that non-linear effects would be an interesting and encouraging extension of our analysis, we are induced to consider this additional evidence as a robustness check. In fact, the specification including a quadratic term basically does not upset the previous results and provides a similar and consistent picture compared to that of linear estimates. Moreover, the range of variation that characterizes the employment density lies completely on the increasing branch of the curve.

The second robustness check involves adding into the baseline specification a distance-weighted density of LLMAs as a potential substitute for the contiguity variable to measure the interactions between local markets. This extra indicator is defined as follows:

$$DWD_r = \sum_{l \neq r} \log\left(\frac{L_l}{S_l}\right) \frac{1}{d_{l,r}} \tag{10}$$

where $d_{l,r}$ is the Euclidean distance between the central locations of LLMAs l and r. The rationale for introducing this new variable is twofold. First, as already explained, we cannot rule out that the positive external economies related to density might cross the border of the LLMAs. Specifically, the distance-weighted density involves that each local market can interact with all the others and the strength of these positive externalities might dissipate with the distance between different local markets, i.e., positive spillovers are stronger the closer are two LLMAs. Moreover, as already explained, the presence of the output price in our empirical proxy for firm TFP could introduce pecuniary externalities between different locations similar to those modeled in the new economic geography literature. These interlinkages are certainly related to some market access that could be proxied (although, sometimes, very loosely proxied) by the sort of market potential *à la Harris* that we have introduced in Eq. (10).

The results are reported in Table 6. Notice that we consider DWD_r as endogenous and, on that ground, the distance-weighted density computed on (the logarithm of) the population density in 1921 was added to the set of previous instruments. First, we observe that the estimated parameter for the density variable is marginally affected by the introduction of DWD_r into the baseline regression. The elasticity is now equal to about 6.7%, a level that is very close to the value of the baseline model. Furthermore, unlike the contiguity variable, the coefficient of the distance-weighted density is never significantly different from zero. We interpret this evidence as in favor of a very localized nature that characterizes the positive externalities occurring between different LLMAs. In any case, our findings seem to be robust to this check.

A third direction that seems important for verifying the robustness of our results concerns the choice of the variable to measure the spatial concentration of economic activity. To this aim, we used as alternative proxies *a*) the logarithm of the population density, *b*) the logarithm of the employment density restricted to the manufacturing sector, *c*) the logarithm of the employment level in all sectors, and *d*) the logarithm of the number of local units as in Henderson (2003). The results of these alternative specifications, reported in Table 6, are stable across the definitions, confirming similar findings in previous literature.

Finally, taking into account the heterogeneity of labor inputs is essential for the estimation of TFP and the assessment of the magnitude of the agglomeration economies. As explained above, we model labor heterogeneity at the firm level using the total wage bill as a proxy. Moreover, we control for input-output simultaneity at the firm level by resorting to the Levinsohn and Petrin (2003) procedure. To check previous findings along this perspective, we measure labor input with *a*) the logarithm of the number of white and blue collars measured as separate inputs[25], or *b*) the logarithm of the total number of firm employees.

For those measures, we get alternative TFP estimations using the methodology by Levinsohn and Petrin (2003). Hence, firm-level data are aggregated for the 610 LLMA as before. The new results, shown in Table 8, deliver an interesting and consistent pattern. When resorting to the white-blue collar counts, a distinction that captures at least some aspects of the workforce heterogeneity in terms of quality, estimated elasticity amounts to 0.084. When labor input is approximated via the total number of employees, i.e., corresponding to the case where all workers are considered to have the same quality and produce the same effort, it rises to 0.106. In other words, the lack of control for the labor quality across firms translates into an upper bias in estimating the agglomeration economies. Notably, firms with better endowments of human capital tend to concentrate in denser areas.

Moreover, there is also the possibility that firms benefit from positive externalities in agglomerated areas because highly skilled workers in other firms or industries concentrate there. To check for this possibility, we introduce into our baseline specification the additional explanatory variable defined by the share of people having a diploma or a university degree in each LLMA. We resort to the same set of instruments as in our baseline, using as alternative dependent variables the three definitions of TFP, i.e., *a*) the

[25] In this case, the measure of TFP is computed according to the specification described in Hellerstein et al. (1999) as follows: $VA_{i,t} = \alpha\, L_{i,t} + \alpha\, \varphi\, (L^w_{i,t}/L_{i,t}) + \beta\, K_{i,t} + \widetilde{\phi}_{i,t}$, where L^w is the number of white collar employees.

one based on labor cost, *b*) the number of white and blue collars, or *c*) the total count of workers (see Table 9). It turns out that the estimated parameter for this additional regressor is not significantly different from zero with the first and third TFP indicator, while it is positive and weakly significant for the remaining proxy. These results might be due either to the fact that we have poor proxies for the differences in terms of human capital across LLMAs, or alternatively, because once properly controlled for the variations in labor quality at the individual firm level, the regressor capturing human capital at the aggregate level is irrelevant. In any case, the estimated elasticity of the density with respect to the productivity is marginally affected by the additional regressor measuring human capital when labor cost is the proxy for the labor input while it does change along the pattern that has been described above when we use the white-blue collar counts and the total employees.

As said, we also tested our baseline specification on different aggregation procedures for the TFP, as defined in Eq. (5). In unreported evidence, we obtain that the coefficient of the density increases, passing from an average weighted by the share of the employment to an average weighted by the share of wages and to the share of value added[26].

As a final check for our estimations, we computed TFP using the alternative methodologies based on the works by Olley and Pakes (1996) and Ackerberg et al. (2015). In unreported evidence, we find that these estimations do not converge and do not deliver reasonable values for the input coefficients. In particular, Kim et al. (2019) observe that, given their non-linear nature, these methods might provide results that are unstable and extremely sensitive to the initial conditions. For these reasons and additional motivations advanced in Fox and Smeets (2011) we keep our estimates based on the approach by Levinsohn and Petrin (2003).

7. Agglomeration economies and the North-South divide

7.1. The productivity gap

Armed with this evidence, we move to the exploration of the North-South divide in Italy. Agglomeration economies can be a relevant factor to explain this gap provided three conditions hold: *i*) there are substantial and persistent imbalances in the agglomeration rates between the North and the South of the country; *ii*) these differences translate into positive TFP differentials in favor of the more agglomerated northern areas because of the positive externalities generated via the concentration of firms and workers in specific local markets; *iii*) productivity differences are also crucial to explain the backwardness of the South in terms of other indicators, such as the GDP per capita or additional welfare measures.

Whatever the initial causes that might explain why the southern regions were not a favorable environment for the concentration of firms and workers, the agglomeration

[26] Similar evidence is reported in Mion and Jacob (2020).

processes might have supported a dynamic effect that self-feeds such differentials of density[27]. For instance, in Toniolo (2013) the interested reader can find a synthesis of the long-standing debate between historians and economists concerning the causes that could explain the relative backwardness of southern regions and its persistence over time. Felice (2018) summarized the aforementioned explanations of the North-South divide into those related to geographical, social capital, and cultural factors, the exploitation of southern regions by the northern ones, and differences in the socio-institutional devices. The author empirically discussed the relative merits of those alternative explanations concluding that the differences between the formal and informal institutions in the North and the South were the main driver of the underdevelopment of the latter. In any case, we want to emphasize that whenever these alternative explanations predicted a pattern where southern regions were less agglomerated than the northern ones, they could be made compatible with our argument based on agglomeration economies.

Coming to the papers dealing with the North-South differences in productivity, a stream of literature used aggregate data to conclude that, even limiting the analysis from the end of WWII onwards, firms in southern regions displayed much lower TFP levels (e.g., Mauro and Podrecca, 1994; Aiello and Scoppa, 2000; Di Liberto et al., 2008; Felice, 2019)[28]. As for the catching-up of the South, evidence was mixed[29].

In recent years a number of studies have employed firm-level data to quantify the magnitude of the North-South TFP productivity gap in Italy. This kind of data allows better control for the differences in the input composition and significantly enlarges the range of the analysis to topics such as market selection and misallocation[30]. Di Giacinto et al. (2014) with a large panel of firms observed during the years between 1995 and 2006 measure productivity disadvantages in the South, controlling for many factors including the degree of urbanization and the presence of industrial districts. Furthermore, Rungi and Biancalani (2019) attribute the negative gap of the southern firms to those in the left tail of the TFP distribution, hinting at an inefficient process of market selection in the South. They find a positive correlation between higher productivity of incumbent firms and entry in local markets.

To evaluate the contribution of agglomeration economies to explaining the North-South productivity gap, we move back to the results for our baseline specification. Remember that in our IV specification with the two geographic variables and 20 regional fixed effects, we got an elasticity of TFP with respect to the employment density equal to 5.9%. Now,

[27] Needless to say, we are not denying the importance of disentangling among those alternative explanations.

[28] For other works comparing the North-South divide in Italy with the West-East one in Germany, see Boltho et al. (1997), Boltho et al. (2018), and Boeri et al. (2020).

[29] There was some consensus on the fact that there was a convergence in TFP levels until the Seventies, mainly due to direct and indirect public intervention and capital accumulation, yet there was much more uncertainty about what happened thereafter.

[30] Selection models will be discussed in the final section of the paper while dealing with the issue of misallocation is beyond our current goals.

if we increase the median density for the South to the level of that for the median in the North (i.e., from 14 to 36 workers per square kilometer), we would get an increase of TFP in the South by approximately 5.7%.

As a complement to the previous results and to shed more light on the relative importance of the variables found in our main specification, we ran a decomposition analysis of the TFP regression results based on the Shapley value[31]. The outcomes of this process have been computed starting from the OLS specifications[32] found in Table 4. In Table 11, we repeated the statistical exercise after estimating a set of four models that include the North-South dummy as well as the macro-areas, regions, and provinces indicator variables, respectively. We find that the spatial fixed effects are the most important determinants of TFP. Their joint contribution to the differences in productivity ranges from 40.6% to 53.6% as the spatial controls increases. The employment density is associated with relative impacts between 31.7% and 24.1%, meaning that about a quarter of the TFP variations are due to the concentration of workers in the focal LLMA. Interestingly, the density of contiguous LLMAs has a substantial effect on TFP that amounts to more than half the weight of the previous variable, from 16.9% to 13.4%. Finally, the geographical features of the local market (i.e., its altitude and the share of its costal surface) provide the residual percentage importance.

So far, we have assumed that the TFP-density elasticity is the same across the two macro-regions. This is not necessarily to be the case. For instance, it might be that the quality of local interactions producing the agglomeration economies in the southern LLMAs is lower than the one in the North due to several factors. Corruption and criminality, lower propensity to cooperate and lack of social capital, weak local institutions and infrastructures, a highly sloped congestion cost curve, can all contribute to reducing the positive externalities produced for a given agglomeration level.

To further investigate this issue, we interact the density term in the baseline with the different sets of territorial dummies. Unreported evidence indicates that the interaction terms are never significant, i.e., we do not detect any heterogeneity in the elasticity of local productivity with respect to the density of economic activity. As an alternative strategy, we carry out the same IV regression as in the baseline by splitting the sample between the LLMAs located in the South and those in the North. Results are reported in Table 10. It turns out that the estimated elasticity in the northern LLMAs is actually a bit larger than that in the South (6.2% versus 5.5%), but the difference is relatively modest and never significant[33].

[31] The notion was first introduced by Shapley (1953) in the context of the co-operative game theory. Such a technique makes it possible to calculate an additive and symmetric breakdown of the contribution attributed to each factor included in a regression model.

[32] Note that we employ the estimates of the OLS regressions since it is a technical requirement of the methodology. We acknowledge that it is a limitation of the decomposition technique, given the presence of endogenous variables in the models.

[33] We cannot reject the null hypothesis on the equality of the two coefficients based on the p-value of the Wald test (i.e., 0.692).

7.2. Agglomeration versus selection

The evidence collected so far shows that the substantial differences between the northern and southern LLMAs in terms of TFP are both due to a lower agglomeration of the latter regions, on the one hand, and to alternative factors, captured by spatial fixed effects as well as extra controls (e.g., the natural advantages), on the other hand.

In this section, we explore one of these alternative determinants of the North-South TFP gaps, related to the hypothesis that market selection processes could operate differently in the northern and southern territories (due to competitive or institutional issues, or both), determining specific patterns of productivity for the active firms. Recent models show that if companies are *ex-ante* heterogeneous in terms of productivity, then more competitive markets will lead a larger share of inefficient businesses to exit, thereby increasing local efficiency through a selection effect[34].

Notice that this selection effect, as already discussed in Section 2, would be observationally equivalent to the explanations of the North-South TFP differences based on agglomeration economies or natural advantages. In particular, a positive link between concentration and local productivity could be compatible with both agglomeration economies and selection models. Similar arguments hold for the other controls, the negative spatial effects detected for the southern regions could result from natural disadvantages or a less severe selection mechanisms.

A possible solution to the observational equivalence problem has been proposed by Combes et al. (2012), who designed a method that can disentangle between multiple effects moving from an analysis based on the differences between conditional means to one exploring the discrepancies in the entire productivity distribution (for the northern and southern firms, in our case). Specifically, agglomeration forces or natural advantages rightward shift the TFP distribution as they equally affect all the firms located in those areas. In contrast, selection processes would affect the TFP distribution by operating a leftward shift of its (left) truncation point because they cut a larger share of inefficient firms in the North.

More in detail, Combes et al. (2012) predict that, under general assumptions, the selection processes left-truncate a share S_i of the distribution of the log-productivity (with S growing as the density of the territory increases), while common productivity advantages (including agglomeration) cause the distribution of the log-productivity to be right-shifted by a factor A_i and dilated by a factor D_i (with A and D growing as the density of location i increases). By extending the predictions of Combes et al. (2012), first and second-nature advantages, in principle, are also reasonably expected to generate a shift of the TFP distribution, in the same way as proper agglomeration economies. Furthermore, sorting will tend to dilate the distribution (maybe asymmetrically).

The test developed by Combes et al. (2012) uses non-parametric techniques exploiting only the information conveyed by the empirical cumulative distribution of log-productivity

[34] See Melitz (2003) and Melitz and Ottaviano (2008).

for a specific set of firms[35]. The estimation procedure is based on the idea that two TFP distributions can differ due to a dilating factor D, a shifting factor A, and a left-truncating factor S that influence the values of some underlying distribution with cumulative density function F. The advantage of this method is twofold. First, it does not impose any parametric assumption on the shape of F. Second, unlike the traditional quantile regression approach, it compares all the quantiles of the two distributions and not only specific percentiles, thereby improving the robustness and the efficiency of parameter estimation. However, this degree of generality is achieved at a cost since the procedure only allows to analyze the differences between two distributions. In this sense, the technique essentially implements a univariate test: we will be able to compare the distributions of two regions (North versus South, in particular) or according to the value assumed by a specific variable (e.g., more agglomerated versus less agglomerated LLMAs).

This approach allows us to verify whether differently working selection and sorting mechanisms can explain part of the productivity gap between the northern and southern regions of Italy. Why should selection in the North be more severe than those in the South? One obvious answer is that the LLMAs in the former territories host a larger number of firms and workers per square kilometer; in other words, they are more agglomerated. Let aside the differences in terms of concentration, competition in the North could be tougher due to other manifold factors. Higher efficiency of the local courts might also result in better bankrupt procedures for the northern areas, facilitating the selectivity effects induced through competition.

To implement the above-mentioned methodology, we split the sample of individual firm log-productivity obtained from Eq. (4) into two groups associated with northern and southern firms[36]. Unlike previous literature, we do not compare firms according to the size or density of the local markets where they are located[37]. From previous sections, we know that LLMAs in the South are, on average, smaller and less dense than those in the North, meaning that even in the way we perform the test, we are comparing agglomerated territories to non-agglomerated ones. However, the splitting rule based on macro-regions allows the other forces that may contribute to a shift in the TFP distribution to play a role (e.g., natural advantages, *et similia*, see more on this below).

[35] The underlying tests have been carried out by resorting to the `estquant` library (Kondo, 2017).

[36] As done in the previous sections, data is averaged across years and netted out for sectoral effects.

[37] This approach has been applied by Combes et al. (2012) in comparing the productivity of establishments located in large versus small metropolitan areas of France. Their methodology assumes that the local productivity cut-off point (i.e., the toughness of selection forces) depends on local market size. They conclude that firm selection cannot explain the heterogeneity of spatial productivity. The same test has been subsequently applied to analyze the productivity differentials in Ukraine before and after market deregulation (Shepotylo and Vakhitov, 2015), and for plants belonging to the same industry (silk-reeling) located in a more or less concentrated fashion (Arimoto et al., 2014). In both cases, the selection turns out to have a little role in explaining productivity differentials. In contrast, Ding and Niu (2019) and Howell et al. (2020) show that the higher productivity observed in large Chinese urban agglomerations is shaped by both selection and agglomeration forces.

Moving to the outcome of this comparison (Table 12, Panel A), results are quite clear-cut. A great deal of the North-South differences in TFP is explained by the three parameters A, D, and S. The pseudo R-squared of the baseline model in column (1) is 97.8%. By observing the constrained models in columns from (2) to (4) that alternately exclude dilation and truncation effects (i.e., D and S are respectively set to 1 and 0), the agglomeration parameter, which corresponds to a rightward shift, turns out to be the main determinant. We still find evidence consistent with a contraction of the distribution (D is significantly lower than one but has negligible explanatory power), while the selection parameter S is not statistically significant.

In other words, based on this evidence, we do not have elements to confirm the hypothesis that northern firms are more efficient than those located in the South due to the tougher competition they are exposed to in the LLMAs where they are localized. This result seems at odds with those obtained by Rungi and Biancalani (2019), who find that North-South productivity gaps are driven by the presence of relatively more inefficient firms on the left tail of the distributions. Apart from dilation, our findings are largely in line with Combes et al. (2012) for France and Accetturo et al. (2018) for Italy, whereas they seem to be in contrast with the tenets of selection models, or, more in general, with the mechanisms that leverage firms heterogeneity by dilating or truncating the productivity distribution[38].

We further employ the test developed by Combes et al. (2012) to verify whether the right shift of the productivity distribution of the northern firms is mainly driven by the significant differentials in terms of density that we discussed in the previous sections. We then compare the TFP distributions between the firms located in agglomerated LLMAs (i.e., those with an employment density above the seventy-fifth percentile of the distribution) and the firms located in non-agglomerated ones (i.e., those with density below the seventy-fifth percentile). Results are shown in Table 12, Panel B. Also in this case, shift, dilation, and truncation explain almost all of the differences between the two distributions, as the pseudo R-squared is 97.0%. Again, the results clearly speak to a dominant role of the A factor. The productivity distribution of agglomerated areas is mostly obtained by shifting rightward the distribution of those LLMAs with a low degree of concentration. However, the A coefficient is less than half of the value obtained when comparing the TFP of the northern and southern areas, thus indicating that relevant differences other

[38] Accetturo et al. (2018) argue that the selection effects devised in the Melitz-Ottaviano type of models cannot be observed at the spatial scale of LLMAs. The reason is that these selection effects depend on the size of the market for the output of the manufacturing products, which are frequently defined nationwide or even internationally. The selection forces caused by competition processes on the output market can be observed at a local scale only if supply is geographically concentrated (e.g., the cement market). LLMAs represent local labor markets that we know are among the main sources of local productivity advantages related to agglomeration. Then, whenever we compare a large (or dense) LLMA to a smaller (or less dense one), this is the right context to look for agglomeration effects, but it is not for detecting the selection effects conveyed though the output markets for manufacturing products. However, one can think about selection mechanisms operating though channels that are different from local competition, and that, however, display a different intensity across local markets. In particular, institutional mechanisms (e.g., local courts) can determine very distinct selection outcomes at a local scale.

than agglomeration economies uniformly affect the productivity of firms: first-nature advantages, and possibly heterogeneous infrastructural or institutional endowments.

8. Final remarks

The evidence collected in this paper confirms what a long strand of contributions has found for other countries, i.e., geographic concentration generates productivity advantages for the local incumbent firms. In our preferred estimation, the elasticity of TFP with respect to density in Italy is about 6%, a magnitude comparable to that measured for other developed countries by scholars using methodologies similar to the ones employed in our work. Moreover, we show that the TFP-density nexus can contribute to explaining a large share of the substantial productivity differences between the North and South in Italy. In other words, firms in Southern Italy are less productive than those in the North to a large extent because they compete in local environments that are less concentrated, and, hence, less capable of producing the positive externalities observed in the agglomerated areas where northern firms are located.

We also deliver two other new and somewhat surprising facts complementing this evidence. First, the TFP-density elasticities are not significantly different across the southern and northern macro-regions. Second, we clearly exclude market selection (possibly associated with different intensity of market competition) as an alternative mechanism for explaining the productivity gap in favor of the firms located in the North. Grounded in all this evidence, it would be tempting to conclude that a substantial share of the southern gap in terms of productivity might be traced back to a lack of concentration. However, considering the specific and complex historical events that were at the origin of the agglomeration economies in Italy (briefly described in Footnote 8), we can conclude that the concentration of economic activities is just one precondition to generate an industrial take-off in the southern less developed areas of the country. A further note of caution is necessary in interpreting our findings. The latter refer to the manufacturing industries and, due to a lack of data, do not consider other important sectors such as construction or services. In this perspective, some of our conclusions cannot be generalized to the entire economy, also considering the importance of some of those sectors for the the southern LLMAs.

In any case, the findings of this study can be interpreted in the light of the longstanding debate about the *questione meridionale* (i.e., southern question) in Italy. Our results also speak to the literature on agglomeration economies in developing countries: the Italian South is a special case of a developing region that coexists in a unified country with other areas that are among the most advanced European regions.

References

Accetturo, A., Di Giacinto, V., Micucci, G., and Pagnini, M. (2018). Geography, productivity, and trade: does selection explain why some locations are more productive than others? *Journal of Regional Science*, 58(5):949–979.

Accetturo, A. and Mocetti, S. (2019). Historical origins and developments of Italian cities. *Italian Economic Journal*, 5(2):205–222.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: a synthesis. *Journal of Urban Economics*, 111(C):93–107.

Aiello, F. and Scoppa, V. (2000). Uneven regional development in Italy: explaining differences in productivity levels. *Giornale degli Economisti e Annali di Economia*, 59(2):270–298.

Arimoto, Y., Nakajima, K., and Okazaki, T. (2014). Sources of productivity improvement in industrial clusters: the case of the pre-war Japanese silk-reeling industry. *Regional Science and Urban Economics*, 46(C):27–41.

Au, C. C. and Henderson, V. J. (2006). Are Chinese cities too small? *The Review of Economic Studies*, 73(3):549–576.

Becattini, G. (1990). The Marshalian industrial district as a socio-economic notion. In Pyke, F., Becattini, G., and Sengenberger, W., editors, *Industrial districts and inter-firm co-operation in Italy*, pages 37–51. International Institute for Labour Studies, Geneva.

Boeri, T., Ichino, A., Moretti, E., and Posch, J. (2020). Wage equalization and regional misallocation: evidence from Italian and German provinces. Working Paper No. 25612, National Bureau of Economic Research.

Boltho, A., Carlin, W., and Scaramozzino, P. (1997). Will east Germany become a new Mezzogiorno? *Journal of Comparative Economics*, 24(3):241–264.

Boltho, A., Carlin, W., and Scaramozzino, P. (2018). Why east Germany did not become a new Mezzogiorno. *Journal of Comparative Economics*, 46(1):308–325.

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

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

Colella, F., Lalive, R., Sakalli, S. O., and Thoenig, M. (2019). Inference with arbitrary clustering. Discussion Paper No. 12584, Institute of Labor Economics.

Combes, P. P. (2011). The empirics of economic geography: how to draw policy implications? *Revew of World Economics*, 147(3):567–592.

Combes, P. P., Gilles, D., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Combes, P. P., Gilles, D., Gobillon, L., and Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In Glaeser, E. L., editor, *Agglomeration Economies*, pages 15–66. University of Chicago Press, Chicago, IL.

Combes, P. P., Gilles, D., and Overman, H. G. (2005). Agglomeration and the adjustment of the spatial economy. *Papers in Regional Science*, 84(3):311–349.

Combes, P. P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, V. J., and Strange, W., editors, *Handbook of regional and urban economics*, volume 5, pages 247–348. Elsevier, Amsterdam.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Daniele, V., Malanima, P., and Ostuni, N. (2018). Geography, market potential and industrialization in Italy, 1971-2001. *Papers in Regional Science*, 97(3):639–662.

De La Roca, J. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

Di Giacinto, V., Gomellini, M., Micucci, G., and Pagnini, M. (2014). Mapping local productivity advantages in Italy: industrial districts, cities or both? *Journal of Economic Geography*, 14(2):365–394.

Di Liberto, A., Pigliaru, F., and Mura, R. (2008). How to measure the unobservable: a panel technique for the analysis of TFP convergence. *Oxford Economic Papers*, 60(2):343–368.

Ding, C. and Niu, Y. (2019). Market size, competition, and firm productivity for manufacturing in China. *Regional Science and Urban Economics*, 74(C):81–98.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, V. J. and Thisse, J. F., editors, *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier, Amsterdam.

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

Felice, E. (2018). The socio-institutional divide: explaining Italy's long-term regional differences. *Journal of Interdisciplinary History*, 49(1):43–70.

Felice, E. (2019). The roots of a dual equilibrium: GDP, productivity, and structural change in the Italian regions in the long run, 1871-2011. *European Review of Economic History*, 23(4):499–528.

Fox, J. T. and Smeets, V. (2011). Does input quality drive measured differences in firm productivity? *International Economic Review*, 52(4):961–989.

Fujita, M., Krugman, P. R., and Venables, A. (1999). *The spatial economy: cities, regions, and international trade.* Massachusetts Institute of Technology Press, Cambridge, MA.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–3153.

Graham, D. J. (2007). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, 62(1):103–120.

Graham, D. J. (2009). Identifying urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science*, 88(1):63–84.

Hellerstein, J. K., Neumark, D., and Troske, K. R. (1999). Wages, productivity, and worker characteristics: evidence from plant-level production functions and wage equations. *Journal of Labor Economics*, 17(3):409–446.

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

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

Henderson, V. J., Squires, T., Storeygard, A., and Weil, D. (2018). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Howell, A., Liu, C., and Yang, R. (2020). Explaining the urban premium in Chinese cities and the role of place-based policies. *Environment and Planning A: Economy and Space*.

Kim, K. I., Luo, Y., and Su, Y. (2019). A robust approach to estimating production functions: replication of the ACF procedure. *Journal of Applied Econometrics*, 34(4):612–619.

Kondo, K. (2017). Quantile approach for distinguishing agglomeration from firm selection in Stata. Technical Paper No. 17-T-001, Research Institute of Economy, Trade and Industry.

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

Locatelli, A., Ciani, E., and Pagnini, M. (2019). TFP differentials across Italian macro-regions: an analysis of manufacturing corporations between 1995 and 2015. *Politica Economica*, 35(2):209–242.

Mauro, L. and Podrecca, E. (1994). The case of Italian regions: convergence or dualism? *Economic Notes*, 23(3):447–472.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *The Review of Economic Studies*, 75(1):295–316.

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

Mion, G. and Jacob, N. (2020). On the productivity advantage of cities. Discussion Paper No. 14644, Centre for Economic Policy Research.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In Henderson, V. J. and Thisse, J. F., editors, *Handbook of regional and urban economics*, volume 4, pages 2119–2171. Elsevier, Amsterdam.

Rosenthal, S. S. and Strange, W. C. (2020). How close is close? The spatial reach of agglomeration economies. *The Journal of Economic Perspectives*, 34(3):27–49.

Rungi, A. and Biancalani, F. (2019). Heterogeneous firms and the North–South divide in Italy. *Italian Economic Journal*, 5(3):325–347.

Segal, D. (1976). Are there returns to scale in city size? *The Review of Economics and Statistics*, 58(3):339–350.

Shapley, L. S. (1953). Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100.

Shepotylo, O. and Vakhitov, V. (2015). Services liberalization and productivity of manufacturing firms: evidence from Ukraine. *Economics of Transition*, 23(1):1–44.

Signorini, L. F. (1994). The price of Prato, or measuring the industrial district effect. *Papers in Regional Science*, 73(4):369–392.

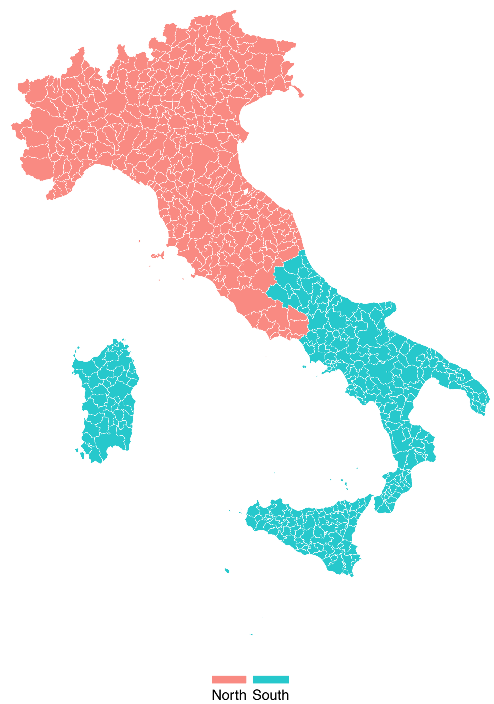
Signorini, L. F., editor (2000). *Lo sviluppo locale: un'indagine della Banca d'Italia sui distretti industriali*. Meridiana Libri, Roma.

Sveikauskas, L. (1975). The productivity of cities. *The Quarterly Journal of Economics*, 89(3):383–413.

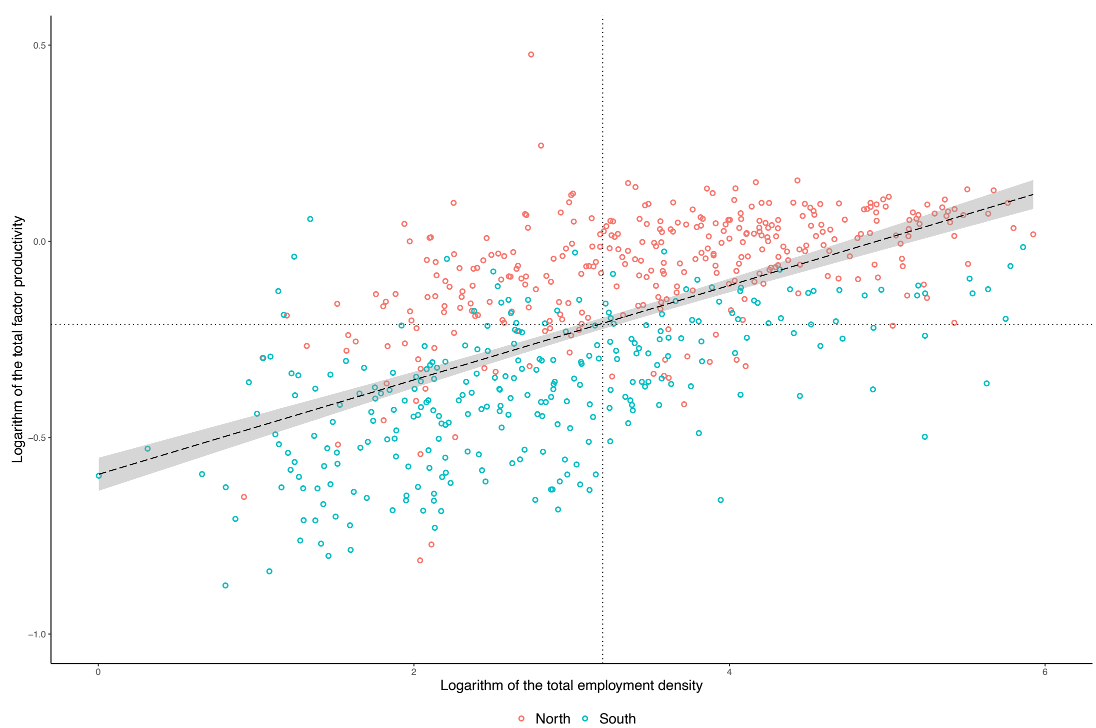
Toniolo, G. (2013). *The Oxford handbook of the Italian economy since unification*. Oxford University Press, Oxford.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Massachusetts Institute of Technology Press, Cambridge, MA.

Figure 1: Territorial reference grid (a) and productivity versus employment density (b) in Italy

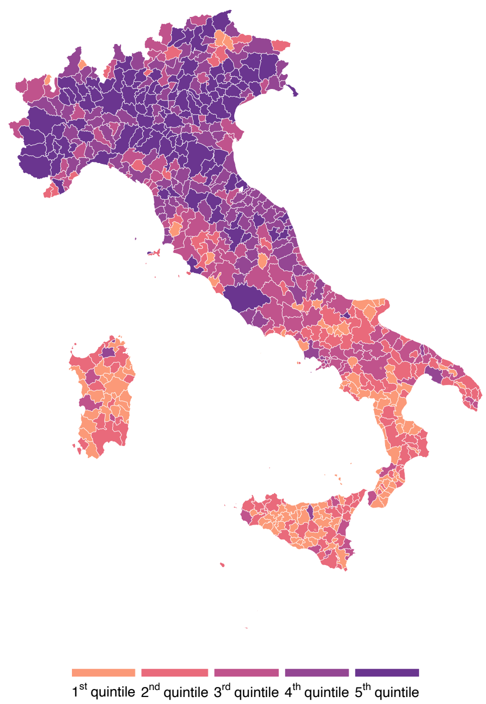


(a) LLMAs belonging to the North and the South

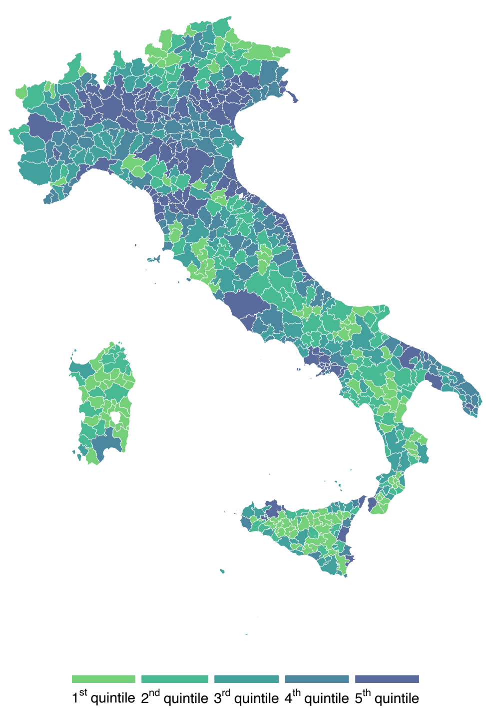


(b) Scatter plot, regression line, and confidence interval of density versus productivity

Figure 2: Spatial distributions of total factor productivity (a) and total employment density (b) by LLMA



(a) Choropleth map of (the logarithm of) total factor productivity



(b) Choropleth map of (the logarithm of) total employment density

Table 1: Descriptive statistics on the differences in TFP and total employment density by macro-zone

Total factor productivity	**Mean**	**1st perc.**	**25th perc.**	**50th perc.**	**75th perc.**	**99th perc.**
North	0.947	0.522	0.868	0.963	1.044	1.166
South	0.694	0.432	0.600	0.687	0.783	0.975
Difference	-0.253***	-0.090***	-0.268***	-0.275***	-0.261***	-0.192***
	(0.011)	(0.021)	(0.000)	(0.000)	(0.000)	(0.009)
Employment density	**Mean**	**1st perc.**	**25th perc.**	**50th perc.**	**75th perc.**	**99th perc.**
North	62.038	3.748	15.383	36.307	74.201	318.607
South	33.567	1.930	7.133	14.139	29.582	323.856
Difference	-28.471***	-1.818***	-8.251***	-22.168***	-44.619***	5.250
	(5.556)	(0.184)	(0.048)	(0.111)	(0.083)	(7.427)

Differences in total factor productivity (TFP) and total employment density across LLMAs located in each macro-zone are tested by regressing the two variables on the South dummy. Standard errors robust to heteroskedasticity and intra-cluster correlation at the level of the LLMA are reported in parentheses. Mean differences are tested with OLS regressions whereas differences for all percentiles of the distributions are tested with quantile regressions. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 2: Descriptive statistics

Variable	Count	Mean	Median	SD	Min	Max
Regressands on productivity						
TFP estimated with the cost of labor (ln)	610	-0.211	-0.176	0.234	-1.086	0.476
TFP estimated with the number of workers (ln)	610	-0.383	-0.336	0.395	-1.652	0.724
TFP estimated with the number of white and blue collars (ln)	610	-0.291	-0.229	0.371	-1.552	0.793
Regressors on agglomeration						
Total employment density in 2001 (ln)	610	3.199	3.143	1.178	0.003	6.714
Manufacturing employment density in 2001 (ln)	610	1.689	1.634	1.479	-2.784	5.208
Population density in 2001 (ln)	610	4.771	4.742	0.975	2.339	8.037
Total employment in 2001 (ln)	610	9.136	9.096	1.326	5.849	14.230
Local units in 2001 (ln)	610	8.076	8.023	1.135	5.509	12.749
Contiguous total employment density in 2001	610	16.426	15.413	7.674	0.000	46.954
Contiguous manufacturing employment density in 2001	610	8.939	8.052	7.380	-5.614	34.727
Contiguous population density in 2001	610	24.396	23.550	9.486	0.000	59.656
Contiguous total employment (thousands) in 2001	610	169.004	90.351	271.019	0.000	1,999.622
Contiguous local units (thousands) in 2001	610	46.284	27.437	65.461	0.000	458.801
Distance-weighted total employment density in 2001	610	0.008	0.008	0.002	0.004	0.012
Regressors on geological characteristics						
Altitude (ln)	610	5.470	5.762	1.180	0.000	7.326
Share of coastal surface	610	0.187	0.000	0.302	0.000	1.000
Regressors on education						
Share of population with a diploma	610	0.292	0.292	0.051	0.149	0.464
Instruments						
Population density in 1921 (ln)	610	4.638	4.694	0.801	2.376	7.380
Population in 1921 (ln)	610	10.574	10.531	0.935	7.555	14.202
Contiguous population density in 1921	610	23.840	23.345	9.358	0.000	57.358
Contiguous population (thousands) in 1921	610	399.173	299.801	370.747	0.000	2,238.407
Distance-weighted population density in 1921	610	0.011	0.011	0.002	0.006	0.017
Mean ruggedness	610	816.797	709.409	613.205	5.592	2,863.851
Mean depth-to-rock	610	3.046	4.000	1.381	1.000	5.000
Mean rainfall in the spring of 1921 (ln)	610	5.331	5.292	0.370	4.434	6.298
Mean temperature in 1921	610	11.951	12.644	3.972	-1.780	17.718

TFP has been estimated with the procedure described by Levinsohn and Petrin (2003) and then weighted (in logarithms) using the share of the firm employment with respect to the total by LLMA, reference year, and sector. Geological data has been collected at the level of the single municipality and subsequently aggregated using the share of municipality surface with respect to the total by LLMA.

Table 3: Correlation matrix

	Variable	1	2	3	4	5	6	7	8
1	Total factor productivity (ln)	1.000							
2	Population density (ln)	0.192	1.000						
3	Total employment density (ln)	0.464	0.908	1.000					
4	Manufacturing employment density (ln)	0.535	0.786	0.928	1.000				
5	Contiguous employment density	0.409	0.324	0.455	0.554	1.000			
6	Distance-weighted employment density	0.626	0.287	0.518	0.619	0.603	1.000		
7	Altitude (ln)	-0.210	-0.511	-0.509	-0.456	-0.201	-0.271	1.000	
8	Share of coastal surface	-0.270	0.254	0.108	-0.078	-0.409	-0.362	-0.320	1.000
9	Share of population with a diploma	0.234	0.376	0.495	0.383	0.268	0.290	-0.078	0.073

The main TFP variable is the one estimated with the cost of labor. Both contiguous and distance-weighted densities are those computed with the total employment.

Table 4: IV estimation for each group of spatial controls and corresponding OLS estimation

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Estimator	IV	IV	IV	OLS	OLS	OLS
Employment density (ln)	0.065***	0.059***	0.062***	0.081***	0.078***	0.080***
	(0.008)	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)
Contiguous employment density	0.004***	0.004***	0.004***	0.004**	0.003**	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Spatial controls	Macro-areas	Regions	Provinces	Macro-areas	Regions	Provinces
Observations	610	610	610	610	610	610
Partial adjusted R-squared (density)	0.673	0.663	0.537			
Partial adjusted R-squared (contiguous density)	0.871	0.878	0.854			
Minimum eigenvalue statistic	157.423	153.024	103.421			
Overidentifying restrictions test (p-value)	0.001	0.335	0.242			
Endogeneity test (p-value)	0.002	0.000	0.003			
Adjusted R-squared				0.652	0.677	0.679

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models (from 1 to 3) are computed with the two-stage least squares (2SLS) estimator whereas the ordinary least squares (OLS) models (from 4 to 6) are computed with a robust variance estimator clustered at the level of the spatial controls (i.e., macro-areas, regions, and provinces). The total employment densities of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 5: IV estimation with quadratic term

Model	**(1)**	**(2)**	**(3)**
Estimator	IV	IV	IV
Employment density (ln)	0.130***	0.116***	0.146***
	(0.033)	(0.033)	(0.034)
Squared employment density (ln)	-0.009**	-0.008*	-0.013**
	(0.005)	(0.005)	(0.005)
Contiguous employment density	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
Spatial controls	Macro-areas	Regions	Provinces
Observations	610	610	610
Partial adjusted R-squared (density)	0.498	0.486	0.450
Partial adjusted R-squared (squared density)	0.528	0.524	0.487
Partial adjusted R-squared (contiguous density)	0.873	0.882	0.858
Minimum eigenvalue statistic	67.352	64.428	65.343
Overidentifying restrictions test (p-value)	0.001	0.391	0.264
Endogeneity test (p-value)	0.020	0.007	0.043

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The total employment densities of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities (measured in 1921). The squared total employment density is instrumented with the square of the linear prediction of $\log(L_r/S_r)$ regressed on all exogenous variables as in Wooldridge (2010). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 6: IV estimation with distance-weighted employment density

Model	**(1)**	**(2)**	**(3)**
Estimator	IV	IV	IV
Employment density (ln)	0.073***	0.067***	0.066***
	(0.008)	(0.008)	(0.009)
Distance-weighted employment density	7.172	10.982	18.208
	(5.243)	(7.152)	(11.266)
Spatial controls	Macro-areas	Regions	Provinces
Observations	610	610	610
Partial adjusted R-squared (density)	0.671	0.671	0.546
Partial adjusted R-squared (distance-weighted density)	0.940	0.954	0.918
Minimum eigenvalue statistic	155.562	158.015	106.227
Overidentifying restrictions test (p-value)	0.002	0.438	0.521
Endogeneity test (p-value)	0.000	0.000	0.007

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The total employment density and the distance-weighted total employment density of the focal and other LLMAs (measured in 2001) are instrumented with the corresponding lagged population density and distance-weighted population density (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 7: IV estimation with alternative definitions of density

Model	**(1)**	**(2)**	**(3)**	**(4)**
Estimator	IV	IV	IV	IV
Population density (ln)	0.069∗∗∗			
	(0.009)			
Contiguous population density	0.003∗∗∗			
	(0.001)			
Manufacturing employment density (ln)		0.047∗∗∗		
		(0.008)		
Contiguous manufacturing employment density		0.006∗∗∗		
		(0.002)		
Total employment (ln)			0.062∗∗∗	
			(0.006)	
Contiguous total employment			-0.000	
			(0.000)	
Local units (ln)				0.104∗∗∗
				(0.015)
Contiguous local units				-0.000∗∗
				(0.000)
Spatial controls	Regions	Regions	Regions	Regions
Observations	610	610	610	610
Partial adjusted R-squared (density)	0.759	0.460	0.777	0.142
Partial adjusted R-squared (contiguous density)	0.979	0.504	0.791	0.018
Minimum eigenvalue statistic	241.022	63.080	247.366	5.052
Overidentifying restrictions test (p-value)	0.326	0.311	0.377	0.287
Endogeneity test (p-value)	0.009	0.000	0.000	0.058

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The density of contiguous LLMAs has been computed with the logarithm of the population density, the logarithm of the manufacturing employment density, the logarithm of the total employment, and the logarithm of local units in models (1), (2), (3), and (4), respectively. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The densities and levels of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities and levels (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 8: IV estimation with alternative measures of TFP

Model	**(1)**	**(2)**
Estimator	IV	IV
Employment density (ln)	0.084***	0.106***
	(0.012)	(0.012)
Contiguous employment density	0.005***	0.007***
	(0.001)	(0.001)
Labor input	Number of white and blue collars	Number of employees
Observations	610	610
Partial adjusted R-squared (density)	0.651	0.651
Partial adjusted R-squared (contiguous density)	0.901	0.901
Minimum eigenvalue statistic	146.171	146.171
Overidentifying restrictions test (p-value)	0.423	0.274
Endogeneity test (p-value)	0.000	0.000

The dependent variable is the logarithm of the total factor productivity (TFP) estimated with the number of white and blue collars and the number of employees in models (1) and (2), respectively. All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The densities of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 9: IV estimation with both alternative measures of TFP and education controls

Model	**(1)**	**(2)**	**(3)**
Estimator	IV	IV	IV
Employment density (ln)	0.058***	0.075***	0.099***
	(0.009)	(0.013)	(0.014)
Contiguous employment density	0.004***	0.004***	0.006***
	(0.001)	(0.001)	(0.001)
Share of population with a diploma	0.037	0.408*	0.334
	(0.149)	(0.214)	(0.219)
Labor input	Cost of labor	Number of white and blue collars	Number of employees
Observations	610	610	610
Partial adjusted R-squared (density)	0.604	0.604	0.604
Partial adjusted R-squared (contiguous density)	0.919	0.919	0.919
Minimum eigenvalue statistic	120.205	120.205	120.205
Overidentifying restrictions test (p-value)	0.340	0.498	0.321
Endogeneity test (p-value)	0.000	0.000	0.000

The dependent variable is the logarithm of the total factor productivity (TFP) estimated with the cost of labor, the number of white and blue collars, and the number of employees in models (1), (2), and (3), respectively. All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The densities of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 10: IV estimation with interaction between density and macro-areas spatial controls and different sub-samples of LLMAs

Model	**(1)**	**(2)**	**(3)**
Estimator	IV	IV	IV
Employment density (ln)	0.065***	0.062***	0.055***
	(0.010)	(0.009)	(0.014)
Contiguous employment density	0.004***	0.003***	0.006***
	(0.001)	(0.001)	(0.002)
Employment density (ln) × South	0.008		
	(0.012)		
South	-0.238***		
	(0.040)		
Sample	All LLMAs	LLMAs of North	LLMAs of South
Observations	610	329	281
Partial adjusted R-squared (density)	0.713	0.721	0.555
Partial adjusted R-squared (contiguous density)	0.865	0.922	0.836
Partial adjusted R-squared (interaction)	0.719		
Minimum eigenvalue statistic	0.000	110.919	49.767
Overidentifying restrictions test (p-value)	0.000	0.529	0.241
Endogeneity test (p-value)	0.007	0.064	0.002

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models are computed with the two-stage least squares (2SLS) estimator. The densities of both the focal and contiguous LLMAs (measured in 2001) are instrumented with the corresponding lagged population densities (measured in 1921). Standard errors are reported in parentheses. The partial adjusted R-squared of the first stages are reported to assess the relevance of the excluded exogenous variables. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan's chi-squared statistic is reported to test the hypothesis that the additional instruments are exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 11: Shapley decomposition of total factor productivity for each group of spatial controls

Regressor	**North-South**	**Macro-areas**	**Regions**	**Provinces**
Spatial controls	40.6%	42.8%	48.0%	53.6%
Employment density (ln)	31.7%	30.5%	27.7%	24.1%
Contiguous employment density	16.9%	16.3%	14.9%	13.4%
Share of coastal surface	6.8%	6.4%	5.8%	5.5%
Altitude (ln)	4.1%	4.0%	3.6%	3.4%
Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**

The Shapley value is computed with the results of the OLS regressions on the determinants of the logarithm of the total factor productivity (TFP).

Table 12: Estimates of shift (A), dilation (D), and truncation (S)

Model	**(1)**	**(2)**	**(3)**	**(4)**
Constrained specification	No	Yes	Yes	Yes
Excluded factors	None	Truncation	Dilation	Truncation and dilation
Panel A: firms in northern versus southern LLMAs				
Relative shift (A)	0.181∗∗∗	0.182∗∗∗	0.191∗∗∗	0.195∗∗∗
	0.002	0.002	0.003	0.002
Relative dilation (D)	0.938∗∗∗	0.947∗∗∗		
	0.014	0.006		
Relative truncation (S)	-0.001		0.002∗	
	0.002		0.001	
Pseudo R-squared	0.978	0.975	0.963	0.960
Number of firms in northern LLMAs	148,994	148,994	148,994	148,994
Number of firms in southern LLMAs	39,123	39,123	39,123	39,123
Total number of firms	188,117	188,117	188,117	188,117
Panel B: firms in agglomerated versus non-agglomerated LLMAs				
Relative shift (A)	0.087∗∗∗	0.087∗∗∗	0.093∗∗∗	0.096∗∗∗
	0.002	0.002	0.002	0.002
Relative dilation (D)	0.942∗∗∗	0.947∗∗∗		
	0.007	0.005		
Relative truncation (S)	-0.000		0.001	
	0.000		0.001	
Pseudo R-squared	0.970	0.964	0.915	0.905
Number of firms in agglomerated LLMAs	139,705	139,705	139,705	139,705
Number of firms in non-agglomerated LLMAs	48,412	48,412	48,412	48,412
Total number of firms	188,117	188,117	188,117	188,117

Bootstrap standard errors are reported in parentheses, the number of replications is 200. The baseline model (1) is constrained to ignore truncation effects ($S = 0$) in model (2), dilation effects ($D = 1$) in model (3), or both in model (4). Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Appendix A. Aggregation of TFP

Assume the following production function:

$$Y_r = A_r\, K_r^{\alpha}\, L_r^{1-\alpha} = A_r \left(\frac{K_r}{L_r}\right)^{\alpha} L_r \tag{A.1}$$

where Y_r is the aggregated output produced in the area r, and A_r is the TFP of aggregated production of the area. Note that sector and year subscripts are dropped for the sake of notation simplicity. L_r and K_r are the amount of labor and capital inputs by all the firms belonging to area r, i.e., $L_r = \sum_{i\in r} l_i$ and $K_r = \sum_{i\in r} k_i$. Assuming constant returns to scale, α stands for the production function parameter. Y_r can be also defined as:

$$Y_r = \sum_{i\in r} y_i = \sum_{i\in r} A_i\, k_i^{\alpha}\, l_i^{1-\alpha} = \sum_{i\in r} A_i \left(\frac{k_i}{l_i}\right)^{\alpha} l_i \tag{A.2}$$

By equating Eq. (A.1) with Eq. (A.2), and considered that profit maximization requires that $k_i/l_i = K_r/L_r \;\forall i$, we obtain:

$$A_r \left(\frac{K_{r,t}}{L_{r,t}}\right)^{\alpha} L_r = \sum_{i\in r} A_i \left(\frac{k_i}{l_i}\right)^{\alpha} l_i \;\rightarrow\; A_r = \sum_{i\in r} \frac{l_{i,t}}{L_{r,t}} A_i \tag{A.3}$$

That is, under constant returns to scale, the TFP of aggregated production for the area coincides with the sum of individual TFPs computed at the firm level and weighted by their share of labor input.

Appendix B. Tables and figures

Table B.13: List of manufacturing industries and corresponding two-digit ATECO classification

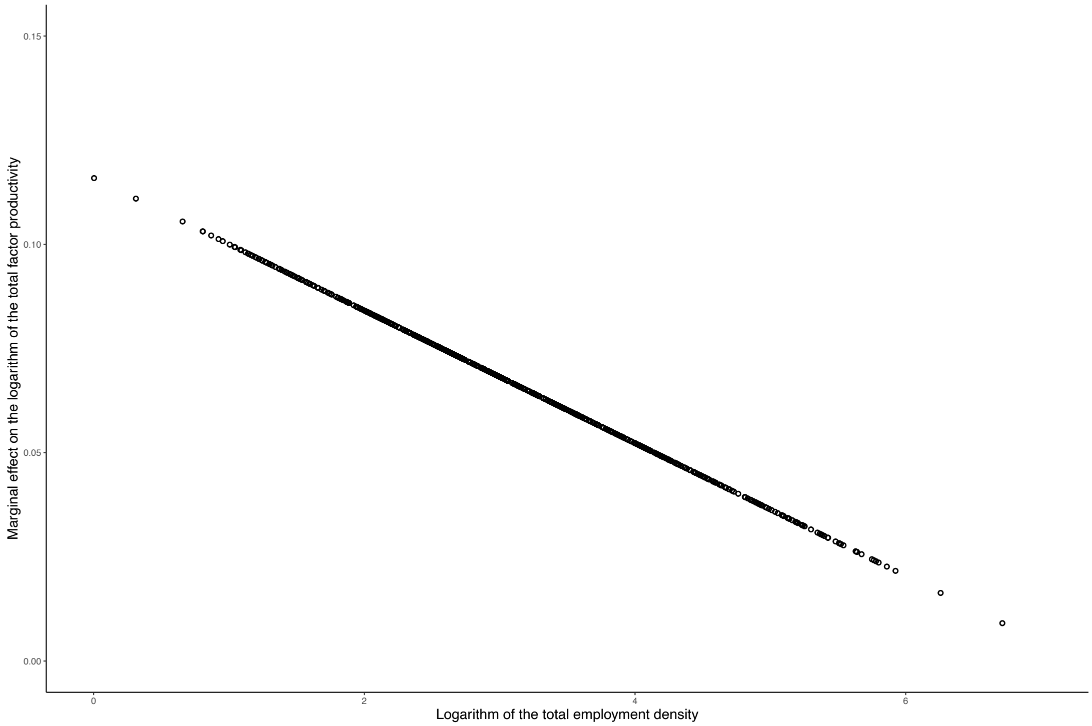
Section	**Division**
CA	Manufacture of food products (10)
CA	Manufacture of beverages (11)
CA	Manufacture of tobacco products (12)
CB	Manufacture of textiles (13)
CB	Manufacture of wearing apparel (14)
CB	Manufacture of leather and related products (15)
CC	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials (16)
CC	Manufacture of paper and paper products (17)
CC	Printing and reproduction of recorded media (18)
CE	Manufacture of chemicals and chemical products (20)
CG	Manufacture of rubber and plastic products (22)
CG	Manufacture of other non-metallic mineral products (23)
CH	Manufacture of basic metals (24)
CH	Manufacture of fabricated metal products, except machinery and equipment (25)
CI	Manufacture of computer, electronic and optical products (26)
CJ	Manufacture of electrical equipment (27)
CK	Manufacture of machineryand equipment not elsewhere classified (28)
CL	Manufacture of motor vehicles, trailers and semi-trailers (29)
CL	Manufacture of other transport equipment (30)

Table B.14: estimation without contiguous employment density

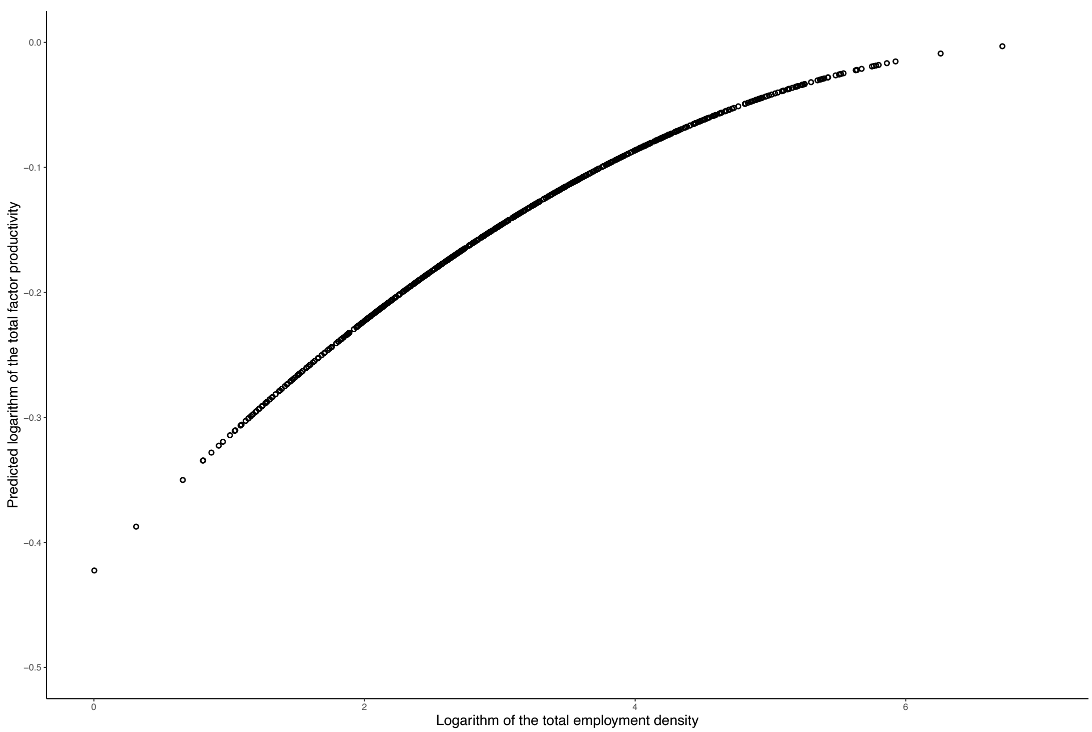
Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Estimator	IV	IV	IV	OLS	OLS	OLS
Employment density (ln)	0.077***	0.070***	0.071***	0.090***	0.086***	0.087***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.006)	(0.008)
Spatial controls	Macro-areas	Regions	Provinces	Macro-areas	Regions	Provinces
Observations	610	610	610	610	610	610
Partial R-squared (density)	0.695	0.694	0.653			
F statistic	193.972	188.354	132.194			
Overidentifying restrictions test (p-value)	0.001	0.307	0.516			
Endogeneity test (p-value)	0.001	0.000	0.003			
Adjusted R-squared				0.645	0.671	0.675

The dependent variable is the logarithm of the total factor productivity (TFP). All regressions include the constant term and controls for the share of coastal surface as well as the logarithm of the altitude of each LLMA. The instrumental variable (IV) models (from 1 to 3) are computed with the two-stage least squares (2SLS) estimator whereas the ordinary least squares (OLS) models (from 4 to 6) are computed with a robust variance estimator clustered at the level of the spatial controls (i.e., macro-areas, regions, and provinces). The total employment density of the focal LLMAs (measured in 2001) is instrumented with the corresponding lagged population density (measured in 1921). Standard errors are reported in parentheses. The partial R-squared and the F statistic of the first stage are reported to assess the relevance of the excluded exogenous variable. The p-value of the Durbin chi-squared statistic is reported to assess whether the logarithm of the total employment density can be considered as endogenous in the models. The p-value of the Sargan chi-squared statistic is reported to test the hypothesis that the additional instrument is exogenous. Stars from one to three indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Figure B.3: Marginal effects (a) and predicted total factor productivity (b)



(a) Marginal effects of (the logarithm of) total employment density



(b) Predicted total factor productivity as a function of (the logarithm of) density