



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/223009/>

Version: Published Version

Article:

Corradini, C., Morris, D. and Vanino, E. (2025) Marshallian agglomeration, labour pooling and skills matching. *Cambridge Journal of Economics*, 49 (3). beaf010. ISSN: 0309-166X

<https://doi.org/10.1093/cje/beaf010>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Marshallian agglomeration, labour pooling and skills matching

Carlo Corradini[⊗], David Morris[⊗] and Enrico Vanino[⊗]

Better skills matching has long been proposed as one of the key advantages of agglomeration economies. Yet, support for this improved matching has remained largely founded upon indirect proxies for skills such as wages and education. This paper contributes to the literature by offering novel empirical evidence on the relationship between specific measures of localised skills deficiencies and agglomeration economies, in the form of industrial density. Developing an instrumental variable approach and controlling for unobserved heterogeneity and other region-industry idiosyncratic effects across a panel dataset for the period 2009–2019 in England and Wales, our analysis reveals a positive effect of agglomeration economies in reducing both skills gaps within the employed workforce and skills shortages in the labour market external to the firm. We consider these findings in the context of persistent regional imbalances and the importance of strengthening skills provision within current regional industrial strategies.

Key words: Skills matching, Skills gaps, Skills shortage vacancies, Agglomeration, Labour market pooling
JEL classification: J24, L25, R11, R12

1. Introduction

Since the foundational insights by Alfred Marshall (1890, 1919) on the positive externalities that arise from the spatial co-location of companies, agglomeration economies have come to constitute a foundational concept across the literature. They underlie modern theories of industrial districts and clusters (Becattini, 1990; Porter, 1998), urban growth (Black and Henderson, 1999), new economic geography (Krugman, 1992), evolutionary economic geography (Boschma and Martin, 2010) as well as seminal contributions on the geography of innovation (Jaffe *et al.*, 1993; Audretsch and

Manuscript received 26 October 2023; final version received 26 July 2024

Address for correspondence: Carlo Corradini, Henley Business School, University of Reading, Reading, UK; email: c.corradini@henley.reading.ac.uk

[⊗]Henley Business School, University of Reading, Reading, UK (CC); University of Nottingham Business School, Nottingham, UK (DM); and University of Sheffield, Sheffield, UK (EV). We are grateful for the comments received at the AAG 2019 Annual Meeting. The statistical data used here is from the Office of National Statistics (ONS) and is Crown copyright and reproduced with the permission of the controller of HMSO and Queens Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. The analysis upon which this paper is based uses research datasets which may not exactly reproduce National Statistics aggregates.

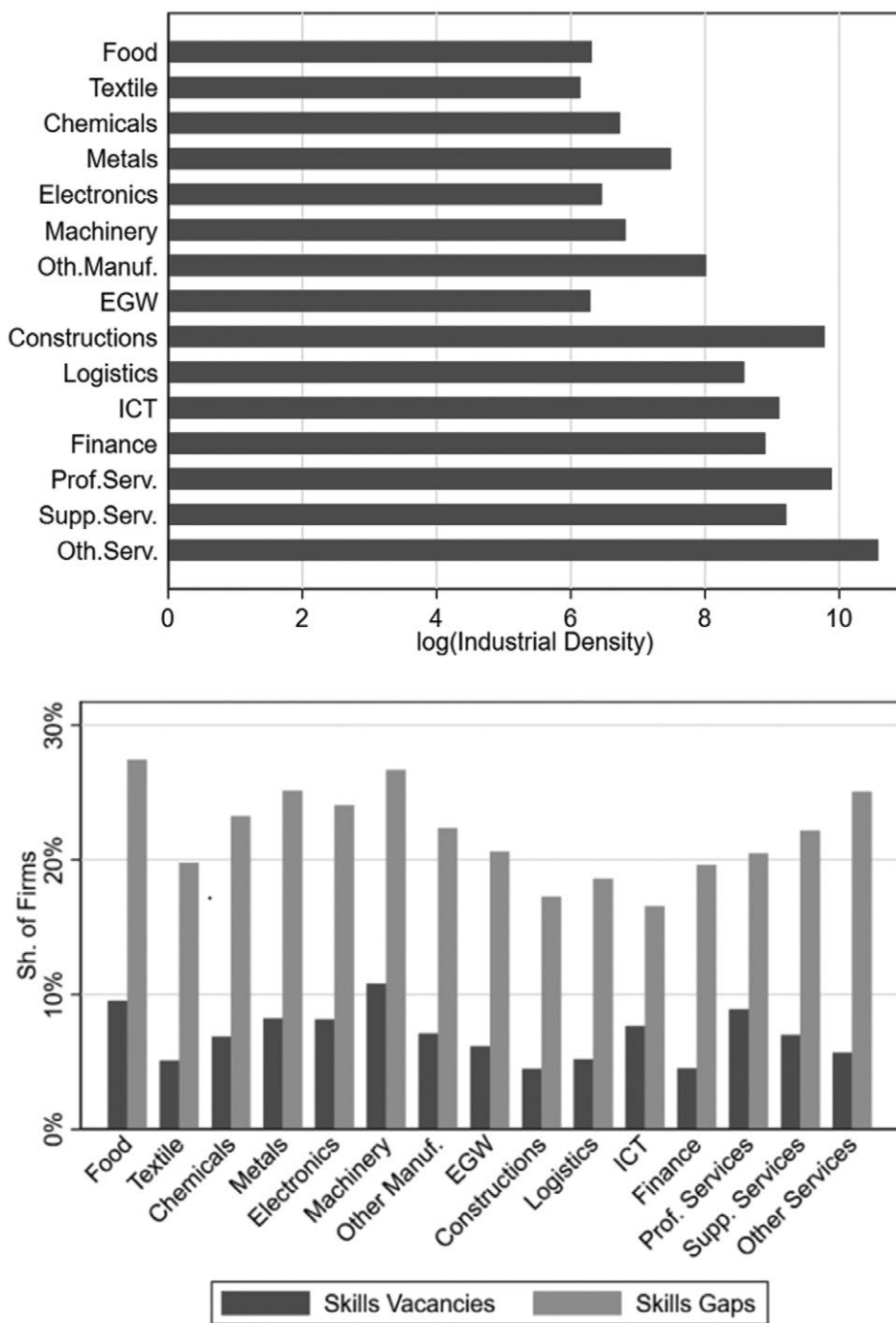


Fig. 1. Industry agglomeration, skills vacancies and gaps. Notes: Statistics elaborated using data from the Employer Skills Survey and the Business Structure Database over the period 2009–2019 for 15 aggregated industries based on the SIC 2007 industrial classification.

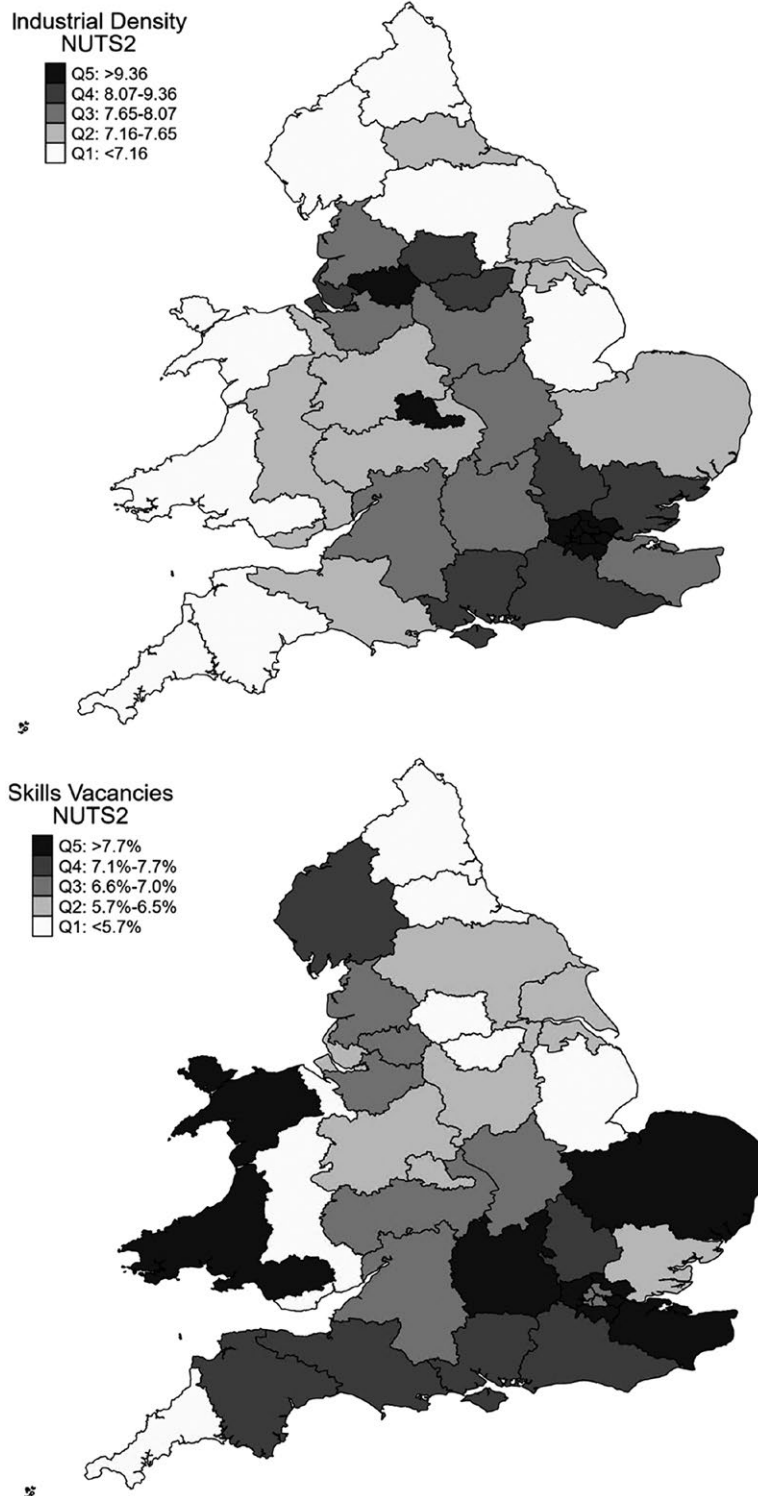


Fig. 2. Regional agglomeration, skills vacancies and gaps. Notes: Statistics elaborated using data from the Employer Skills Survey and the Business Structure Database over the period 2009–2019 for NUTS2 regions in England and Wales.

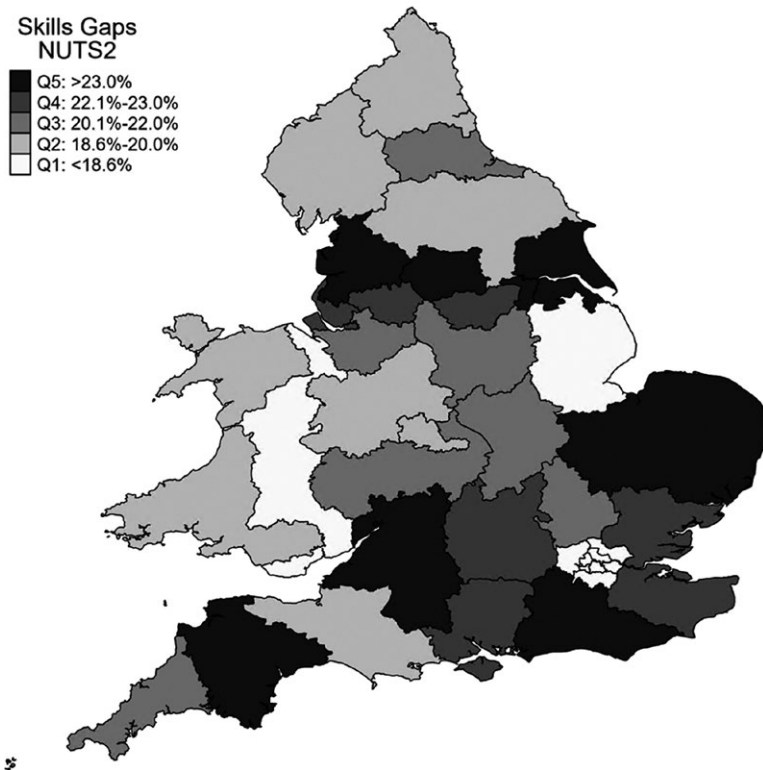


Fig. 2. Continued

mismatch across space. Industrial density is particularly high in regions characterised by large urban centres, mainly in the Greater London area, the West Midlands and the Greater Manchester area. Skill vacancies and gaps are scattered across peripheral regions, with more intense mismatch in the East of England, Wales, and Yorkshire, in line with evidence from previous work (Morris *et al.*, 2020).

3.2 Methodology

To analyse the effect of agglomeration forces on skills mismatch, we start by estimating an ordinary least-squares (OLS) model as follows:

$$y_{rit} = \beta_0 + \beta_1 id_{rit} + \beta_2 sd_{rit} + \beta_3 X_{rit} + \lambda_{ri} + \lambda_{rt} + \varepsilon_{rit}$$

where y_{rit} represents the share of firms in a region r , industry i and time t reporting each of the four different measures of skills mismatch we define in Section 3.1: *skills gaps*, *vacancies*, *hard-to-fill vacancies* and *skills shortages vacancies*. The main parameter of interest will be β_1 , identifying the relationship between the industrial density in a region-industry id_{rit} and the various measures of skills mismatch listed above. In addition, following the Marshallian prediction of agglomeration driven by sharing inputs suppliers, we also analyse the potential spillover originating from the density of related and vertically integrated industries located within the same region which are linked through supply chain relationships (sd_{rit}) (Ellison *et al.*, 2010). This is calculated by

weighting the industrial density of sector j in region r at time t (id_{rit}) by the average share of inputs of production that industries i and j source from each other at the national level (α_{ij}), measured using data from the ONS UK input–output tables for year 2014, and summing it up for all j industries related to i . In addition, we also distinguish between spillovers originating from the density of upstream and downstream industries, where α_{ij} in that case represents the share of inputs of production supplied by sector j over the total demand of inputs of production by sector i , or the share of demand from sector j over total demand for inputs produced by sector i , respectively.

$$sd_{rit} = \sum_{j \neq i} d_{jrt} \times \alpha_{ij}.$$

We control for time-variant region–industry characteristics X_{rit} by including the HHI of market concentration and the average region–industry labour productivity. In addition, to reduce the risk of omitted unobservable bias⁴ we control for region–industry time-invariant characteristics (λ_{ri}) and for any time-variant region-specific characteristics (λ_{rt}) by including fixed-effects.

Although the inclusion of appropriate control variables and extensive fixed-effects mitigates potential issues related to omitted variables bias, there are still threats to the identification of a causal relationship between agglomeration and skills mismatch. As noted in the literature on agglomeration and skills matching (Combes *et al.*, 2008; Glaeser and Resseger, 2010; Berlingieri, 2019), these concerns are due to the potential for reverse causality and further unobserved factors which might be correlated with agglomeration and skills shortages.⁵ To address these issues, we also implement a two-stage least-squares (2SLS) approach, instrumenting our agglomeration variable with a plausibly exogenous instrumental variable. To do this, we use a Bartik (1991) instrument formed by interacting the exposure to a common shift with a differential localised exposure. In our case, we take into account the employment density in a region–industry almost 100 years before the beginning of our period of analysis, using information from the 1921 Census of England and Wales digitised by the National Archives. This provides information on the sector of employment of all residents across census areas in 1921, which we link to the region–industry classification we use in this analysis to identify the initial local conditions—the total employment in industry i and region r in 1921 (e_{ri}^{1921}) divided by the area of region r (a_r) (the ‘share’). We then interact this with the overall number of firms in an industry at the national level in each year in our sample (b_{it}) (the ‘shift’):

$$hid_{rit} = \frac{e_{ri}^{1921}}{a_r} \times b_{it}.$$

As a result, our Bartik instrument (hid_{rit}) would provide the predicted weighted average of the exogenous growth in a given industry nationwide in which the weights are dependent on the initial local distribution. Following the recent methodological literature on the use of Bartik instruments, we assess the validity of our instrument

⁴ In all our estimations, standard errors are robust to heteroskedasticity and clustered at the region–industry level.

⁵ Measurement error is less of a concern in our case, as we use administrative data on the population of businesses in the UK to build our measures of agglomeration, as well as utilising numerous measures of skills shortages to alleviate the issues related to having to rely on self-reported measures from surveys.

Table 2. Relationship between agglomeration and skills vacancies and gaps—2SLS estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skill gaps		Vacancies		Skill-shortage vacancies		Hard-to-fill vacancies	
Ind. density	-0.0256** (0.0123)	-0.0263** (0.0125)	-0.0398*** (0.0119)	-0.0382*** (0.0120)	-0.0294*** (0.00698)	-0.0290*** (0.00707)	-0.0316*** (0.00738)	-0.0311*** (0.00746)
Density spillover	-0.00165 (0.00905)		0.00388 (0.0146)		-0.0116* (0.00702)		-0.00867 (0.0104)	
Downstream density		0.00843 (0.0133)		-0.0189 (0.0123)		-0.0119 (0.00989)		-0.0103 (0.0104)
Upstream density		-0.00474 (0.00833)		0.0108 (0.0101)		-0.00640** (0.00327)		-0.00197 (0.00691)
HH Index	-0.0413 (0.0490)	-0.0438 (0.0492)	-0.0877** (0.0364)	-0.0821** (0.0365)	-0.0284 (0.0283)	-0.0269 (0.0283)	-0.0620* (0.0334)	-0.0605* (0.0335)
Lab. productivity	0.0109* (0.00614)	0.0108* (0.00613)	-0.00137 (0.00761)	-0.00111 (0.00761)	-0.00192 (0.00418)	-0.00185 (0.00418)	-0.00650 (0.00512)	-0.00643 (0.00512)
NUTS2-IND FE	Y	Y	Y	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y
NUTS2-YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,083	2,083	2,081	2,081	2,082	2,082	2,081	2,081
R-squared	0.196	0.196	0.304	0.303	0.343	0.343	0.380	0.380

Notes: Results estimated using a 2SLS model. Robust standard errors clustered at the region-industry level reported in parentheses.

Source: Authors' elaboration based on ESS and BSD data.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3. *Continued*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skill gaps				Vacancies			
	Manufacturing		Services		Manufacturing		Services	
Observations	(0.00967)	(0.00978)	(0.00498)	(0.00504)	(0.0114)	(0.0115)	(0.00603)	(0.00607)
R-squared	797	797	1285	1285	796	796	1285	1285
NUTS2-IND FE	0.436	0.436	0.399	0.399	0.444	0.444	0.44	0.44
YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y
NUTS2-YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Results estimated using a 2SLS model. Robust standard errors clustered at the region-industry level reported in parentheses.

Source: Authors' elaboration based on ESS and BSD data.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

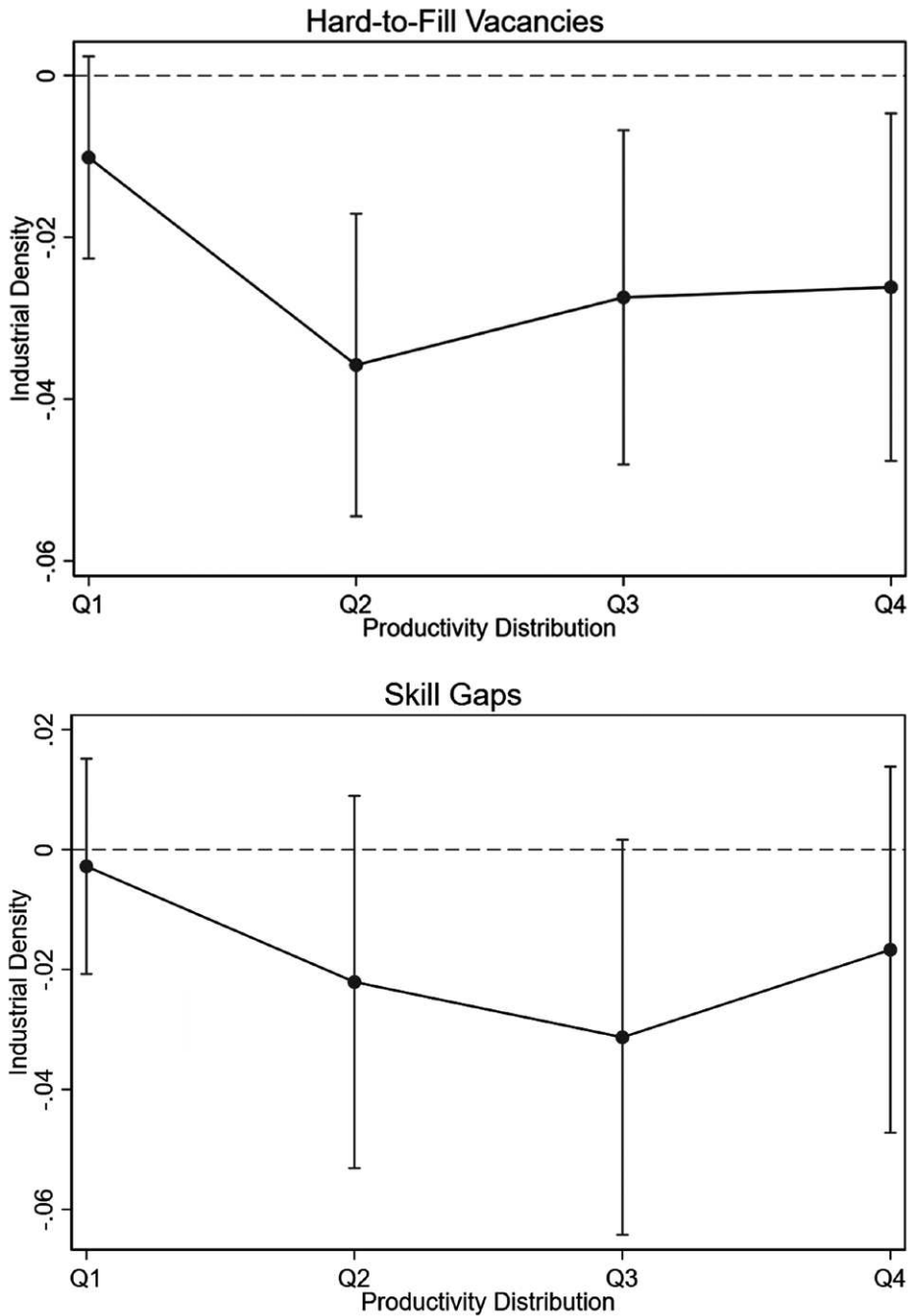


Fig. 3. Relationship between agglomeration and skills vacancies and gaps across productivity distribution—2SLS estimates. Notes: Results estimated using a 2SLS model with robust standard errors clustered at the region-industry level. 95% confidence interval reported. Estimates from bottom (Q1) to top (Q4) quartile of region-industry productivity.

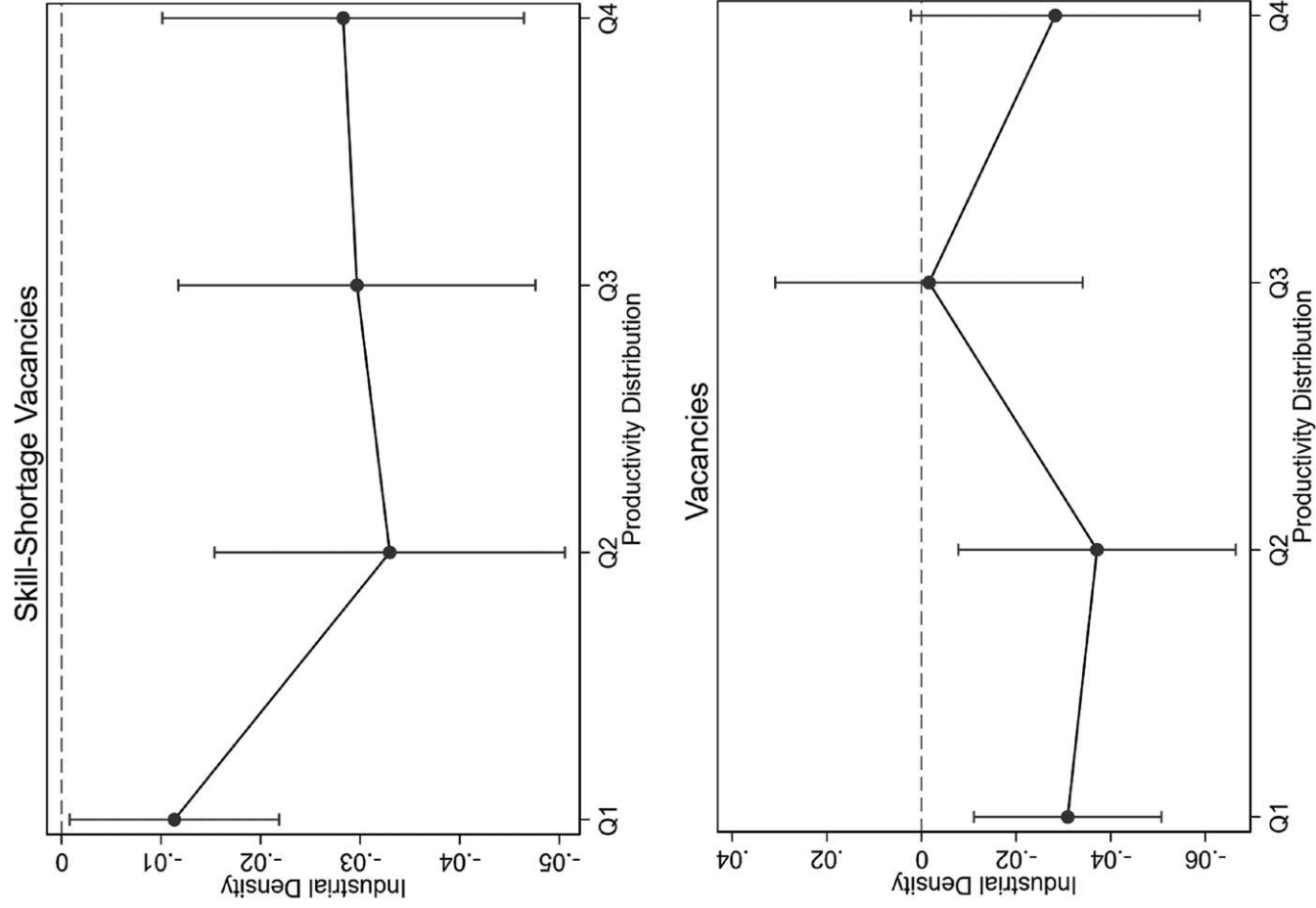


Fig. 3. Continued

Appendix

Table A1. Results of the first-stage regression for the 2SLS estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	Ind. density	Ind. density	Dens. spillover	Ind. density	Downstr. dens.	Upstr. dens.
Hist. ind. density	1.012*** (0.0252)	1.012*** (0.0253)	0.0156** (0.00723)	1.009*** (0.0252)	0.0246** (0.0107)	0.0143 (0.0146)
Hist. dens. spillover		0.00558 (0.0508)	1.001*** (0.0783)			
Hist. downstr. dens.				0.0448 (0.056)	0.908*** (0.0921)	-0.361 (0.308)
Hist. upstr. dens.				-0.0235 (0.0272)	-0.538 (0.429)	1.017*** (0.0781)
HH Index	-0.0478 (0.0967)	-0.0476 (0.0961)	0.0777** (0.0387)	-0.0533 (0.0968)	0.160* (0.0914)	0.00502 (0.0616)
Lab. productivity	0.00149 (0.00898)	0.0014 (0.00897)	0.00840** (0.00426)	0.000887 (0.00892)	0.00571 (0.00685)	0.0129 (0.009)
NUTS2-IND FE	Y	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y	Y
NUTS2-YEAR FE	Y	Y	Y	Y	Y	Y
F-Stat.	1609.18	1609.18	751.748	1609.18	309.146	505.13
Observations	2051	2051	2051	2051	2051	2051
R-squared	0.998	0.998	0.999	0.998	0.996	0.994

Notes: Results of the first-stage of the 2SLS model. Robust standard errors clustered at the region-industry level reported in parentheses.

Source: Authors' elaboration based on ESS and BSD data.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2. Relationship between agglomeration and skills vacancies and gaps—OLS and 2SLS estimates using employment density

OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skill gaps		Vacancies		Skill-shortage vacancies		Hard-to-fill vacancies	
Empl. density	-0.0174* (0.00943)	-0.0174* (0.00944)	-0.0158* (0.00829)	-0.0156* (0.00825)	-0.0155*** (0.00535)	-0.0156*** (0.00535)	-0.0149** (0.00614)	-0.0150** (0.00614)
Density spillover	-0.0133 (0.018)		0.00265 (0.0321)		-0.0149 (0.0139)		-0.0105 (0.0195)	
Downstream density		-0.00826 (0.0217)		-0.0189 (0.0315)		-0.00478 (0.0129)		-0.00275 (0.0168)
Upstream density		-0.00547 (0.0126)		0.0154 (0.0179)		-0.00926 (0.0086)		-0.007 (0.0105)
HH Index	0.00878 (0.0478)	0.00878 (0.0477)	0.0161 (0.0368)	0.0185 (0.0366)	0.0446* (0.0266)	0.0441* (0.0267)	0.0398 (0.0275)	0.0393 (0.0276)
Lab. productivity	0.0109 (0.00714)	0.0109 (0.00714)	0.00376 (0.00729)	0.0033 (0.00726)	-0.00183 (0.00502)	-0.00173 (0.00504)	-0.00561 (0.00552)	-0.00553 (0.00555)
Observations	2412	2412	2409	2409	2409	2409	2410	2410
R-squared	0.428	0.428	0.561	0.562	0.484	0.484	0.502	0.502
2SLS	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Skill Gaps		Vacancies		Skill-Shortage Vacancies		Hard-to-Fill Vacancies	
Empl. density	-0.0299** (0.0147)	-0.0302** (0.0148)	-0.0438*** (0.0147)	-0.0416*** (0.0148)	-0.0328*** (0.00839)	-0.0323*** (0.00856)	-0.0342*** (0.00885)	-0.0334*** (0.00911)
Density spillover	-0.00089 (0.0171)		-0.0501** (0.0225)		-0.0165 (0.0145)		-0.017 (0.0164)	
Downstream density		0.00706 (0.0227)		0.0188 (0.0176)		-0.00175 (0.0222)		-0.03 (0.0245)
Upstream density		-0.00303 (0.0117)		-0.0527* (0.028)		-0.0223** (0.0114)		0.00111 (0.0131)
HH Index	0.0322 (0.058)	0.0314 (0.0581)	0.0213 (0.0479)	0.0268 (0.0473)	0.053 (0.0339)	0.0544 (0.0337)	0.0235 (0.0382)	0.0256 (0.0379)
Lab. productivity	0.0106* (0.00619)	0.0106* (0.00619)	-0.0018 (0.00784)	-0.00192 (0.00779)	-0.00253 (0.00418)	-0.00256 (0.00419)	-0.00708 (0.00516)	-0.00713 (0.00516)

Table A2. *Continued*

OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skill gaps		Vacancies		Skill-shortage vacancies		Hard-to-fill vacancies	
F-stat	2302.81	505.13	2302.81	505.13	2302.81	505.13	2302.81	505.13
Observations	2083	2083	2081	2081	2082	2082	2081	2081
R-squared	372	372	372	372	372	372	372	372
NUTS2-IND FE	Y	Y	Y	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y
NUTS2-YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Results estimated using OLS and 2SLS models. Robust standard errors clustered at the region-industry level reported in parentheses.

Source: Authors' elaboration based on ESS and BSD data.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3. *Continued*

OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Skill gaps		Vacancies		Skill-shortage vacancies		Hard-to-fill vacancies	
Observations	20252	20252	20252	20252	20252	20252	20252	20252
R-squared	0.069	0.069	0.127	0.127	0.093	0.093	0.111	0.111
TTWA-SIC2 FE	Y	Y	Y	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y
TTWA-YEAR FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Results estimated using OLS and 2SLS models. Robust standard errors clustered at the region-industry level reported in parentheses.

Source: Authors' elaboration based on ESS and BSD data.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.