The sources of local productivity advantages: theory and evidence

Marcello Pagnini*

*Bank of Italy, Economic Research Unit, Bologna and Rimini center for economic analysis (RCEA)

Lecture notes,

Bologna, April 7th, 2022

Three questions and three papers:

1)How big are local productivity advantages associated to geographic concentration in Italy? What type of geographic concentration is more relevant?

V. Di Giacinto, M. Gomellini, G. Micucci & M. Pagnini (2014) "Mapping local productivity advantages in Italy: Industrial districts, cities or both?", *Journal of Economic Geography* 14.2 (2013): 365-394.

2)What are the forces explaining such differences and how to devise an effective identification strategy?

Accetturo, A., Di Giacinto, V., Micucci, G., & 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.

3) What is the strength of agglomeration economies in Italy? How much do they explain of the north-south productivity gap?

Buzzacchi, L., De Marco, A., & Pagnini, M. (2021). Agglomeration and the Italian North-South divide. *Bank of Italy Occasional Paper*, n. 637.

A bit of theory

•PRE Melitz (2003) & Melitz-Ottaviano (2008)
Agglomeration economies as the outcome of pure externalities:
Sharing
Matching
Learning

New economic geography models
Interplay of trasport costs, internal economies of scale and CES preferences

•POST M & MO

Selection models based on firm heterogenity combined with trasport costs and NON CES preferences

Misallocation models

Toward a synthesis?

First question and first paper:

Motivations

Much of emphasis attributed to ID in Italy and much less attention paid to UA. Is this choice empirically motivated?

Moreover: can we infer something about the mechanisms governing agglomeration economies (eg specialization vs diversification) from the previous question?

Defining ID (Sforzi-Istat algorithm) and UA in Italy

Step1: defining local labour markets areas (LLMA)

the Italian territory is divided into a number of LLMAs . These are aggregations of two or more neighboring municipalities based on daily commuting flows from place of residence to place of work as recorded in the Population Census. The share of resident LLMA employees in total LLMA employees and the share of resident LLMA employees in total LLMA residents must be at least 75 percent.

Step2: defining ID

ID are LLMAs matching the following conditions:

- Specialization in the manufacturing sector as a whole (location quotient (LQ) > 1)
- Share of manufacturing employment in small and medium sized firms (less than 250 employees) above the corresponding share at national level
- The same share has to be above .5 in the 'dominant industry' ie the single manufacturing activity with the largest number of employees out of those industries with LQ >1

- Step3 : defining UA

LLMAs with a population above 500,000

ID and UA in 1991

According to the Sforzi-Istat algorithm: 199 out of 784 LLMA in Italy were sorted out as ID

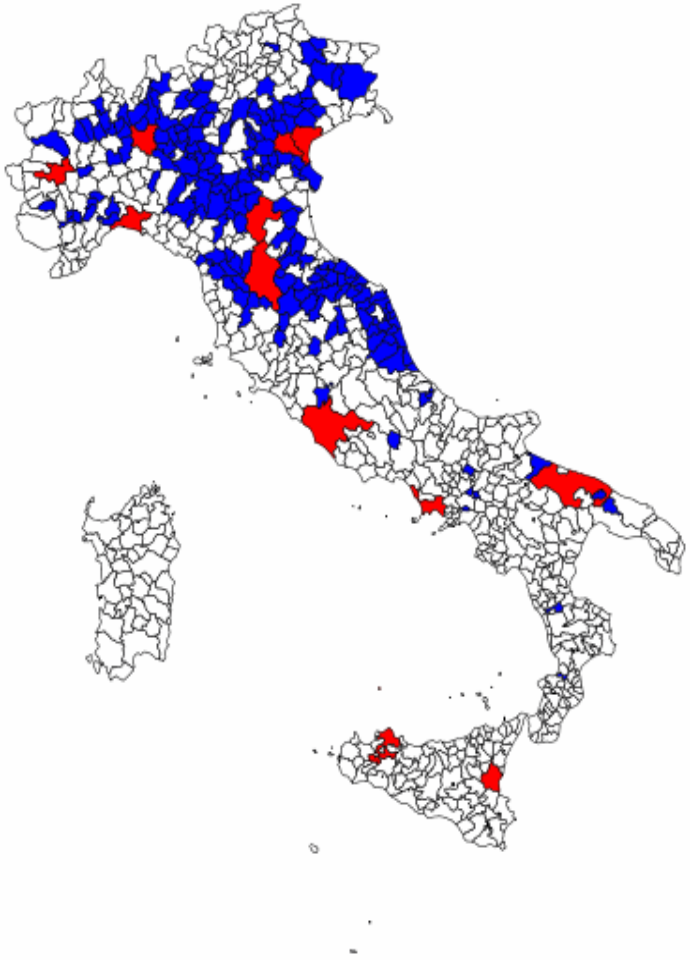
According to previous definition: 12 LLMA can be defined as UA in 1991

The two definitions overlap for only one LLMA

(Padua) included into UA

Mapping ID and UA in Italy

(ID=blue areas; UA= red areas)



Tfp estimation

Consider the following Cobb-Douglas production function:

$$Q_{i\in(r,s)t} = \Phi_{it} L_{it}^{\alpha_s} K_{it}^{\beta_s} \quad (1)$$

where L and K denote labour and capital inputs used to produce the amount of output Q for firm i belonging to sector s, located in LLMA r and in year t, α_s and β_s are the production coeff. varying across sectors

After log transformation, estimate the following:

$$q_{it} = \alpha_s l_{it} + \beta_s k_{it} + \phi_{it} \quad (2)$$

tfp is then obtained as a residual:

$$\phi_{it} = q_{it} - \hat{\alpha}_s l_{it} + \hat{\beta}_s k_{it} \quad (3)$$

Data

Sample of approx. 500,000 manufacturing firms (Corporations, not plants) per year observed in the period 1995-2006
Source: Chamber of commerce-Company Accounts Data Service

Balance sheet data on value added, consumption of intermediate goods, stock of capital,

Additional information at firm level includes: number of employees, sector of ec. activity (up to the 4 digits SIC sector classification), location (municipality where the firm locates)

Some additional empirical issues

Only few firms report L , missing data are imputed by using a regression analysis where total labour cost is used as the main predictor

Capital stock is estimated through the perpetual inventory method

Value added and consumption of intermediate goods are deflated by using specific industry price indexes

firms with $L < 4$ and with extreme values for K/L are excluded

The final sample consist approx. of 29,000 firms per year

Equation (2) is estimated by industry through ols (LS) and through fixed effect (FE) and Levinsohn and Petrin (LP) methods to control for input-output simultaneity

Baseline specification

Consider the following regression:

$$\hat{\phi}_{it} = \delta UA + \eta ID + \rho\, flagimp_{it} + \Sigma_h \mu_h\, firmsize_{it}^h + \gamma_g + \lambda_s + \varpi_t + \varepsilon_{it} \quad (4)$$

where the first two terms are dummies indicating firms located in UA or ID, the following vars signal wether L_{it} has been inputed and firm size, γ_g, λ_s and ω_t are area, industry and year fixed effects and the last term is an error term for which we assume:

$$\varepsilon_{it} = \iota_r + \eta_{it} \quad (5)$$

Given the assumptions about the error term in (5), we estimate eq. (4) by clustering error terms for LLMAs

Baseline regression: dependent variable log of tfp measured through LP method (1)

(Estimation period: 1995-2006; standard errors in brackets)

	Model I	Model II	Model III (2)
UA	0.102***	0.108***	0.092***
	(0.01)	*(0.01)*	*(0.01)*
ID	0.029***	0.036***	0.016*
	(0.01)	*(0.01)*	*(0.01)*
Medium size	0.033***		0.037***
	(0.01)		*(0.01)*
Large size	0.160***		0.164***
	(0.01)		*(0.01)*
Northeast	-0.001	-0.001	
	(0.01)	*(0.01)*	
Centre	-0.035**	-0.036**	
	(0.01)	*(0.01)*	
South	-0.242***	-0.242***	
	(0.01)	*(0.01)*	
UA*medium		-0.039*	
		(0.02)	
UA*large		0.030	
		(0.03)	
ID*medium		-0.037**	
		(0.01)	
ID*large		-0.001	
		(0.03)	
N	344353	344353	344353
r2_a	0.677	0.678	0.679

(1) All specifications include year and industry fixed effects, a control for inputed data.-(2) It includes 20 region fixed effects.

Robustness

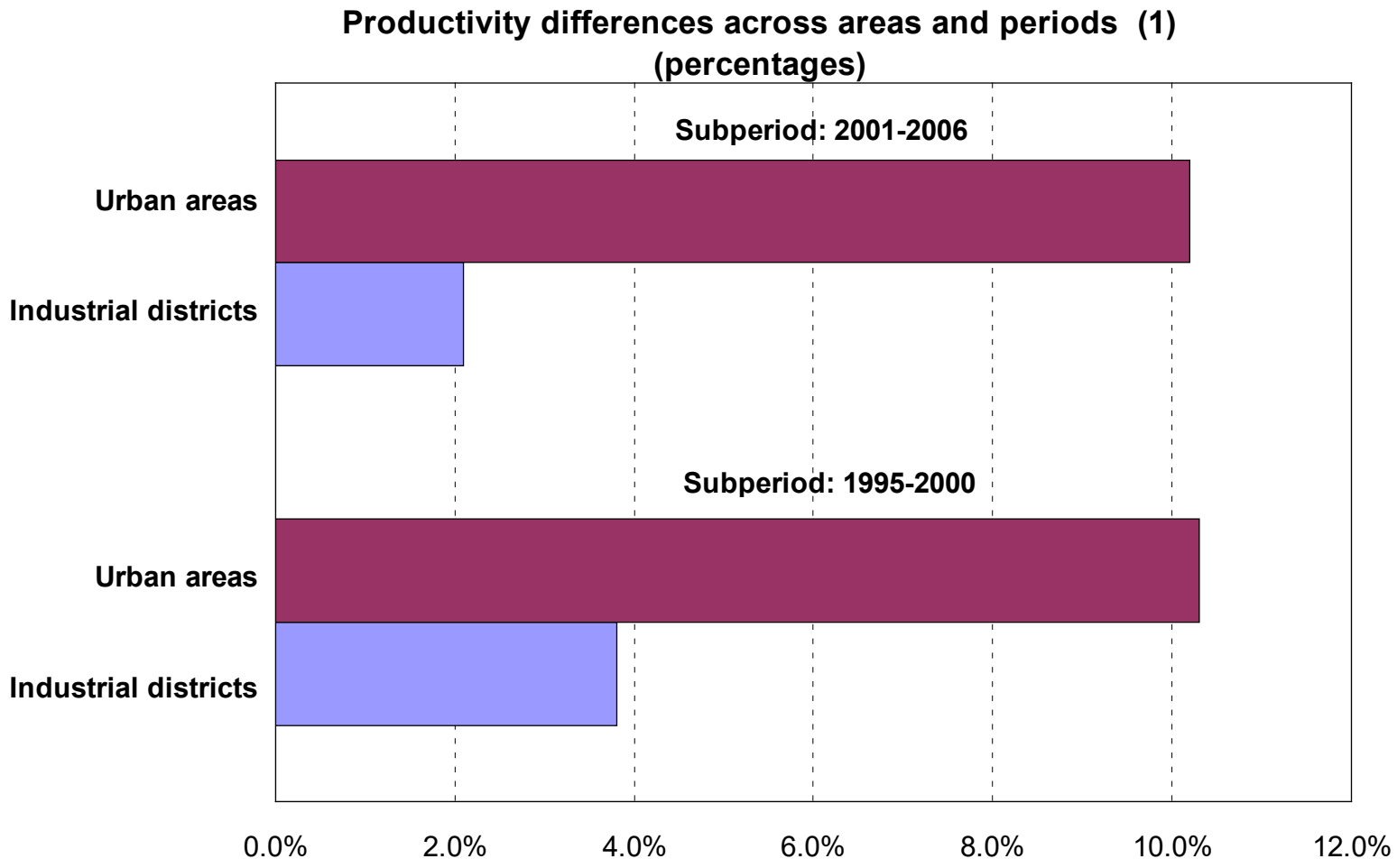
- Using the 2001 map for ID

- IV estimation: potential endog. vars ID and UA dummies; Instruments: Population density in 1921 and the share population with an university or secondary school degree in 1971 in each LLMA; the share of LLMA's land near the coastline and the log of the LLMA average altitude

- Monoplants firms

- Human capital het. Across ID and UA (white and blue collars shares)

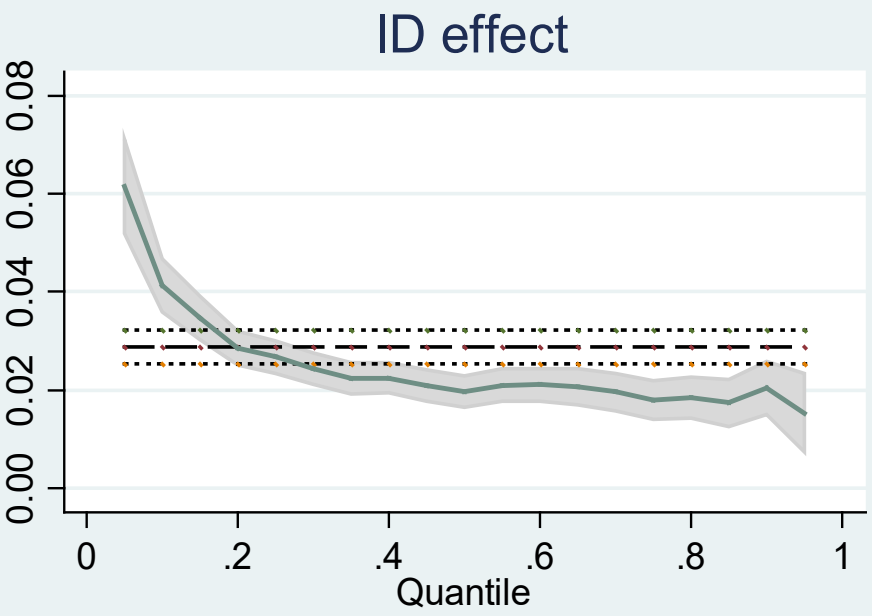
- Other controls

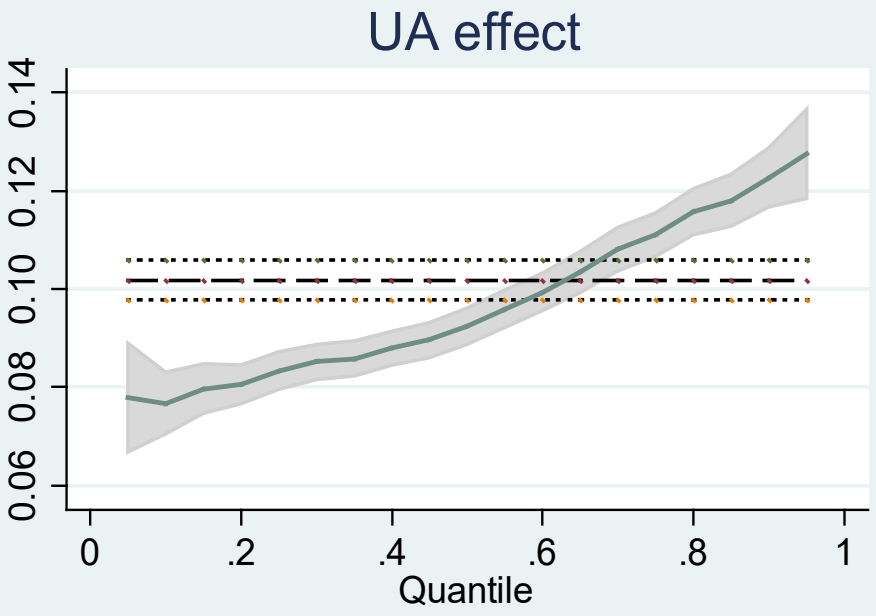


(1) This graph reports the estimated parameters for urban areas and industrial districts based on specification in equation [1]. All the estimated coefficients are significant at 1% level. For an illustration of the full set of results refer to Di Giacinto et al. (2012).
Source: Our elaborations based on data from Centrale dei Bilanci, Cerved and INPS.

Fig. 2 – Quantile regression: ID and UA effects (1)

The horizontal line corresponds to OLS coefficients, grey areas and horizontal dotted lines denote 95% confidence intervals for the estimated parameters. Quantiles vary from .05 to .95 and are incremented by .05.





Second question and second paper

Why does spatial concentration generate local productivity advantages?

Three (observationally equivalent?) explanations:

- Agglomeration economies:

Fujita and Ogawa (1982); Lucas and Rossi-Hansberg (2002); Rosenthal and Strange (2004) and Melo, Graham, and Noland (2009)

- Selection:

Melitz (2003) and Melitz and Ottaviano (2008); Syverson (2004 a and b); Del Gatto, Ottaviano and Pagnini (2008)

- Firm or worker spatial sorting/dilation/differential absorptive capacity:

firms (Baldwin and Okubo,2006 ; Okubo, Picard and Thisse ,2009; Nocke, 2006) and workers (Gould, 2007, Matano and Naticchioni, 2011 Venables, 2011)

Agglomeration and selection effects

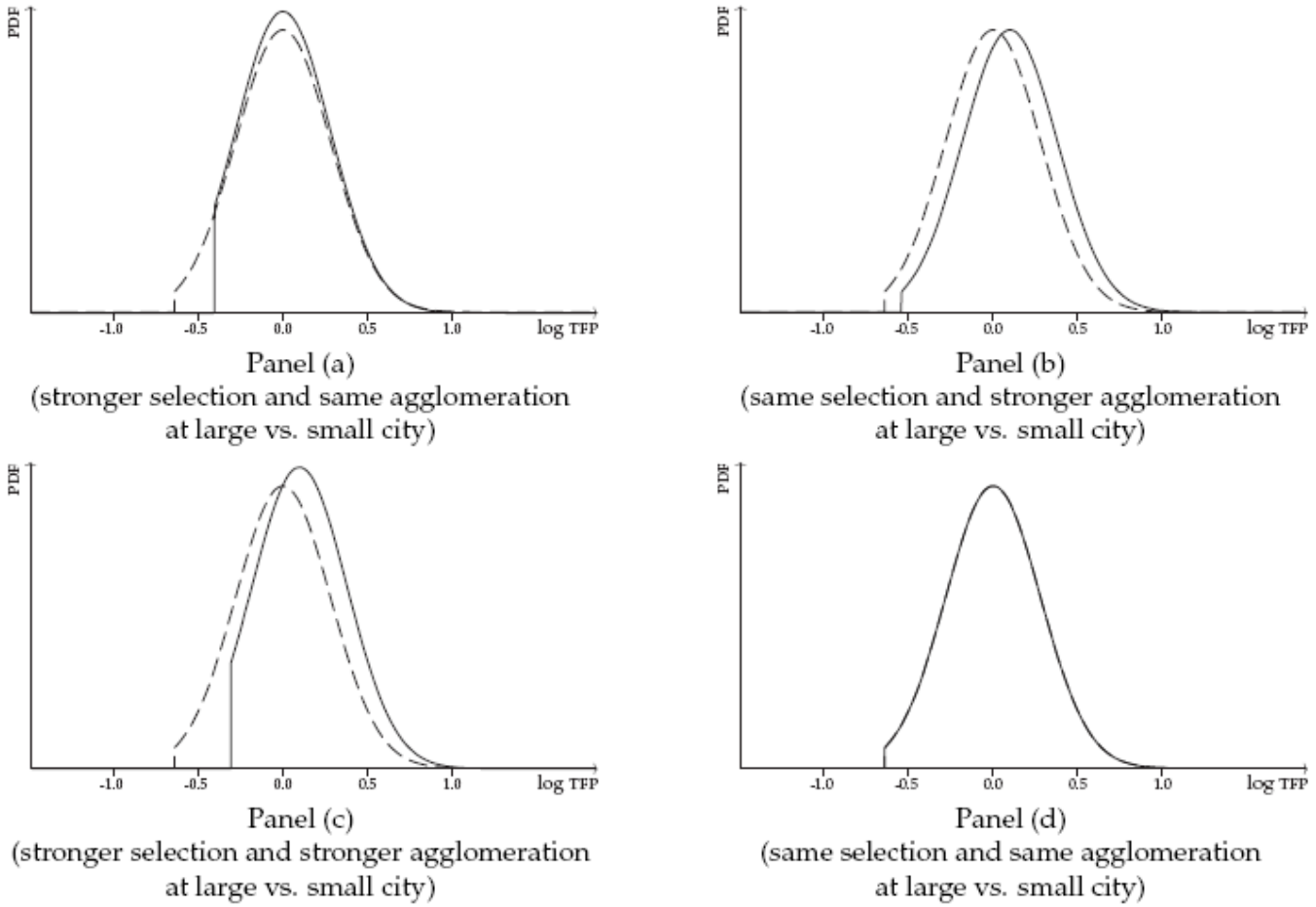


Figure 1: Log productivity distributions in large (solid) and small cities (dashed)

Source: Combes et al (2012)

Combes, Duranton, Gobillon, Puga, and Roux (2012)

Main findings:

Theory: Melitz and Ottaviano(2008) model with a generic tfp distribution and with agglomeration, selection & dilation nested within the same model):

•Dense markets generate larger agglomeration economies,

•More efficient firms gain more from these positive externalities (dilation effect)

•Large markets generate stronger selection effects: ie a larger share of less efficient firms is forced to exit from the market

Empirics (new empirical methodology to estimate agglomeration, truncation and dilation):

•Productivity differences across 341 french employment areas are mostly explained by agglomeration.ie tfp distribution is right shifted at the above median population density areas.

•Selection play no role: no stronger left truncation in denser areas

• Moderate evidence of a dilation effect

Our Contribution:Why does selection play no role? I theory

Three explanations all related to some kind of additional sources of market asymmetries:

• Provided entry costs are positively correlated with market size, there might exist equilibrium configurations where large markets will exhibit less intense selection effects than those displayed in small markets

• Asymmetric trade costs: local market access to other locations may foster competition and induce stronger selection effects beyond those generated by local market size

•Agglomeration and selection operate at a different spatial scale: rapid spatial decay of agglomeration economies; definition of a relevant market for manufacturing products

Our Contribution:Why does selection play no role? II empirical strategy

• Detecting potential confounding factor and netting out for their effects: dropping LLMAs with large elasticity of entry costs to market size (local population)

•Grouping LLMAs according to their market access to domestic & foreign markets and compare selection effects

•Estimating agglomeration and selection at different spatial scales

Steps in the empirical strategy

- Grouping firms according to a unique characteristic of the LLMA's where they locate. Baseline: LLMAs with pop. > 200,000 vs LLMAs with pop < 200,000

- Comparing tfp distributions betweent the two groups of firms or LLMA's

- Parameter estimation by industry:

i. A (>0): agglomeration, rightward shift of the tfp distribution

ii. S(>0) : selection, rightward shift of the truncation point,

iii. D(>1): dilation: high productivity firms benefit more from agglomeration economies

The zoning system: defining UA in Italy

Step1: defining local labor markets areas (LLMA)

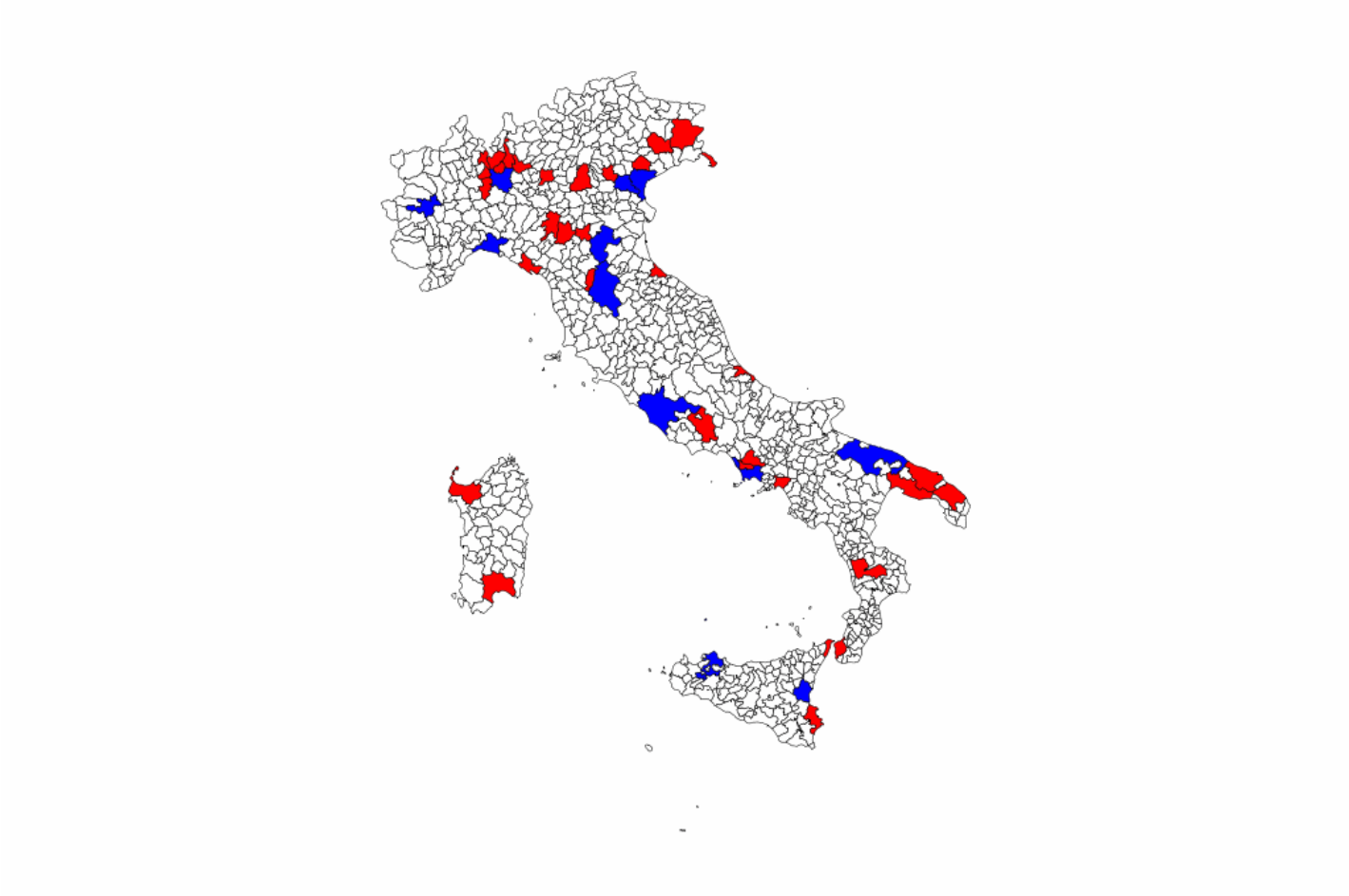
the Italian territory is divided into a number of LLMAs . These are aggregations of two or more neighboring municipalities based on daily commuting flows from place of residence to place of work as recorded in the Population Census. The share of resident LLMA employees in total LLMA employees and the share of resident LLMA employees in total LLMA residents must be at least 75 percent.

Step2: obtaining 686 LLMA in Italy in 2001

Step3: defining UA as those LLMAs with a population above 200,000 (base line) alternative threshold (500,000) and alternative proxies for local market size (population density)

Mapping Urban areas in Italy: 686 local labor market areas in 2001

(pop. >500,000 blue areas; 200,000 <pop<500,000 red areas; pop < 200,000 white areas)



Baseline specification

Estimates of Agglomeration (A), Selection (S) and Dilation (D).
Urban areas: population > 200,000

Sectors	A *(s.e.)*	S *(s.e.)*	D *(s.e.)*	Obs. for non UA	Obs. for UA	R2
Food products, beverages and tobacco	**0.111** *(0.02)**	**0.012** *(0.01)*	**1.143** *(0.05)**	1,826	1,200	0.967
Textiles and textile products	**0.058** *(0.01)**	**0.003** *(0.01)*	**1.086** *(0.03)**	2,781	2,686	0.970
Leather and leather products	**0.029** *(0.01)**	**0.012** *(0.01)*	**1.121** *(0.07)*	1,639	704	0.907
Wood and products of wood and cork (except furniture)	**0.060** *(0.02)**	**-0.002** *(0.01)*	**1.013** *(0.07)*	872	508	0.937
Pulp, paper and paper products; recorded media; printing services	**0.069** *(0.01)**	**-0.003** *(0.00)*	**1.150** *(0.05)**	994	1,735	0.966
Coke, refined petroleum products and nuclear fuel	**-0.049** *(0.30)*	**-0.006** *(0.75)*	**1.191** *(0.41)*	60	87	0.895
Chemicals, chemical products and man-made fibres	**0.107** *(0.04)**	**0.007** *(0.03)*	**1.160** *(0.09)*	521	966	0.922
Rubber and plastic products	**0.047** *(0.01)**	**0.012** *(0.01)*	**1.058** *(0.06)*	1,288	1,193	0.912
Other non metallic mineral products	**0.047** *(0.01)**	**0.009** *(0.01)*	**0.994** *(0.05)*	1,710	916	0.958
Basic metals and fabricated metal products	**0.046** *(0.00)**	**0.002** *(0.00)*	**0.996** *(0.02)*	5,090	4,809	0.981
Machinery and equipment n.e.c.	**0.049** *(0.01)**	**0.000** *(0.00)*	**1.045** *(0.03)*	2,917	3,261	0.983
Electrical and optical equipment	**0.085** *(0.01)**	**0,000** *(0.00)*	**1.183** *(0.04)**	1,702	2,466	0.989
Transport equipment	**0.045** *(0.01)**	**0.007** *(0.01)*	**1.061** *(0.06)*	701	767	0.929
Other manufactured goods n.e.c.	**0.052** *(0.01)**	**0.002** *(0.01)*	**1.118** *(0.04)**	2,148	1,599	0.961

Source: elaborations on Centrale dei Bilanci, Cerved.
Estimations of Average Total Factor Productivity per firm, by adopting the procedure of Levinsohn and Petrin. Period: 1995-2006. The t statistics are obtained from 50 bootstrapped replications. *: for A and S significantly different from 0 at 5%, for D significantly different from 1 at 5%.

Dropping LLMA with high entry cost elasticity

Estimates of Agglomeration (A), Selection (S) and Dilation (D).
Urban areas: population > 200,000, excluding LLMA with s'> 75th percentile

Sectors	A *(s.e.)*	S *(s.e.)*	D *(s.e.)*	Obs. for non UA	Obs. for UA	R2
Food products, beverages and tobacco	**0.076** *(0.08)*	**0.013** *(0.48)*	**1.091** *(0.10)*	1346	671	0.905
Textiles and textile products	**0.052** *(0.01)**	**0.011** *(0.01)*	**1.027** *(0.04)*	1975	1941	0.928
Leather and leather products	**0.013** *(0.02)*	**0.012** *(0.02)*	**1.137** *(0.14)*	1229	310	0.796
Wood and products of wood and cork (except furniture)	**0.037** *(0.04)*	**-0.004** *(0.13)*	**1.002** *(0.11)*	626	346	0.830
Pulp, paper and paper products; recorded media; printing services	**0.035** *(0.02)**	**-0.001** *(0.01)*	**1.021** *(0.07)*	738	747	0.745
Coke, refined petroleum products and nuclear fuel	**-0.048** *(0.29)*	**0.081** *(0.47)*	**1.102** *(0.37)*	42	31	0.159
Chemicals, chemical products and man-made fibres	**0.018** *(0.04)*	**0.060** *(0.04)*	**1.164** *(0.14)*	374	375	0.891
Rubber and plastic products	**0.030** *(0.02)*	**0.012** *(0.02)*	**1.011** *(0.09)*	887	731	0.869
Other non metallic mineral products	**0.055** *(0.05)*	**-0.000** *(0.23)*	**0.916** *(0.10)*	1170	567	0.899
Basic metals and fabricated metal products	**0.044** *(0.00)**	**0.002** *(0.00)*	**0.939** *(0.03)**	3689	3116	0.988
Machinery and equipment n.e.c.	**0.031** *(0.01)**	**0.003** *(0.00)*	**1.000** *(0.03)*	2165	1970	0.950
Electrical and optical equipment	**0.034** *(0.02)*	**-0.001** *(0.04)*	**1.066** *(0.07)*	1274	1157	0.887
Transport equipment	**0.017** *(0.06)*	**0.006** *(0.21)*	**0.982** *(0.16)*	549	441	0.882
Other manufactured goods n.e.c.	**0.026** *(0.03)*	**-0.023** *(0.07)*	**1.053** *(0.10)*	1602	1022	0.828

Source: elaborations on Centrale dei Bilanci, Cerved.
Estimations of Average Total Factor Productivity per firm, by adopting the procedure of Levinsohn and Petrin. Period: 1995-2006. The t statistics are obtained from 50 bootstrapped replications. *: for A and S significantly different from 0 at 5%, for D significantly different from 1 at 5%.

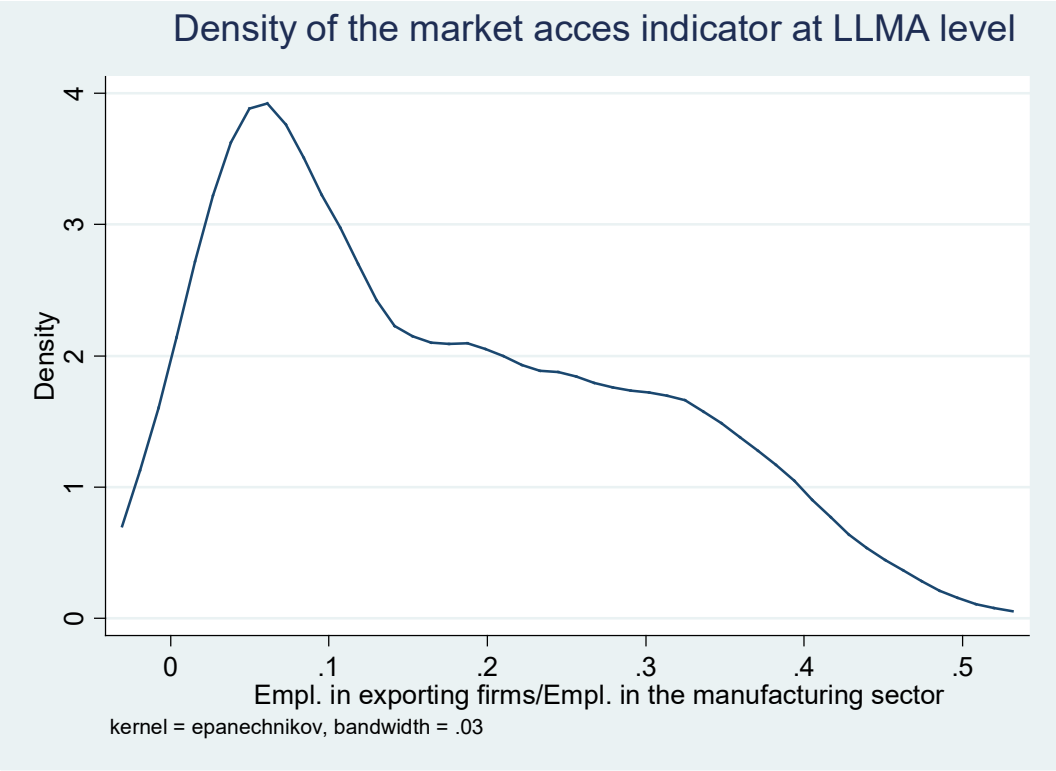
Market access to domestic markets: LLMAs with market potential ($MP_i = \sum_j \frac{Pop_j}{d_{ij}}$) above the 75° perc

Sectors	A *(s.e.)*	S *(s.e.)*	D *(s.e.)*	Obs. for non UA	Obs. for UA	R2
Food products, beverages and tobacco	0.152 (0.018)*	0.012 (0.008)	1.034 (0.049)	1469	1566	0.980
Textiles and textile products	0.076 (0.014)*	0.022 (0.01)*	0.997 (0.039)	1662	3780	0.951
Leather and leather products	0.058 (0.015)*	0.002 (0.009)	0.990 (0.052)	1019	1341	0.965
Wood and products of wood and cork (except furniture)	0.082 (0.015)*	0.006 (0.014)	1.046 (0.08)	741	636	0.943
Pulp, paper and paper products; recorded media; printing services	0.088 (0.015)*	0.029 (0.013)*	0.975 (0.053)	963	1747	0.969
Coke, refined petroleum products and nuclear fuel	0.066 (0.183)	0.019 (0.403)	1.177 (0.308)	82	64	0.786
Chemicals, chemical products and man-made fibres	0.101 (0.025)*	0.002 (0.022)	0.948 (0.07)	427	1065	0.951
Rubber and plastic products	0.115 (0.017)*	0.017 (0.017)	0.924 (0.062)	797	1686	0.966
Other on metallic mineral products	0.133 (0.017)*	0.011 (0.017)	0.837 (0.058)	1180	1450	0.945
Basic metals and fabricated metal products	0.062 (0.007)*	0.017 (0.007)*	0.955 (0.019)	3578	6274	0.975
Machinery and equipment n.e.c.	0.055 (0.009)*	0.015 (0.005)*	0.949 (0.031)	1668	4492	0.990
Electrical and optical equipment	0.079 (0.013)*	0.013 (0.008)	0.971 (0.04)	1553	2598	0.968
Transport equipment	0.032 (0.019)	0.012 (0.013)	0.872 (0.054)	758	706	0.929
Other manufactured goods n.e.c.	0.086 (0.013)*	0.011 (0.01)	1.024 (0.039)	1892	1842	0.941

Source: elaborations on Centrale dei Bilanci, Cerved.
Estimations of Average Total Factor Productivity per firm, by adopting the procedure of Levinsohn and Petrin. Period: 1995-2006. The t statistics are obtained from 50 bootstrapped replications. *: for A and S significantly different from 0 at 5%, for D significantly different from 1 at 5%. LMMA are defined in 2001.

Proxy for local market access to foreign markets

Employees in the LLMA working in plants exporting to foreign markets over the total employees in the manufacturing activities in the LLMA



Grouping LLMA according to market access to foreign markets

Estimates of Agglomeration (A), Selection (S) and Dilation (D).
LLMA with better access to foreign markets: LMMA with a ratio between local employees in exporting plants and total employees > 0.2705 (the 75^ percentile of this variable across LMMA)

Sectors	A *(s.e.)*	S *(s.e.)*	D *(s.e.)*	Obs. for non UA	Obs. for UA	R2
Food products, beverages and tobacco	**0.194** *(0.02)**	**0.031** *(0.02)**	**1.067** *(0.06)*	1,397	1,613	0.984
Textiles and textile products	**0.062** *(0.02)**	**0.008** *(0.01)*	**0.922** *(0.04)**	1,986	3,477	0.937
Leather and leather products	**0.121** *(0.07)*	**0.012** *(0.08)*	**1.020** *(0.13)*	678	1,674	0.952
Wood and products of wood and cork (except furniture)	**0.087** *(0.02)**	**0.020** *(0.02)*	**0.892** *(0.06)*	454	917	0.967
Pulp, paper and paper products; recorded media; printing services	**0.115** *(0.02)**	**0.021** *(0.02)*	**0.866** *(0.04)**	763	1,961	0.955
Coke, refined petroleum products and nuclear fuel	**0.129** *(0.23)*	**0.019** *(0.25)*	**1.203** *(0.38)*	79	67	0.854
Chemicals, chemical products and man-made fibres	**0.169** *(0.03)**	**0.019** *(0.01)*	**0.873** *(0.06)**	344	1,142	0.994
Rubber and plastic products	**0.142** *(0.02)**	**0.015** *(0.01)*	**0.867** *(0.05)**	590	1,897	0.988
Other non metallic mineral products	**0.149** *(0.02)**	**0.015** *(0.02)*	**0.784** *(0.05)**	1,097	1,529	0.975
Basic metals and fabricated metal products	**0.102** *(0.01)**	**0.026** *(0.01)**	**0.930** *(0.03)**	2,494	7,355	0.986
Machinery and equipment n.e.c.	**0.092** *(0.01)**	**0.016** *(0.01)**	**0.974** *(0.03)*	1,119	5,048	0.988
Electrical and optical equipment	**0.106** *(0.02)**	**0.003** *(0.02)*	**0.884** *(0.05)**	1,077	3,093	0.964
Transport equipment	**0.072** *(0.02)**	**0.020** *(0.01)*	**0.881** *(0.05)**	570	892	0.974
Other manufactured goods n.e.c.	**0.102** *(0.02)**	**0.004** *(0.01)*	**0.874** *(0.06)**	1,017	2,733	0.920

Source: elaborations on Centrale dei Bilanci, Cerved.
Estimations of Average Total Factor Productivity per firm, by adopting the procedure of Levinsohn and Petrin. Period: 1995-2006. The t statistics are obtained from 50 bootstrapped replications. *: for A and S significantly different from 0 at 5%, for D significantly different from 1 at 5%. LMMA are defined in 2001.

Changing the spatial scale

Baseline: 686 LLMA in 2001

New estimation:103 Italian provinces (nuts 2 classification)

STRUCTURAL CHARACTERISTICS OF LLMA AND PROVINCES

LLMA

variable	mean	p50	sd	min	max	N
POPT	83372.47	32901	225817.9	4368	3452792	686
SURFKM2	439.8955	353.42	350.1348	10.36	3661.14	686
dens	183.5222	97.11239	290.9731	12.06682	3979.991	686

Provinces

variable	mean	p50	sd	min	max	N
POPT	554467.4	372712	614983.9	87827	3783193	103
SURFKM2	2925.519	2561.73	1750.512	211.82	7519.93	103
dens	242.1266	168.8291	331.249	35.91648	2631.724	103

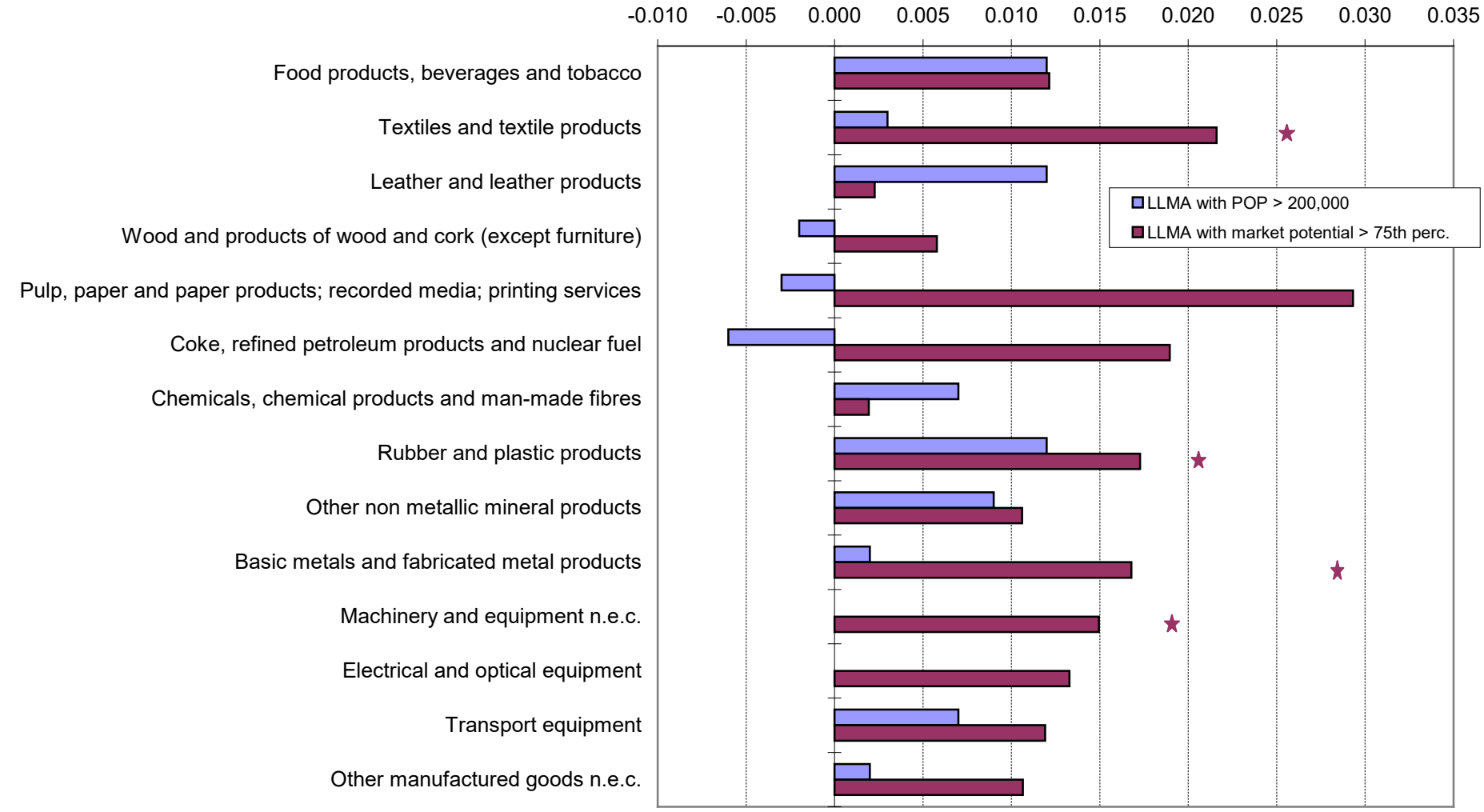
Changing the spatial scale: 103 Italian Provinces

Estimates of Agglomeration (A), Selection (S) and Dilation (D).
Urban Areas: Italian provinces with population density above the mean level (242.12)

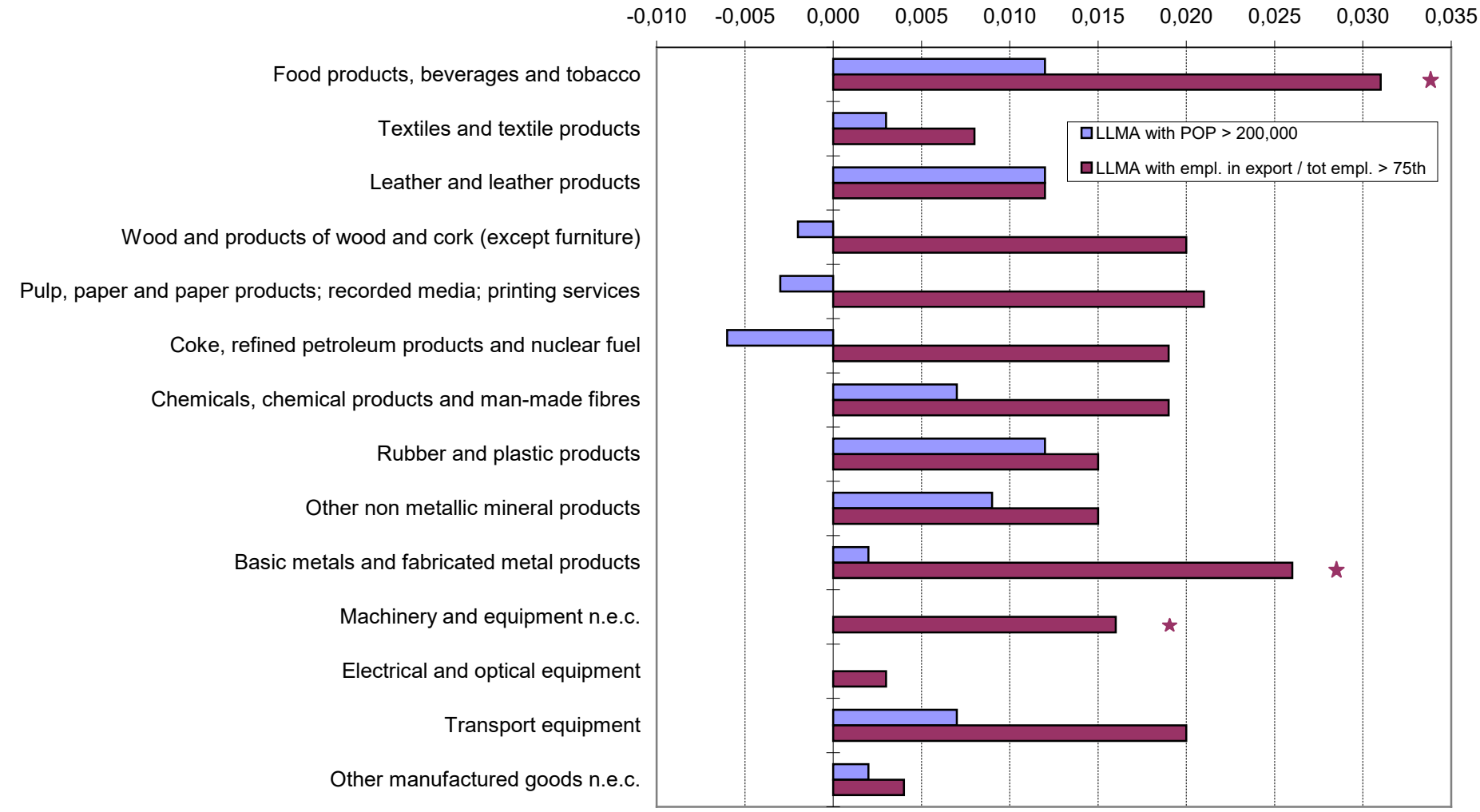
Sectors	A *(s.e.)*	S *(s.e.)*	D *(s.e.)*	Obs. for non UA	Obs. for UA	R2
Food products, beverages and tobacco	**0.126** (0.018)*	**0.001** (0.006)	**1.042** (0.038)	1722	1329	0.985
Textiles and textile products	**0.050** (0.009)*	**0.014** (0.007)*	**1.036** (0.032)	1878	3573	0.936
Leather and leather products	**0.032** (0.015)*	**0.017** (0.015)	**1.056** (0.058)	1149	1192	0.910
Wood and products of wood and cork (except furniture)	**0.066** (0.019)*	**0.003** (0.03)	**0.944** (0.071)	707	672	0.929
Pulp, paper and paper products; recorded media; printing services	**0.093** (0.015)*	**-0.004** (0.006)	**1.009** (0.043)	833	1899	0.912
Coke, refined petroleum products and nuclear fuel	**0.049** (0.258)	**-0.017** (0.506)	**1.210** (0.428)	63	83	0.781
Chemicals, chemical products and man-made fibres	**0.105** (0.056)	**0.011** (0.051)	**1.091** (0.126)	398	1088	0.883
Rubber and plastic products	**0.076** (0.015)*	**0.020** (0.014)	**0.991** (0.055)	942	1537	0.943
Other non metallic mineral products	**0.060** (0.015)*	**0.019** (0.011)	**0.991** (0.049)	1525	1088	0.971
Basic metals and fabricated metal products	**0.053** (0.005)*	**0.013** (0.004)*	**0.997** (0.023)	4202	5660	0.971
Machinery and equipment n.e.c.	**0.046** (0.008)*	**-0.002** (0.004)	**1.005** (0.025)	2526	3652	0.960
Electrical and optical equipment	**0.093** (0.011)*	**0.007** (0.006)	**1.144** (0.034)*	1341	2825	0.968
Transport equipment	**0.061** (0.018)*	**0.011** (0.01)	**1.046** (0.062)	594	873	0.971
Other manufactured goods n.e.c.	**0.043** (0.01)*	**0.007** (0.006)	**1.005** (0.04)	1722	2019	0.909

Source: elaborations on Centrale dei Bilanci, Cerved.
Estimations of Average Total Factor Productivity per firm, by adopting the procedure of Levinsohn and Petrin. Period: 1995-2006. The t statistics are obtained from 50 bootstrapped replications. *: for A and S significantly different from 0 at 5%, for D significantly different from 1 at 5%. LMMA are defined in 2001.

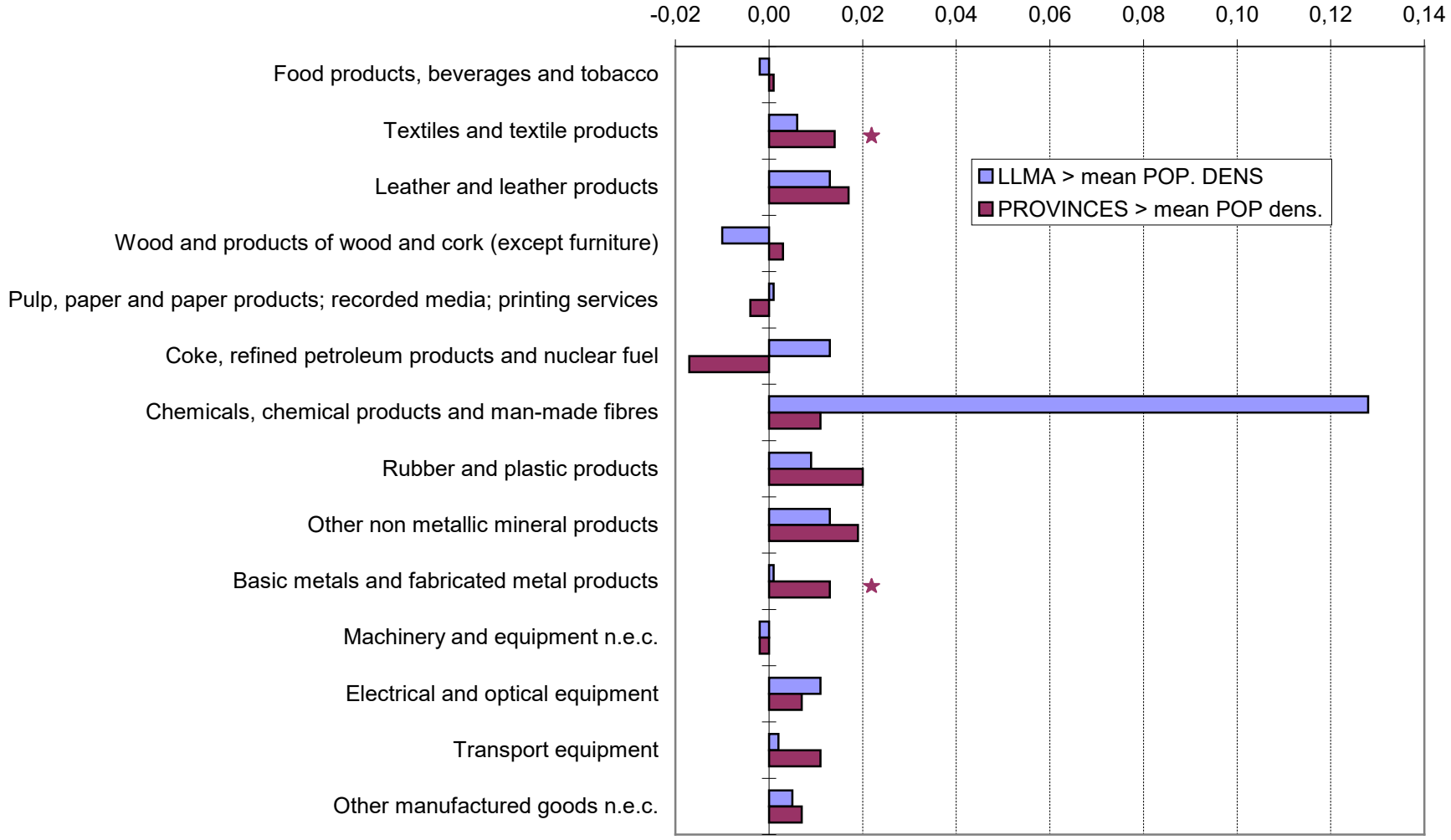
Comparing selection effects: local market size vs domestic market access



Comparing selection effects: market size vs foreign market access



Comparing selection effects: LLMA versus Provinces



Summing up and …

•Relevant local productivity advantages are detected for firms in urban areas

•Those advantages are higher and more resiliant compared to productivity premia in Industrial districts

•when using LLMA as geographical units, agglomeration is the main driver of the local productivity advantages in cities. Also a relevant dilation effect is detected.

• When grouping LLMA according to their market access to domestic & foreign markets, we find stronger selection effects in areas with a better access

•When using larger geographical units (Italian provinces), agglomeration is still the main force generating productivity advantages in denser areas however a selection effect also does emerge

…what to do in the next future

- Improving the proxies for market access both to foreign and to domestic markets; definition of a relevant market for manufacturing products

- Estimating A,D and S for specific subindustries (eg concrete)

- Theory: what role for alternative preference structures?

Third question and paper:

Research goals

- We provide a measure of the elasticity of TFP with respect to an indicator of local economic density for the Italian manufacturing sector:
  - the results may be considered as a benchmark for the comparison between the Italian case and similar estimates carried out for other countries

- We study whether agglomeration economies can contribute to the explanation of the productivity gap between the firms located in the northern and southern regions of Italy:
  - firms located in the South of Italy might be less productive due to the fact that local markets display a lower density of economic activity
  - the positive productivity gap might stem from a lower TFP elasticity with respect to density as compared to the North

Motivation

- The concentration of workers, firms, or institutions in specific areas might generate productivity advantages for the companies located within their borders

- Scholars argue that agglomeration (i.e., proximity) 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*'

- Three channels: matching, sharing and learning

- Empirical studies have investigated this nexus showing that the Total Factor Productivity (TFP) of firms increases with the density of economic activities in the local markets

Literature

Two (not fully connected) streams of literature:

•Agglomeration economies:

Duranton and Puga (2004), Ciccone and Hall (1996), Rosenthal and Strange (2004), Melo, Graham, and Noland (2009), Ahlfeldt and Pietrostefani (2019)

In Italy: Signorini (ed 2000), Cingano and Schivardi (2004), Di Giacinto, Gomellini, Micucci and Pagnini (2014), Accetturo, Di Giacinto, Micucci and Pagnini (2018)

•North-South divide in Italy:

Toniolo (2013) , Felice (2018) , Locatelli, Ciani, Pagnini (2019)

Identification challenges (1)

- Usually estimations of agglomeration economies refer to net (benefits-costs) returns and not to the gross components

- The three channels: matching, sharing and leanrning are difficult to be detected and identified

- Agglomeration economies are a black box

Identification challenges (2)

- Three main classes of mechanisms can determine an emerging correlation between density and economic advantages (e.g., productivity), even if agglomeration has no causal role:
  - ❶ first-nature advantages turn into better local outcomes that attract firms and workers in specific locations, thus affecting their performance
  - ❷ a positive relationship between productivity and density will be observed if stronger selection effects emerge in agglomerated markets
  - ❸ sorting mechanisms rooted in the idea that more productive 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, etc.

Main results preview

- Elasticity is around 6%, stable across the different specifications, incl. the one based on the 110 provinces fixed effects.
- Controlling for sector composition only, the southern LLMAs have a lower TFP, on average, by 26.7% compared to North (28.7% in terms of median)
- Moving median density for the South to the level of that for the median in the North (i.e., from 14 to 36 workers per sk), augments TFP in the South by approx. 5.7%
- Differences in density contribute to explain one fourth of the variance of TFP across LLMAs
- No significant differences of the density elasticity between the north and the south
- No evidence of stronger selection due to tougher competition in the northern LLMAs

Methodology and data

- We measure TFP at the firm level with a dataset that includes a large sample of Italian manufacturing firms (approx. 75,000 firms per year) observed in the years between 1995 and 2015 (Locatelli, Ciani and Pagnini, 2019)

- We aggregate TFP firm level data into the 611 Local Labor Market Areas (LLMA) as defined in 2011 by the Italian National Institute of Statistics (ISTAT)

- We estimate the presence of agglomeration economies by regressing the logarithm of TFP on the logarithm of the number of employees per square kilometer for each LLMA

- We address the problem of endogeneity by resorting to Two-Stages Least Squares (2SLS) regressions with a rich set of historical and geological instruments

TFP estimation

- Cobb-Douglas:

$$Y_{i,t} = A_{i,t}\left(\sum_{h} s^{h} L^{h}\right)_{i,t}^{\alpha} K_{i,t}^{\beta}$$

$$\log\left(P_{i,t}Y_{i,t}\right) = \alpha \log\left(W_{i,t}\right) + \beta \log\left(K_{i,t}\right) + \varepsilon_{it} \quad (3)$$

$$TFP_{i,t} = \log\left(P_{i,t}Y_{i,t}\right) - \hat{\alpha} \log\left(W_{i,t}\right) - \hat{\beta} \log\left(K_{i,t}\right) \quad (4)$$

Wages used as a proxy for labor input; Levinsohn and Petrin (2003) to control for output-input simultaneity; Estimation by 10 distinct man. Industries

Aggregation: from firm to LLMA TFP

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

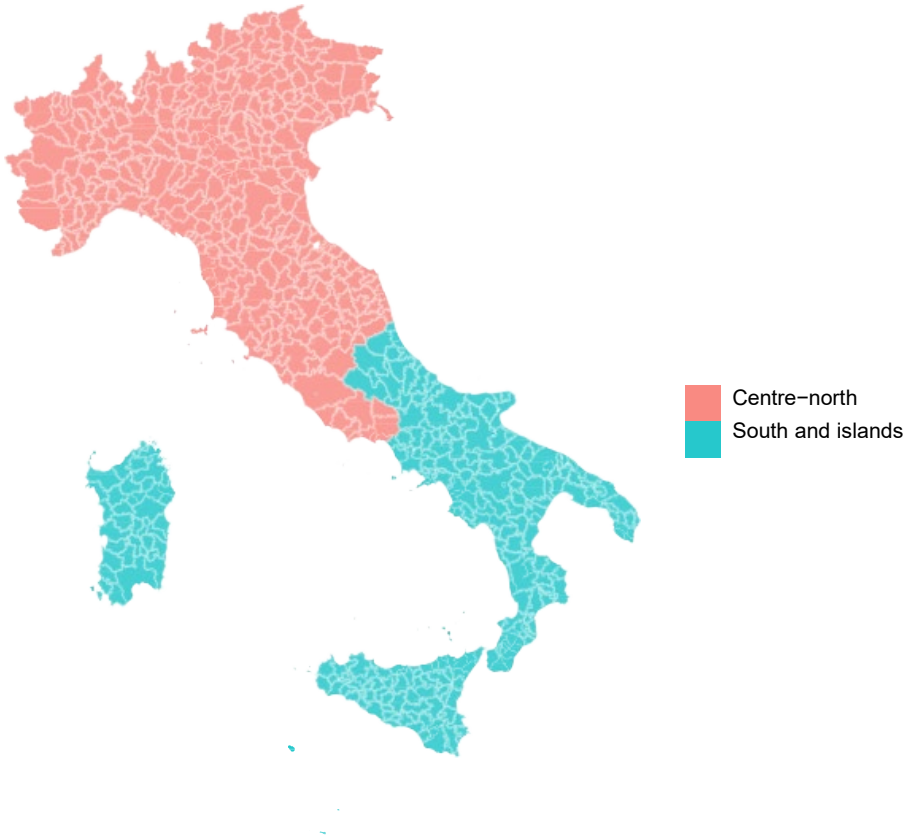
$$TFP_{r,s,t} = \delta_s + \theta_{r,s,t}$$

$$TFP_r = \sum_{t=1995}^{t=2015} \left(\frac{TFP_{r,t}}{T} \right)$$

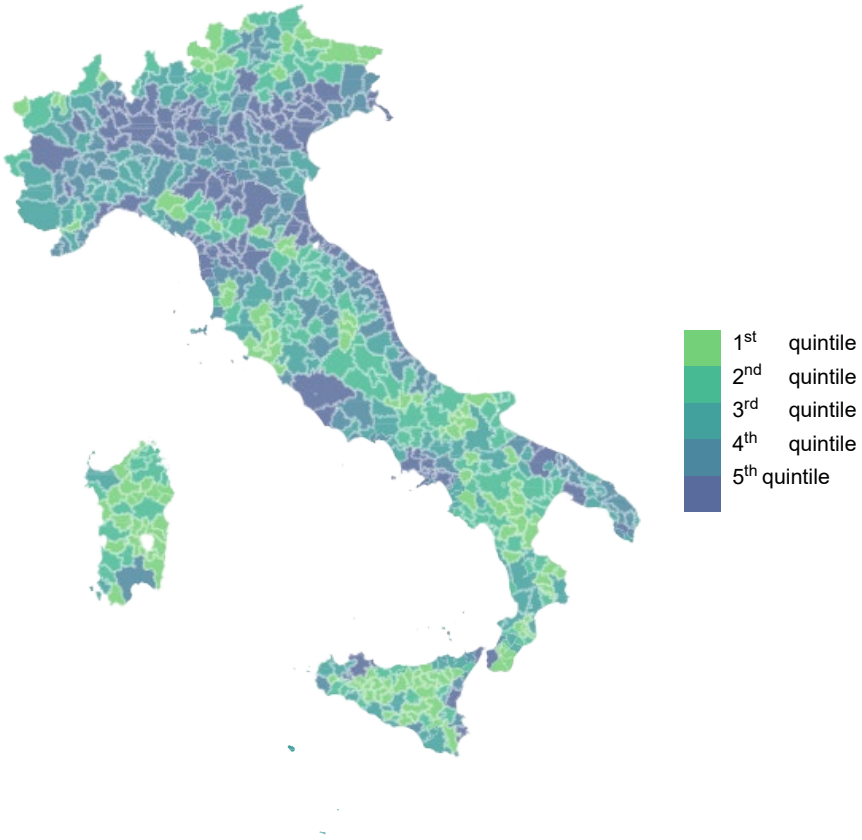
Local Labor Market Areas

- LLMA are built by aggregating municipalities based on their spatial contiguity, and the self-containment of daily commuting flows for work reasons
- Such a partition is produced every ten years as the data that are needed in this respect comes from the census of the population and economic activities carried out by ISTAT at the beginning of each decade
- The adopted observation unit represent an ideal reference for the analysis of agglomeration economies since many of the externalities mentioned by the theory tend to occur at the level of a local labor market

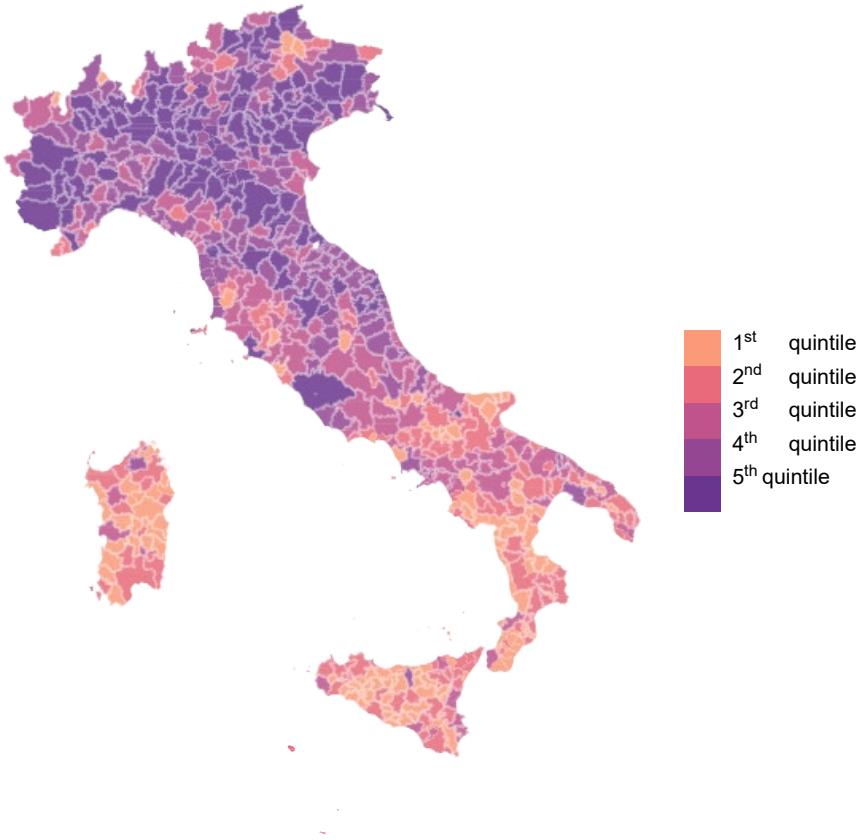
LLMA by geographical breakdown



Spatial distribution of total employment density by LLMA



Spatial distribution of total factor productivity by LLMA

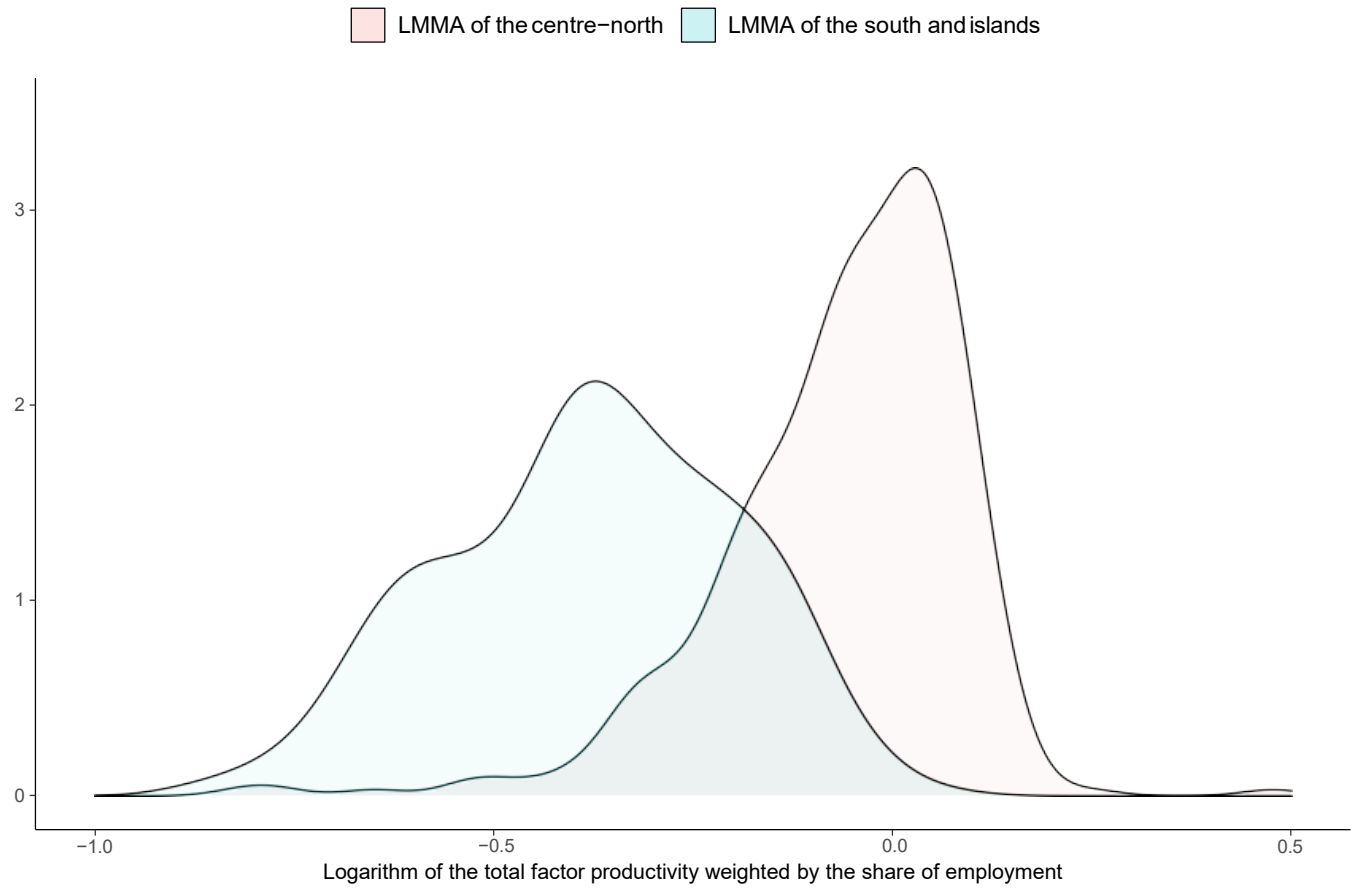


Differences in TFP and total employment density by macrozone

TFP	Mean	1st perc.	25th perc.	50th perc.	75th perc.	99th perc.
Centre-north	0.947	0.522	0.868	0.963	1.044	1.166
South and islands	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)
Empl. density	Mean	1st perc.	25th perc.	50th perc.	75th perc.	99th perc.
Centre-north	62.038	3.748	15.383	36.307	74.201	318.607
South and islands	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)

Note: differences in TFP and total employment density across LLMAs located in each macrozone are tested by regressing each variable 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; differences for other percentiles of the distributions are tested with quantile regressions; stars from one to three indicate statistical significance at 10%, 5%, and 1% levels.

Distributions of TFP by geographical breakdown



Scatterplot of density and TFP

Baseline and Instruments for density

$$Log(TFP_r) = \gamma \log\left(\frac{L_r}{S_r}\right) + \rho\left(\sum_{j \in B(r)} log\left(\frac{L_j}{S_j}\right)\right) + \alpha\,(shr\ of\ coastal\ land) + \beta \log(alt_r) + spat.fixed\ effects \ldots$$

Instruments History: (log of) *Population density in 1921*; Geology: *ruggedness and depth to rock*; Climate*: (log) of rainfall and the average temperature in 1921*

Empirical results: estimation with instrumental variables (2SLS)

Model	(1)	(2)	(3)
Total employment density (ln)	0.065∗∗∗	0.059∗∗∗	0.062∗∗∗
	(0.008)	(0.008)	(0.009)
Contiguous total employment density	0.004∗∗∗	0.004∗∗∗	0.004∗∗∗
	(0.001)	(0.001)	(0.001)
Spatial controls	4 Macro-areas	20 Regions	110 Provinces
Observations	610	610	610
Partial adj. R-squared (density)	0.673	0.663	0.537
Partial adj. R-squared (contiguous density)	0.871	0.878	0.854
Minimum eigenvalue statistic	157.423	153.024	103.421
Overidentifying restrictions test (p-value) Endogeneity	0.001	0.335	0.242
test (p-value)	0.002	0.000	0.003

Note: the dependent variable is the logarithm of the TFP weighted by the share of employment in the sector and estimated with the cost of labor; all the regressions include the constant term as well as controls for the share of coastal surface and the altitude of the LLMA; the densities of both the focal and contiguous LLMA are instrumented with the corresponding lagged population densities.

Robustness

- Functional form: adding quadratic term for density
- Market potential as an alternative to contig. LLMAs
- Alternative definitions of the agglomeration variable
- Estimation with alternative measures of TFP
- Additional controls based on aggregate human capital measures

Contributions to explained TFP variance across LLMA : Shapley decomposition

Regressor	Shapley value	Percentage
:Macro-areas	0.2807	42.8%
Total employment density (ln)	0.2006	30.5%
Contiguous total employment density	0.1073	16.3%
Share of coastal surface	0.0422	6.4%
Altitude (ln)	0.0259	4.0%
Total	0.6568	100.0%
Regions	0.3309	48.0%
Total employment density (ln)	0.1908	27.7%
Contiguous total employment density	0.1026	14.9%
Share of coastal surface	0.0402	5.8%
Altitude (ln)	0.0248	3.6%
Total	0.6894	100.0%
Provinces	0.3957	53.6%
Total employment density (ln)	0.1778	24.1%
Contiguous total employment density	0.0989	13.4%
Share of coastal surface	0.0406	5.5%
Altitude (ln)	0.0250	3.4%
Total	0.7380	100.0%

Empirical results: differences in aggl. returns across macroreas

Model	(1)	(2)	(3)
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 adj. R-squared (density)	0.713	0.721	0.555
Partial adj. R-squared (contiguous density)	0.865	0.922	0.836
Partial adj. 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

Agglomeration and selection effects

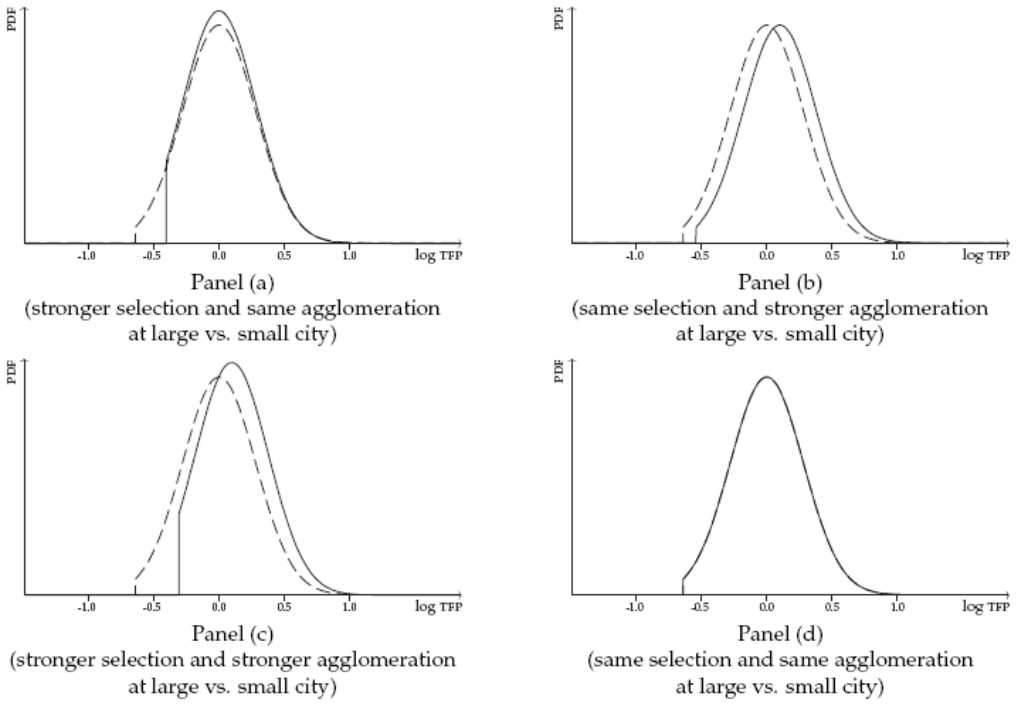


Figure 1: Log productivity distributions in large (solid) and small cities (dashed)

Source: Combes et al (2012)

Empirical results: 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
Firms in northern LLMAs	148,994	148,994	148,994	148,994
Firms in southern LLMAs	39,123	39,123	39,123	39,123
Total number of firms	188,117	188,117	188,117	188,117

Final remarks

- Intensity of agglomeration economies in Italy is similar to that for other developed countries
- Lack of agglomeration contributes to explain a part of north-south tfp gap in manufacturing
- Agglomeration returns seem to be similar between north and south. Problems are with the lack of agglomeration and not that its returns are lower
- No evidence for lower selection in the southern LLMAs
- To do: controls for sorting (mig. Flows), analysys by subperiods …

Thank you …