BANCA D'ITALIA

EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

An anatomy of Italian cities: evidence from firm-level data

by Andrea Lamorgese and Andrea Petrella

October 2016

Number 362

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional papers)

An anatomy of Italian cities: evidence from firm-level data

by Andrea Lamorgese and Andrea Petrella

Number 362 – October 2016

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

AN ANATOMY OF ITALIAN CITIES: EVIDENCE FROM FIRM-LEVEL DATA

by Andrea R. Lamorgese* and Andrea Petrella*

Abstract

Economic activity is concentrated in urban areas and during the 20th century the urban population grew at a faster rate than the national average in many countries. Why is this the case? In this paper we focus on the urban premium in firms' productivity. We run two exercises. First, we document an urban premium in the level of firms' productivity and we inspect its determinants among both firm and city characteristics. Second, we corroborate the evidence on the determinants of the urban productivity gap, studying the heterogeneity in productivity across urban and non-urban areas by means of a Blinder–Oaxaca decomposition. The results point to a sizable urban productivity premium, mostly explained by firm size and by some characteristics of urban areas, in particular the average levels of educational attainment and labor market participation.

JEL Classification: D24, O47, R30.
Keywords: urban productivity premium, agglomeration, urban growth.

Contents

* Bank of Italy, Directorate General for Economics, Statistics and Research.

1 Introduction[1]

Two stylized facts emerge strongly from casual observation and the empirical literature in urban economics.

First, economic activity around the world concentrates in urban areas. Large US cities generated almost 85 percent of the country's GDP in 2010, compared with 78 percent for large cities in China, 76 percent in Latin America and just under 65 percent for those in Western Europe during the same period (Manyika et al., 2012).[2][3] The same report predicts that in the next 15 years, the 259 large US cities are expected to generate more than 10 percent of global GDP growth —a share bigger than that of all such cities in other developed countries combined.

The second notable stylized fact is than urban population has outgrown the national average in many countries during the XX century. Table 1 shows that the gap is large in US and Spain, less so in Italy and France.

Why is this the case?

In this paper we focus on productivity differences between urban and non urban areas as determinants of agglomeration and look into static productivity advantages of urban areas using a novel dataset on the universe of Italian firms.

We first establish whether Italian firms located in urban areas feature an advantage in the levels of productivity (measured as value added per employee) with respect to firms located outside urban areas. Our results point toward a sizable urban productivity advantage in the period between 2005 and 2013, both for manufacturing and (up to a larger extent) services firms. Using detailed data on the universe of Italian firms, we can control for a great deal of heterogeneity across firms; in so doing we verify that a large part of the urban productivity premium is explained by sector specific factors and firm size. Nevertheless, a sizable amount of the gap depends on local characteristic of the city, which we group into five sets: Urban economies, Size of the local market and demographic composition, Education, Mobility and Real Estate, and Labor market participation. For each of them we provide a measure of the contribution to the total urban productivity premium.

The local characteristics which are associated with higher levels of firms' productivity are the average level of education and of labor market participation in the area; besides these ones also the mobility within the area and the elasticity of housing supply favor higher productivity. Urban areas appear to be better endowed of such characteristics than non urban ones. A relevant contribution to explaining the urban productivity premium

[1]We wish to thank Antonio Accetturo, Vernon Henderson and Andrea Linarello for valuable advice, as well as seminar participants at the Bank of Italy. The usual disclaimer applies. The views expressed herein are those of the author and are not necessarily those of the Bank of Italy.

[2]Large cities are defined as those with 150,000 or more inhabitants in U.S.A (where they were 259 in 2010) and Western Europe (186); as those with more than 200,000 inhabitants in China (710) and Latin America (289).

[3]In Italy, the 12 largest cities (those with more than 200,000 inhabitants) together with their commuting zones produce 1/3 of the national value added; another third is produced in the 67 cities (plus their commuting zones) with population between 150,000 and 200,000 inhabitants. The remaining third is produced in the remaining (about 500) cities.

also comes from urban economies, here proxied by diversification in the area and average firm's sales, both higher in urban than in non urban areas.

To confirm this evidence, we then analyze heterogeneity in productivity levels across urban and non urban areas by the means of a Blinder–Oaxaca decomposition. This amounts to estimating the effect of groups of covariates on productivity in both urban and non urban areas and allows us to decompose the urban productivity gap into the contribution deriving from different endowments of the covariates (endowment effect), the one arising from heterogeneous elasticities across groups (coefficient effect), and a residual. Average productivity in manufacturing, net of sector and time fixed effect, is larger in urban than in non urban areas, and it is due more to differences in endowments than to differences in coefficients: firm's size, education and labor market characteristics come out of the picture as the main determinants of the urban productivity premium, whereas no difference in the coefficients comes out as prevalent. The picture is similar in services, where labor market participation does not weigh differently on productivity in urban and non urban areas, but part of the urban productivity premium is explained by the different elasticity of productivity to firm size across urban and non urban areas.

The paper is organized as follows. Section 2 introduces the ASIA dataset and the other data used in the paper. Section 3 analyzes the urban premium in productivity levels, and gives an assessment of the relative contribution of firms' and areas' characteristics. Section 4 provides additional evidence from an empirical exercise exploiting the Blinder–Oaxaca decomposition. Section 5 concludes.

2 Data

In this paper we use three sets of data: the definition of urban areas, which we build borrowing on the OECD–Eurostat definition and the map of Italian Local Labor Markets (henceforth, LLMs); firm level data on value added, sales and the number of workers; data on the characteristics of the Italian urban and non-urban areas. The next three subsections describe the three sources of data.

2.1 Italian urban areas

What is meant by city in the empirical urban literature is vague. Urban areas indeed do not necessarily overlap with the administrative borders of single municipalities and in general not even with those of the administrative units at a lower level of breakdown (NUTS3 regions, provinces in Italy). So urban areas end up being constituted by more than one municipality, but their territory does not coincide with the one of a province. In a word, cities are in between the level of aggregation of two administrative units, the municipality and the province.

A functional agglomeration which stays in between municipalities and provinces is the Local Labor Market (LLM), which is conventionally deemed as a good representation of a spatial agglomeration. LLMs are the result of a partition of the national territory made of subset of municipalities chosen in such a way that they contain both the place of residence

and the workplace of (a majority of) residents.[4]

The analyses in this paper use a definition of urban and non urban areas based on the OECD–Eurostat definition and adapted to keep into account the Italian Bureau of Census' (Istat) definition of commuting areas.

OECD–Eurostat methodology defines a urban area as a homogeneous set of areas whose density of population exceeds a certain threshold.[5] This definition is consistent with the traditional view that urban agglomerations are the places where production and knowledge spillovers take place, for the simple reason that density creates thick markets and favors the matching between demand and supply. It is therefore consistent with the traditional sources of agglomeration (*labor pooling*, *cost sharing* and *knowledge spillovers*), which are at the core of the birth of the industrial cities in XIX century.

This definition is also consistent with international standards set by the Urban Audit program, thus allowing international comparisons.

To define a urban area Eurostat performs a three-step procedure. First, it considers a partition of the territory of the European Union in a grid of 1 square km cells, select all those with a population density of at least 1,500 inhabitants per square km and clusters together —in what it calls an urban center— all the neighboring dense cells reaching a population of 50,000 inhabitants or more. Second, it considers the administrative borders and aggregate to the urban center all (LAU2) municipalities whose at least half of the population is resident in the urban center. The so formed agglomeration is candidate to belong to the urban agglomeration. This is defined such that: i) there is a administrative link; ii) half of the population of the urban area lives within one of the urban centers therein; iii) at least 75% of the population of the urban centers therein live in the urban agglomeration (fig. 1). The urban agglomeration thus defined is the metropolitan *core*.

Third, OECD-Eurostat defines an urban area as the union of the metropolitan core and its commuting area. Such urban area is called a Larger Urban Zone (LUZ).[6]

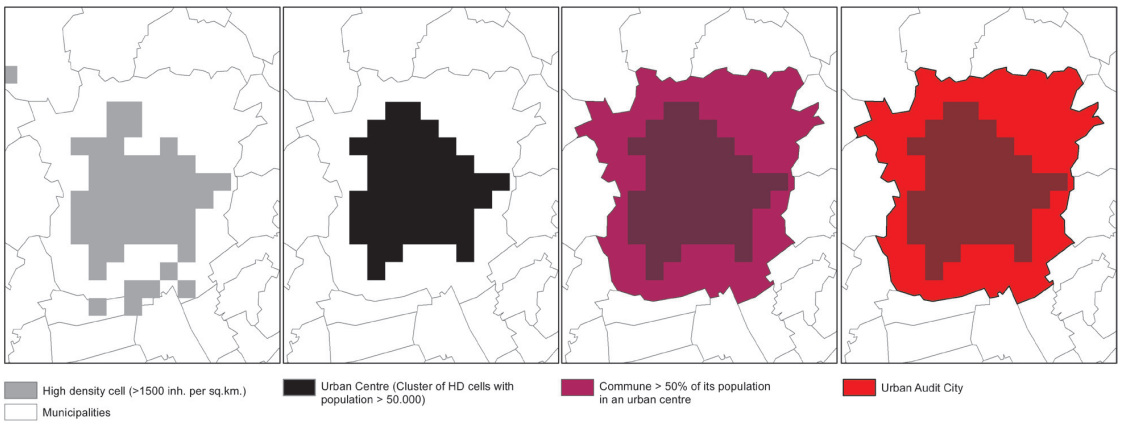
In this work we adopt the first two steps of the OECD-Eurostat definition, while we consider Italian LLM, rather than OECD-Eurostat LUZ in the third step, so that we define a urban area as an LLM containing a urban center as defined by OECD-Eurostat. This methodology identifies 73 urban areas (or urban LLMs) over a total number of 611 LLMs

[4] At the end of 2014 Istat issued the fourth classification of LLMs based on commuting flows of the 2011 Census (see `http://www.istat.it/it/strumenti/territorio-e-cartografia/sistemi-locali-del-lavoro`). The three classification before (corresponding to the commuting flows of the 1981, 1991, and 2001 Census) used the same kind of data but slightly different definitions. Starting in 2011 the definition has changed rather radically to be consistence with the European definition of LLM. For the sake of comparability, for 2001 the new and the old definition coexist.

[5] See `http://ec.europa.eu/eurostat/statistics-explained/index.php/European_cities_%E2%80%93_the_EU-OECD_functional_urban_area_definition`.

[6] The commuting area is similar to the LLM, with slightly different thresholds. Namely, OECD–Eurostat's commuting zones is constituted by all municipalities with at least 15% of residents who work in a neighboring municipality, such that the LUZ is continuous and self contained. A LLM is characterized by the fact that: i) people commute for work reasons; ii) LLMs are self contained (at least 75% of resident work within the LLM; 25% at most outside it); iii) municipalities within a LLM are contiguous (commuting takes place between contiguous municipalities, non contiguous ones are excluded); iv) the *core* of the LLM is the municipality toward which commuting flows are maximum.

Figure 1 Urban centers and urban cities



Source: Eurostat.
Note: Municipalities are white-colored; cells of the density grid that are denser than 1500 inhabitants per square km are gray-colored; urban centers (i.e. *cluster* of dense cells with more that 50,000 inhabitants) are black-colored; the sets of municipalities that contain the urban center are crimson-colored; urban areas are red-colored.

in 2011. Non urban areas are the remaining LLMs. While the 73 urban areas have grown both in population and size, having absorbed an increasing number of municipalities, non urban areas have diminished over time from 880, to 710 to 612 and finally to 538 over the 4 Censuses between 1981 and 2011 (table 2).

2.2 Firm-level data

The main source of the data used in this paper is the *Archivio statistico delle imprese attive* (ASIA).[7]

Data refer to the population of Italian firms of the private sector excluding those operating in agriculture and in the financial sector for the years 2005, 2008 and 2011–13. The sectors for which information are gathered are: Manufacturing; Construction; Retail and wholesale trade; Transportation and storage (Logistics); Accommodation and food service activities; ICT; Real Estate; Professional, scientific and technical activities; Travel agency, rental, cleaning, security and other administrative and support services. Sector information is provided at the 5 digits breakdown.

For each firm the dataset gathers information about the year of birth, the sector the firm belongs to, the number of employees, the sales and the value added.

For the population of Italian firms contained in the ASIA archives, the Italian national institute of statistics computes the value added using different sources: the SCI (*Sistema dei Conti delle Imprese*) for all the firms with more than 100 employees; balance sheets

[7]Data have been made available for elaboration at the former school SAES-Istat within the agreement Bank of Italy and Istat.

data (coming from *Camere di Commercio e Cerved*) for incorporated firms with less than 100 employees; taxes data (from *Agenzia delle Entrate*) for all other firms.

With these data sources it is possible to recover the value added for more than 80% of firms operating in Italy between 2005 and 2008. For the remaining ones the value is imputed computing the median value added per employee in the same region, the same size class (among 5 different ones) and the same sector code (5 digit Ateco classification).[8]

Starting from 2011 the share of firms for which value added can be directly computed is larger than 95%; starting from 2012 the database corresponds to the data of the FRAME-SBS archives, which are also the microeconomic information support for the national accounts.

The dynamics of value added is deflated using cost deflator for the factors used in each 2 digit sector of economic activity (in the 2010 base year). Labor productivity is measured as value added per worker.

On aggregate, the ASIA microdata are consistent with Italian national accounts (CN) data and with the Structural business statistics (SBS) from Eurostat.

In manufacturing the dynamics of VA coming from the aggregation of ASIA microdata closely tracks the dynamics of national accounts, while they slightly differ from SBS, in particular between 2008 and 2011. In the services there is a larger difference among between the dynamics of ASIA, on the one side, and those of CN and SBS on the other one. The difference is particularly accentuated with respect to CN (which takes into account the black market) and in the period between 2011 and 2012, when both SBS and ASIA show a drop in productivity of about 5% (only 1.8% for CN).

Table 3 shows some descriptive statistics computed on the ASIA database. The number of firms in the whole economy has been rising before the crisis (up to 2008) and steadily decreasing afterwards, both in urban and non urban areas. This dynamics is due to manufacturing —where a widespread process of restructuring, churning, and selection had been going on since the early 90's— while the number of firms in services shows smaller fluctuations over the period of observation. There is no much heterogeneity in such dynamics across urban and non urban areas.

Between the first and the last year of observation, the average firm size has increased across the board, steadily in the services, with a hump shape in manufacturing. Urban firms are larger than non urban ones, during the whole period.

Labor productivity increases during the reference period in manufacturing (cumulatively and year on year, but in the aftermath of the sovereign debt crisis in 2012), while decreasing in services. In manufacturing, the growth of productivity reflected a growth of value added larger than the growth of the number of workers in the period 2005-2008, and a larger drop in the number of workers than in value added afterwards. In the services, instead the dynamic of productivity reflects a steady expansion of the number of workers against a drop in value added.

An interesting heterogeneity between urban and non urban areas is apparent. Productivity in manufacturing grows more in non urban areas than in urban areas (11.9% vs.

[8]Imputation of missing values is called for in order to be consistent with aggregate data on value added. Dropping these values doesn't change the estimates contained in this paper.

7.8% cumulate over 2005–2013), while in services the drop is three times as large in non urban areas than in urban ones (-18.4% vs -6.0% cumulate over the same period). The big difference across the two sectors is in the year-on-year dynamics: in manufacturing the dynamics is qualitatively similar across urban and non urban areas, but the magnitudes are larger in non urban areas; instead in services productivity was already dropping in non urban areas before the crisis and continued to drop during it, whereas in urban areas it dropped only in the period 2008–2012, showing some recovering in 2013.

2.3 Data on areas' characteristics

The data on areas' characteristics come from the dataset 8milacensus.[9] We have a very rich set of area-specific controls, which we distinguish in 5 subsets:

i. **Urban economies**, which contains variables proxying productive amenities of the area, as sector composition (shares of real value added) of the area, average firm size, specialization (Herfindahl–Hirschman index of real sales).

ii. **Size of the local market and demographic composition**, which contains population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

iii. **Education**, which contains gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range.

iv. **Mobility and Real Estate**, which contains variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (iii) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility).

v. **Labor market participation**.

Data are provided for each of the 7 census years in the time range 1951–2011. We take values in 2001, so to balance the need to capture areas' characteristics which are still actual with the technical requirement that they should be reasonably predetermined with respect to the dependent variable, which ranges in the interval 2005-2013.

3 The urban productivity premium

In this section we want to document whether firms located in urban areas consistently display larger levels of labor productivity than firms located in non urban areas.

[9] `http://www.8milacensus.it`.

Obviously, part of this gap does depend on observable characteristics of the firms and of the places where the firms are located. That is, firms operating in some sectors might be consistently more productive than firms operating in some other sectors and these firms might be located in urban areas, so that a large part of the urban productivity premium might be explained by the sector composition rather than by any deeper economic mechanism. Analogously, larger firms, which are notoriously more productive, might be located in urban areas, so that composition by size might be the only reason why urban areas are more productive.

On the one hand we net these spurious effects to obtain a measure of the urban productivity premium that conditions out firms' and areas' specificities; on the other hand, we compute the contribution to the unconditional urban productivity premium of a wide range of firms and areas' characteristics.

Operatively, we estimate the following regression model:

$$y_{isc,t} = \alpha + \beta I_{\text{Urban}} + \gamma_i X_{i,t-s} + \gamma_c X_{c,t-s} + \delta_1 I_t + \delta_1 I_s + \varepsilon_{isc,t}, \tag{1}$$

where $y_{isc,t}$ is the value added per worker in firm i which at time t operates in sector s and is located in the area c. I_{Urban} is a dummy which takes value 1 if the firm is located in one of the 73 Italian urban areas which we have defined above, $X_{i,t-s}$ and $X_{c,t-s}$are a set of firm- and area- specific controls, lagged s periods, so to be reasonably predetermined with respect to time t variables. I_t and I_s are respectively time and sector fixed effects, where sector fixed effects are at a very disaggregated breakdown (5 digits of the Ateco classification), so to take into account a large chunk of sectoral idiosyncratic specificities.

While we have a small set of firm specific controls, basically only lagged sales and the sector fixed effect, we have a very rich set of area specific controls distinguished in the 5 subsets: Urban economies, Size of the local market and demographic composition, Education, Mobility and Real Estate, and Labor market participation.

We start estimating a very parsimonious version of model (1), in which we only include the urban area dummy and time fixed effects. This amounts to computing an unconditional average of the gap between labor productivity in firms located in urban areas and firms located elsewhere, only controlling for time-varying idiosyncratic shocks which are common to all firms in the sample.

Between 2005 and 2013 firms located in the urban areas of the country show a premium in the levels of labor productivity of more than 7.7 percentage points in manufacturing and 15.6 in services with respect to non urban firms, controlling for time fixed effects (tables 4 and 5).

Firm's characteristics, like the level of sales in the previous period and the sector fixed effect, explain most of the differences in productivity premium across manufacturing firms and services ones: conditioning them out, the urban productivity premium decreases to about 3% both in services and in manufacturing (although heterogeneity is such that the premium is estimated with a large standard error). Net of sector heterogeneity, larger firms —that is firms with larger sales in the previous period—[10] have a higher level of

[10] As a robustness check also the number of workers in the previous period was considered, without any significant change in the estimate.

productivity.

Adding area's characteristics among the regressors increases the precision of the estimates of the urban productivity premium. In general, the features of the area explain a larger part of this premium in the services sector than in manufacturing.

Proxies for urban economies, that is proxies for the strategic interactions of the economic environment in which firms operate, like the specialization of the area, its sectoral composition, or the average size (in term of sales) of the firms which are thereby located influence firm's productivity in the expected way: the latter is larger in areas where specialization is smaller (that is, in a more diverse environment) and where the average firm's size is larger. Diversity (i.e. lower specialization) seems to be more important for manufacturing than for services, though.

The level of education explains a larger share of the urban productivity premium in manufacturing than in services (column (5)). The effect on the urban productivity premium of better mobility and a more responsive supply of residential buildings are sizable in both sectors (column (6)).

Labor market participation increases productivity more in manufacturing than in services, and explain a larger share of the urban productivity premium in the former than in the latter (column (7)).

Demographic decomposition of the population (elders' dependency ratio, share of foreign resident, foreign residential mobility, foreign school attendance) and proxies for the size of the local market (population and density) contribute negatively to the urban productivity premium, more in manufacturing than in services (column (4)).

All in all, larger and more productive firms localize in denser, less specialized areas, where the average firm dimension is larger, higher skilled people gather, commuting is easier, the supply of residential property is more elastic and the participation to the labor market is larger.

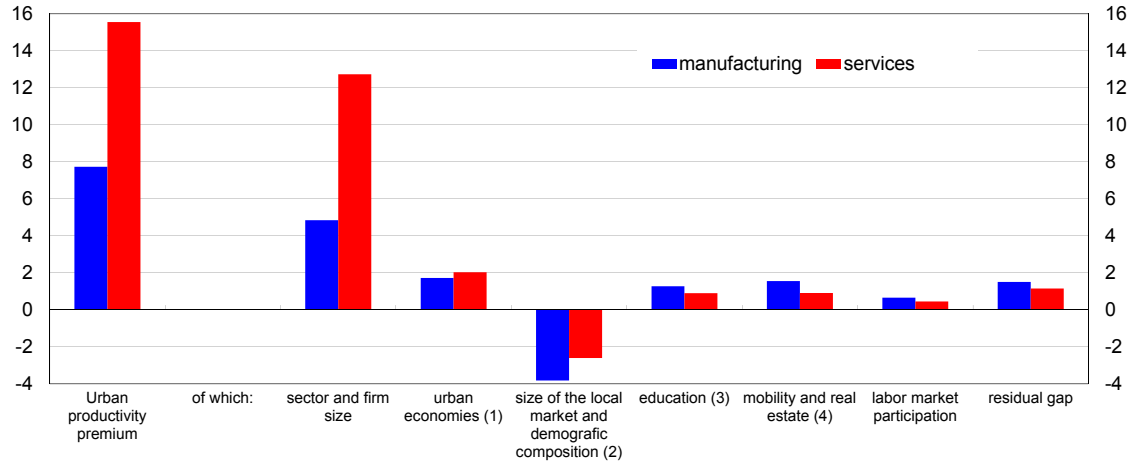
Figure 2 provides a graphical description of the contribution of each set of firms' and areas' controls by depicting the changes in the coefficient β in equation (1) passing from leaner to richer specifications.

3.1 Robustness

The number of areas' controls we have introduced in the previous exercise is large and most of the characteristics have little variability over time and risk to be a catch-all of a number of underlying features linked to other local features. Even if they have been selected from a wider set of variables trying to minimize multicollinearity problems, some of them remain correlated and such correlation may bias estimates of the coefficients of the controls (γ_i and γ_c), but potentially also of the coefficient of interest (β).

In order to check the robustness of the baseline estimates, in this section we adopt a standard technique and replace each group of controls (i.e. Size of the local market and demographic composition, Education, Mobility and Real Estate) with the its first principal component, while keeping own sales and the proxies for Urban economies as

Figure 2 Contributions to the urban productivity premium



Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA)
Notes: (1) Proxied by sector composition (shares of real value added) of the area, average firm sales, specialization (Herfindahl–Hirschman index of real sales). (2) Proxied by population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners. (3) Proxied by the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (4) Contains variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area, increase in residential buildings, residential mobility).

distinct regressors.[11]

Results are shown in tables 6 and 7. The results of the baseline estimation are by and large confirmed, although in column (3) the estimate of the urban productivity premium (β) is larger in magnitude when urban economies are included in the regression as principal components than in the case in which sector composition (shares of real value added) of the area, average sales in the area, and specialization (Herfindahl–Hirschman index of real sales) enter as distinct regressors. The coefficients of the other variables of interest (Sales, Specialization and Area's sales) are very similar across the two estimated models, even if specialization turns out to be less stable across the different specifications.

The estimated model with principal components also allows to assess the effect of groups of area's controls on local productivity, whereas in the baseline the number of regressors and the within group heterogeneity are large and make interpretation of the coefficients cumbersome.

The proxy for a higher level of education (that is a lower gender gap in high school attainment, a lower illiterate ratio, a lower level of school dropouts, a larger share of the

[11]The fifth control for the characteristics of the area is Labor market participation, which is a single control that already summarizes the features of the local labor market, since it is highly collinear with the other variables related to the local labor market that are available in the dataset.

population with high school attainment, larger incidence of education among youngsters) stands out showing the higher correlation with the levels of productivity. Manufacturing firms enjoy a larger effect on productivity of being located in an environment featuring higher levels of education than services firms.

Demographic composition of the population (elders' dependency ration, share of foreign resident, foreign residential mobility, foreign school attendance) and proxies for the size of the local market (population and density) affect the levels of productivity similarly in services and in manufacturing. Finally, better mobility and a more responsive supply of residential buildings increase productivity, more for services than for manufacturing firms. The estimates of both proxies are sometimes imprecise across specification.

Correlation between this characteristics of the area and the levels of productivity are in general stronger in urban than in non urban areas (table 8).

4 A Blinder–Oaxaca decomposition

The previous section evaluates the effects of covariates on the urban productivity premium by analyzing the stability of the estimates of the coefficient β in regression (1) as long as an increasing set of controls are taken into account. The advantage of such a design is that it is simple, but it has two main drawbacks. First, it relies on the assumption that the elasticities of labor productivity with respect to the different covariates are the same across urban and non urban areas.[12] This means that all the productivity differences boil down to the effect of heterogeneity in the covariates ("endowment effect") and in differences in the intercept. The second drawback is that the results depend on the order in which the different set of covariates enter the regression. Obviously some robustness check, changing such order, has been performed, and results end up being qualitatively similar, but the order in which covariates are taken into account remains discretionary.

One natural way to overcome these drawbacks and test for the robustness of the empirical strategy adopted so far is to perform a Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973).

The Oaxaca-Blinder decomposition is used to study heterogeneity in outcomes (typically labor-market ones) by groups (sex, race, and so on) by decomposing mean differences in the log of the outcome based on linear regression models in a counterfactual manner. It divides the outcome differential (i.e. wage differential) between two groups into a part that is "explained" by group differences in some characteristics, which are typically deemed to be determinants of the outcome (i.e. education or work experience), and a residual part that cannot be accounted for by such differences in determinants of the outcome. The "explained" part can be in turn decomposed in a component accounting for the heterogeneity across groups in the endowment of the characteristics and the heterogeneity across groups in the response of the outcome to those characteristics. The "unexplained" part is often used as a measure for discrimination, but it also subsumes the effects of group differences in unobserved predictors.

[12] Adding interaction of covariates with the urban area dummy variable is not viable, since the interacted regressors ends up to be highly collinear with the original covariate.

In our case, the Oaxaca-Blinder decomposition allows to recover the effects of groups of covariates on the urban productivity gap. With respect to the simpler exercise in the previous sections the Blinder–Oaxaca decomposition does not rely on any ordering of the covariates and provides a direct estimation of their contribution on the gap in productivity across the two groups, that is urban and non urban areas.

Operatively, the Oaxaca–Blinder decomposition takes the mean outcome difference

$$R = E(Y_{NUA}) - E(Y_{UA})$$

and decomposes it based on a set of predictors X and a linear model

$$Y_l = X_l'\beta_l + \varepsilon_l, \quad E(\varepsilon_l) = 0 \;\; l = \{NUA, UA\}$$

into

$$\begin{aligned} R =& E(Y_{NUA}) - E(Y_{UA}) = E(X_{NUA})'\beta_{NUA} - E(X_{UA})'\beta_{UA} \\ =& \left[E(X_{NUA}) - E(X_{UA})\right]' \beta_{UA} + E(X_{UA})' \left(\beta_{NUA} - \beta_{UA}\right) \\ &+ \left[E(X_{NUA}) - E(X_{UA})\right]' \left(\beta_{NUA} - \beta_{UA}\right) \\ =& E + C + I. \end{aligned} \tag{2}$$

The first component

$$E = \left[E(X_{NUA}) - E(X_{UA})\right]' \beta_{UA}$$

the so called "endowment effects" amounts to the part of productivity differential that is due to group differences in the predictors, that is to differences in the firms' and areas' characteristics across urban and non urban areas. More specifically, the E component measures the expected change in the urban area's mean productivity if urban areas had the same levels of covariates of non urban areas.

The second component

$$C = E(X_{UA})' \left(\beta_{NUA} - \beta_{UA}\right)$$

the so called "coefficient effects" measures the contribution of differences in the coefficients (including differences in the intercepts), and describes the different returns of the local characteristics. Component C measures the expected change in the urban area's mean productivity if urban areas firms reacted to local characteristics with the same elasticities measured by non urban areas' coefficients.

The third component

$$I = \left[E(X_{NUA}) - E(X_{UA})\right]' \left(\beta_{NUA} - \beta_{UA}\right)$$

is the interaction term accounting for the fact that differences in endowments and coefficients exist simultaneously between the two groups. Potential nonlinearities and omitted variables are also captured by this term.

Performing the Oaxaca–Blinder decomposition with our very rich set of sector controls (with a 5 digits breakdown) is challenging from a computational point of view. We therefore proceed in two different ways.

In a first exercise, we condition out time and sector fixed effect in a preliminary step and then run the Blinder–Oaxaca decomposition on the residuals separately for manufacturing and services firms. In a second exercise we run the Oaxaca–Blinder decomposition in a single step, using sector controls with a two digits breakdown. The drawback of the first exercise is that when conditioning on time and sector fixed effects we are taking out sectorXyear averages that might be different across the urban and non urban areas. The drawback of the second exercise is that we are controlling less precisely for sector heterogeneity.

Tables 9 and 10 present results of the Blinder–Oaxaca decomposition on filtered data for manufacturing and services firms respectively.

Average productivity in manufacturing, net of sector and time fixed effect, is larger in urban than in non urban areas. The largest share of the premium is explained by differences in the local endowments of the different observed characteristics (table 9). The contribution is not the same across different sets of covariates: firm's size, education and labor market characteristics come out of the picture as the main determinants of the urban productivity premium. The amount of premium which is explained by the observables (5.3%) is roughly comparable to the coefficient of the I_{Urban} dummy in equation (1), reported in Table 4 for manufacturing. This is further decomposed in the contribution of endowments, of coefficients and of the interaction terms. Differences in the endowment of observables has a greater relevance in explaining the urban productivity premium (3.8%), whereas none of the coefficients stands out as prevalent, even if collectively their differences across urban and non urban areas explain almost two fifths of the premium.

Average productivity in services, net of sector and time fixed effect, is around 8 percentage larger in urban than in non urban areas and the gap is primarily due to differences in endowments, only partly to differences in coefficients (table 10). Firm's size and education are the main determinants of the urban productivity premium and the elasticity of the productivity to firm's size is also larger in urban areas; firm's size raise the urban productivity premium both through an endowment and a coefficient effect. Finally, being in an area which is specialized in accommodation and food services decreases the urban productivity premium.

Tables 11 and 12 describe results coming from the one step Blinder–Oaxaca decomposition with two digits sector controls. They by and large confirm results from the two steps decomposition with finer (5 digits) sector controls; in services the less tight control for sector idiosyncratic effects shows up in some significant endowment, coefficients and interaction effects for sectors in sections J (Financial activities) and F (Construction), as well as in interaction effects for education and the composition of population

5 Concluding remarks

This paper provides a few stylized facts about labor productivity of firms in Italian urban and non urban areas. We first establish that firms located in urban areas experience a sizable productivity premium with respect to the ones located in non urban areas. Then we describe what are the main drivers of productivity in urban and non urban areas, trying

to single out the ones that explain the greatest share of the urban productivity premium. We verify that a large part of the urban productivity premium is explained by sector-specific factors and firm size. Nevertheless, a sizable amount of the gap depends on local characteristic of the city, mainly connected to the education levels of the population and labor market participation.

The Blinder–Oaxaca decomposition allows us to disentangle the part of the premium deriving from different endowments of the covariates (endowment effect), and the one arising from heterogeneous elasticities across urban and non-urban areas (coefficient effect). We show that the observed productivity premium is due more to differences in endowments than to differences in coefficients: firm's size, education and labor market characteristics emerge as the main determinants of the urban productivity premium, whereas no difference in the coefficients comes out as prevalent.

References

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8:436–455.

Duranton, G. and Puga, D. (2014). The Growth of Cities. In *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, chapter 5, pages 781–853. Elsevier.

Giffoni, F., Gomellini, M., and Pellegrino, D. (2016). Human capital and urban growth. Manuscripts.

Manyika, J., Remes, J., Dobbs, R., Orellana, J., and Schaer, F. (2012). Urban america: Us cities in the global economy. Technical report, McKinsey Global Institute.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(693–709).

Tables

Table 1 Population growth in urban areas in selected countries

	Growth rate of population (10-years averages)		
	time range	urban areas	national average
USA	1920–2010	17.9	12.6
Spain	1920–2010	18.1	8.9
France	1937–2007	7.7	5.7
Italy	1911–2001	7.2	4.9

Source: Giffoni et al. (2016), Duranton and Puga (2014).

Table 2 Evolution of Italian LLMs over time

Census year	Urban	Non Urban	Totale	LLM definition
1981	74	880	954	1981
1991	74	710	784	1991
2001	74	612	686	2001
2001	74	609	683	2011
2011	74	538	612	2011

Source: Istat, *Census* in 1981, 1991, 2001 e 2011.

Table 3 Descriptive statistics

	Non-urban			Urban		
Year	# firms	# workers per firm	VA per worker	# firms	# workers per firm	VA per worker
			Manufacturing			
2005	225,960	8.41	42,818	235,087	9.69	51,666
2008	219,459	8.83	45,617	225,000	10.13	54,691
2011	206,206	8.63	47,516	212,778	10.00	55,958
2012	204,726	8.62	46,639	209,798	9.98	55,122
2013	200,768	8.53	47,909	206,279	9.97	55,681
			Services			
2005	1,052,146	2.52	31,543	1,452,897	3.56	40,388
2008	1,084,041	2.69	28,819	1,490,655	3.78	41,746
2011	1,065,450	2.75	27,379	1,471,030	3.84	39,875
2012	1,069,544	2.75	25,909	1,473,835	3.86	37,813
2013	1,056,372	2.74	25,734	1,460,670	3.86	37,957
			Total economy			
2005	1,278,106	3.56	36,243	1,687,984	4.42	43,833
2008	1,303,500	3.72	35,530	1,715,655	4.62	45,472
2011	1,271,656	3.70	34,998	1,683,808	4.62	44,276
2012	1,274,270	3.70	33,680	1,683,633	4.62	42,473
2013	1,257,140	3.66	33,979	1,666,949	4.61	42,698

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).

Table 4 Determinants of the levels of productivity: Manufacturing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Urban areas	0.077*	0.029	0.012***	0.049***	0.037***	0.022***	0.015**
	[0.045]	[0.018]	[0.005]	[0.009]	[0.007]	[0.007]	[0.007]
Sales		0.265***	0.257***	0.255***	0.254***	0.254***	0.2539***
		[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Specialization			-0.087***	-0.230***	-0.154***	-0.176***	-0.145***
			[0.020]	[0.025]	[0.020]	[0.020]	[0.020]
Area's sales (1)			0.064***	0.074***	0.044***	0.049***	0.0373***
			[0.006]	[0.007]	[0.005]	[0.005]	[0.005]
Labor market participation							0.011***
							[0.001]
Controls for:							
Size of local mkt & demography (2)	no	no	no	yes	yes	yes	yes
Education (3)	no	no	no	no	yes	yes	yes
Mobility & RE (4)	no	no	no	no	no	yes	yes
Sectors (5)	no	yes	yes	yes	yes	yes	yes
Area's sector composition	no	no	yes	yes	yes	yes	yes
Obs.	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035
R^2	0.004	0.379	0.383	0.386	0.388	0.389	0.389

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. (1) Average firm sales in the LLM. (2) Population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners. (3) Gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15–19 age range. (4) Variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (5) Sectors' controls are at 5 digits according to the ATECO 2007 classification.

Table 5 Determinants of the levels of productivity: Services

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Urban areas	0.155***	0.028*	0.008***	0.034***	0.027***	0.016***	0.012**
	[0.036]	[0.015]	[0.003]	[0.007]	[0.005]	[0.005]	[0.005]
Sales		0.334***	0.329***	0.328***	0.327***	0.327***	0.326***
		[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Specialization			-0.016	-0.120***	-0.081***	-0.118***	-0.086***
			[0.023]	[0.019]	[0.018]	[0.017]	[0.015]
Area's sales (1)			0.067***	0.067***	0.049***	0.052***	0.042***
			[0.006]	[0.005]	[0.004]	[0.005]	[0.004]
Labor market participation							0.008***
							[0.001]
Controls for:							
Size of local mkt & demography (2)	no	no	no	yes	yes	yes	yes
Education (3)	no	no	no	no	yes	yes	yes
Mobility & RE (4)	no	no	no	no	no	yes	yes
Sectors (5)	no	yes	yes	yes	yes	yes	yes
Area's sector composition	no	no	yes	yes	yes	yes	yes
Obs.	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403
R^2	0.011	0.353	0.355	0.357	0.357	0.358	0.358

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. (1) Average firm sales in the LLM. (2) Population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners. (3) Gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15–19 age range. (4) Variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (5) Sectors' controls are at 5 digits according to the ATECO 2007 classification.

Table 6 Determinants of the levels of productivity: Manufacturing (principal components)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Urban areas	0.077*	0.029	0.035***	0.043***	0.035***	0.027***	0.018*
	[0.045]	[0.018]	[0.014]	[0.013]	[0.011]	[0.010]	[0.009]
Sales		0.265***	0.257***	0.257***	0.256***	0.256***	0.255***
		[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Specialization			-0.087***	-0.146***	-0.149***	-0.128***	-0.015
			[0.020]	[0.022]	[0.021]	[0.021]	[0.022]
Area's sales (1)			0.064***	0.076***	0.063***	0.059***	0.025***
			[0.006]	[0.006]	[0.005]	[0.005]	[0.006]
Education (2)					0.025***	0.025***	0.026***
					[0.002]	[0.002]	[0.002]
Mobility						0.017***	0.002
& RE (3)						[0.002]	[0.002]
Labor mkt							0.001***
participation							[0.001]
Controls for:							
Sectors (4)	no	yes	yes	yes	yes	yes	yes
Area's sector							
composition	no	no	yes	yes	yes	yes	yes
Size of local mkt							
& demography (5)	no	no	no	yes	yes	yes	yes
Obs.	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035	1,333,035
R^2	0.004	0.379	0.382	0.383	0.384	0.384	0.386

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. (1) Average firm sales in the LLM. (2) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15–19 age range. (3) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (4) Sectors' controls are at 5 digits according to the ATECO 2007 classification. (5) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

Table 7 Determinants of the levels of productivity: Services (principal components)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Urban areas	0.155***	0.028*	0.026***	0.031***	0.027***	0.023***	0.015**
	[0.036]	[0.015]	[0.010]	[0.009]	[0.008]	[0.007]	[0.007]
Sales		0.334***	0.329***	0.329***	0.328***	0.328***	0.328***
		[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Specialization			-0.016	-0.055**	-0.056***	-0.044**	0.051**
			[0.023]	[0.024]	[0.022]	[0.022]	[0.021]
Area's sales (1)			0.067***	0.075***	0.064***	0.062***	0.033***
			[0.006]	[0.006]	[0.006]	[0.006]	[0.005]
Education (3)					0.017***	0.015***	0.018***
					[0.002]	[0.002]	[0.002]
Mobility						0.011***	-0.000
& RE (3)						[0.002]	[0.002]
Labor mkt							0.008***
participation							[0.001]
Controls for:							
Sectors (4)	no	yes	yes	yes	yes	yes	yes
Area's sector							
composition	no	no	yes	yes	yes	yes	yes
Size of local mkt							
& demography (5)	no	no	no	yes	yes	yes	yes
Obs.	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403	7,562,403
R^2	0.011	0.353	0.355	0.355	0.356	0.356	0.357

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. (1) Average firm sales in the LLM. (2) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners. (3) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15–19 age range. (4) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (5) Sectors' controls are at 5 digits according to the ATECO 2007 classification.

Table 8 Determinants of the levels of productivity: interactions

	Manufacturing	Services
Urban areas	0.023**	0.019***
	[0.010]	[0.007]
Sales	0.255***	0.328***
	[0.002]	[0.004]
Specialization	-0.036	0.012
	[0.029]	[0.026]
Specialization X Urban areas	0.030	0.059
	[0.036]	[0.036]
Area's sales (1)	0.024***	0.031***
	[0.006]	[0.005]
Labor mkt participation	0.010***	0.008***
	[0.001]	[0.001]
Size of loc mkt & demography (2)	-0.012***	-0.010***
	[0.003]	[0.002]
Size of loc mkt & demography (3) X Urban areas	0.007	0.013**
	[0.005]	[0.006]
Education (3)	0.024***	0.015***
	[0.002]	[0.002]
Education (3) X Urban areas	0.006	0.006*
	[0.004]	[0.003]
Mobility & RE (4)	0.004	0.002
	[0.003]	[0.002]
Mobility & RE (4) X Urban areas	-0.003	-0.002
	[0.004]	[0.004]
Controls for:		
Sectors (5)	yes	yes
Area's sector composition	yes	yes
Obs.	1,333,035	7,562,403
R^2	0.386	0.357

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. (1) Average firm sales in the LLM. (2) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners. (3) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (4) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (5) Sectors' controls are at 5 digits according to the ATECO 2007 classification.

Table 9 Blinder-Oaxaca 2steps decomposition: Manufacturing

	average productivity	p-value		Observations
overall				1,333,035
non urban	0.021	0.028		648,660
urban	0.073	0.002		684,375
difference	-0.053	0.035		
endowments	-0.038	0.116		
coefficients	-0.019	0.098		
interaction	0.004	0.639		

	Endowments effect		Coefficient effect		Interaction	
	coeff.	p-value	coeff	p-value	coeff.	p-value
Sales	-0.024*	0.059	-0.000	0.740	0.000	0.740
Urban Economies (1)	-0.001	0.177	-0.000	0.859	0.000	0.858
Education (2)	-0.017**	0.003	-0.001	0.387	0.002	0.351
Mobility & RE (3)	-0.000	0.929	-0.001	0.456	0.002	0.334
Size of loc mkt & demography (4)	0.003	0.518	-0.003	0.355	0.006	0.342
Labor mkt participation	-0.011**	0.048	-0.000	0.871	0.000	0.871
sector C	0.005	0.298	-0.000	0.842	0.001	0.841
sector F	-0.001	0.773	-0.002	0.365	0.003	0.350
sector G	0.000	0.704	-0.000	0.847	0.000	0.768
sector H	0.004	0.210	0.002	0.391	-0.004	0.266
sector I	0.003	0.825	-0.001	0.618	0.000	0.617
sector J	0.004	0.238	0.004	0.210	-0.008	0.110
sector M	omitted		-0.000	0.928	0.000	0.928
sector N	-0.000	0.968	-0.000	0.968	0.000	0.968
const			-0.017	0.085		

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. 5 dgts sector FE absorbed. (1) First principal component of area's specialization e average firm's sales in the LLM. (2) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (3) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (4) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

Table 10 Blinder–Oaxaca 2steps decomposition: Services

	average productivity	p-value		Observations
overall				7,562,403
non urban	0.004	0.532		3,152,969
urban	0.084	0.000		4,409,434
difference	-0.079	0.001		
endowments	-0.067	0.005		
coefficients	-0.026	0.012		
interaction	0.016	0.154		

	Endowments effect		Coefficient effect		Interaction	
	coeff.	p-value	coeff	p-value	coeff.	p-value
Sales	-.046***	0.000	0.003**	0.011	-0.004**	0.002
Urban Economies (1)	-0.003*	0.058	-0.000	0.578	0.000	0.526
Education (2)	-0.016**	0.004	-.003	0.198	0.005	0.060
Mobility & RE (3) cong	-0.004	0.254	-0.002	0.244	.006	0.163
Size of loc mkt & demography (4)	-0.004	0.606	-0.006	0.122	0.015	0.084
Labor mkt participation	-0.008	0.096	-0.001	0.464	0.002	0.415
sector C	.005	0.198	0.000	0.927	-0.000	0.926
sector F	.001	0.792	-0.002	0.223	0.005	0.193
sector G	-.001	0.643	-0.000	0.699	0.001	0.655
sector H	.005	0.153	0.003	0.403	-0.007	0.123
sector I	.007**	0.004	-0.001	0.290	0.003	0.283
sector J	.004	0.132	0.007	0.233	-0.015**	0.019
sector M	omitted		0.001	0.305	-0.002	0.280
sector N	-.006	0.220	-0.002	0.253	0.006	0.220
const			-0.021	0.003		

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. 5 dgts sector FE absorbed. (1) First principal component of area's specialization e average firm's sales in the LLM. (2) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (3) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (4) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

Table 11 Blinder-Oaxaca 1 step decomposition: Manufacturing

	average productivity	p-value		Observations
overall				1333035
non urban	9.887	0.000		648,660
urban	9.964	0.000		684,375
difference	-0.077	0.054		
endowments	-0.064	0.105		
coefficients	-0.030	0.016		
interaction	0.017	0.080		

	Endowments effect		Coefficient effect		interaction	
	coeff.	p-value	coeff	p-value	coeff.	p-value
Sales	-0.030	0.182	-0.037	0.475	0.000	0.529
Urban Economies (1)	-0.001	0.146	-0.000	0.985	0.000	0.985
Education (2)	-0.021***	0.001	-0.001	0.386	0.004	0.230
Mobility & RE (3)	-0.000	0.957	0.001	0.437	0.003	0.220
Size of loc mkt & demography (4)	0.003	0.551	-0.002	0.296	0.010	0.140
Labor mkt participation	-0.015**	0.044	-0.061	0.653	0.001	0.659
sector C	0.006	0.264	-0.072	0.247	-0.010	0.307
sector F	-0.002	0.479	-0.004	0.833	-0.001	0.833
sector G	0.000	0.763	-0.054	0.142	0.001	0.748
sector H	0.005	0.151	0.001	0.923	-0.000	0.923
sector I	-0.001	0.422	-0.003	0.736	-0.001	0.737
sector J	0.006	0.152	0.009	0.132	-0.006	0.152
sector M	omitted		-0.020	0.177	0.005	0.188
sector N	0.000	0.801	-0.011	0.294	0.006	0.298
const			0.269	0.165		

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. 2 dgts sector FE absorbed. (1) First principal component of area's specialization e average firm's sales in the LLM. (2) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (3) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (4) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

Table 12 Blinder-Oaxaca 1 step decomposition: Services

	average productivity	p-value		Observations
overall				7,562,403
non urban	9.667	0.000		3,152,969
urban	9.822	0.000		4,409,434
difference	-0.155	0.000		
endowments	-0.144	0.000		
coefficients	-0.040	0.000		
interaction	0.029	0.007		

	Endowments effect		Coefficient effect		interaction	
	coeff.	p-value	coeff	p-value	coeff.	p-value
Sales	-0.006	0.468	0.260***	0.000	-0.001	0.472
Urban Economies (1)	-0.004**	0.042	-0.000	0.859	0.000	0.856
Education (2)	-0.026***	0.001	-0.004	0.147	0.007*	0.059
Mobility & RE cong	0.001	0.764	-0.004	0.242	0.006	0.212
Size of loc mkt & demography (4)	-0.013	0.144	-0.010*	0.065	0.021**	0.044
Labor mkt participation	-0.010*	0.079	-0.076	0.548	0.001	0.567
sector C	0.003	0.423	0.012	0.640	0.002	0.650
sector F	-0.002	0.560	0.022*	0.072	0.008*	0.084
sector G	-0.002	0.384	0.011	0.659	0.001	0.671
sector H	0.005	0.230	0.016	0.152	-0.007	0.212
sector I	0.004	0.144	0.010	0.093	0.006	0.103
sector J	0.007*	0.084	0.031***	0.002	-0.021**	0.011
sector M	omitted		0.017	0.045	-0.005*	0.054
sector N	-0.005	0.248	-0.009	0.243	0.005	0.248
const			-0.185	0.300		

Source: Authors' computation on data from Archivio statistico delle imprese attive (ASIA).
Notes: Clustered SE in parenthesis at LLM level. * significant at 10 per cent, ** at 5 per cent, *** at 1 per cent. 2 dgts sector FE absorbed. (1) First principal component of area's specialization e average firm's sales in the LLM. (2) First principal component of the gender gap in high school attainment, share of illiterates, early school dropout, share of adult with high school attainment, level of education in the 15-19 age range. (3) First principal component of variables proxying the quality of infrastructures (incidence of population who (i) commute daily outside the municipality of residence for work; (ii) use public transportation; (c) commute daily for study or work for longer than 60 minutes) and variable proxying the supply of residential housing (share of home ownership, share of residential buildings in the area increase in residential construction, residential mobility). (4) First principal component of population, population density, elders' dependency ratio, share of foreign residents, index of residential mobility of foreigners, school attendance of foreigners.

A Additional tables

Table A.1 Ateco 2007 classification: Sections

Code	Definition
A	Agriculture and fishing
B	Extraction
C	Manufacturing
D	Utilities: Provision of electric energy, gas, steam, AC
E	Utilities: Provision of water, sewage treatment, recycling
F	Construction
G	Wholesale and retail trade; car and motorcycle repair
H	Transportation and storage services
I	Accommodation & catering services
J	Information and communication services
K	Financial and insurance services
L	Real Estate
M	Professional scientific and technical services
N	Business services
O	Public administration & defense
P	Education
Q	Health services
R	Arts, entertainment and recreation services
S	Other services
T	Personal services

Source: Istat, Ateco 2007.