Gran Sasso Science Institute (GSSI), PhD in Urban Studies and Regional Sciences.
Joint-Degree with Universidad Autonoma de Barcelona (UAB), PhD in Applied Economics.

Productive Factors (Mis)Allocation across regions and cities

Francesca Ghinami

Supervisor:

Sandro Montresor, Full Professor of Applied Economics, *Gran Sasso Science Institute*

Co-Supervisor:

Miquel-Àngel Garcia-López, Full Professor of Economics, *Universitat Autònoma de Barcelona*

Dissertation submitted to the Social Sciences Area of the Gran Sasso Science Institute, and to to the Department of Applied Economics of the Autonomous University of Barcelona, in partial fulfillment of the Requirements for the Joint-Degree of Doctor of Philosophy in Urban Studies and Regional Science (GSSI) and in Applied Economics (UAB).

Declaration

I certify that this thesis is an original work carried out by me under the joint guidance of my supervisors, other than where I have clearly indicated that it is the work of others, as specified in the statement of conjoint work. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Statement of conjoint work

I hereby declare that Chapter 1 is jointly co-authored with Sandro Montresor, Full Professor at Gran Sasso Science Institute.

Francesca Ghinami

March 31, 2023

Abstract

This doctoral thesis focuses on the spatial distribution of productive factors and provides new original evidence about the role that spatial frictions - costs related to distance - have for it and for its degree of allocative efficiency, meant as the output maximizing optimal distribution of scarce resources across users.

This dissertation is motivated by the high policy and research relevance of the spatial distribution of productive factors in presence of distance related costs. The strength of agglomeration economies has been steadily increasing in the last century, owing to the reduction in transport and trade costs associated with economic integration and technological advancements. As a result, smaller and peripheral cities and regions tend to face productivity and population decays, struggling in attracting and retaining productive factors. As highlighted by the extant literature, this outcome is not always socially desirable in terms of aggregate welfare and productivity, as the external nature of agglomeration spillovers is likely to lead to inefficient spatial outcomes.

The thesis addresses this general topic with respect to the efficiency of the spatial distribution of specific productive factors, and to the role of spatial frictions on it, and analyses three research questions in as many chapters: i) the spatial distribution of risk-capital and the role of proximities in reducing regional equity gaps, in Chapter 1; ii) the distribution of workers across-cities as influenced by the adoption of remote-work arrangements, in Chapter 2; iii) the spatial disparities in firms' ability to efficiently allocate human and physical capital, and the productivity

and output losses that these entail, in Chapter 3. By working across these three research questions, the thesis aims at reaching two main objectives. Firstly, it aims at adding to the research at the frontier about how to investigate the influence that spatial frictions and their mitigation have on the mobility and spatial distribution of productive factors. Secondly, it aims at providing new estimates of the magnitude of the welfare and productivity losses that this interplay determines.

Chapter 1 aims to investigate the role that different forms of proximity have in the access to Venture Capital (VC) by Innovative Startup Companies (ISC). By combining VC with economic geography literature, we claim that, while tangible (spatial) proximities are relevant for successful VC deals with young innovative firms, different kinds of intangible proximity between them also matter and could explain the absence of location-mirroring relationships. By referring to the population of Italian innovative startups, and by tracing the VC investments occurred in them, we find that tangible proximities account for this matching, but more in functional than in geographical terms, showing an expected concave relationship with it. Industrial proximity between the two actors matters too, with an atypical convex pattern, and makes the role of functional proximity less binding for the matching. The greatest correlation emerges with respect to a relational kind of proximity, due to the closeness between partners in organisational and social terms. Its effect grows exponentially with the level of proximity, but relational proximity does not moderate the impact of functional proximity on the matching. Research and policy implications are drawn accordingly.

Chapter 2 explores the effects of the adoption of remote-work on the size and competitiveness of US cities. Contributing to the revamp of debate on the topic stimulated by the Covid-19 pandemic, it first predicts these effects by proposing a

Quantitative Spatial Economic model with shipping and commuting costs. Then it evaluates the counterfactual changes in population distribution across US cities given remote-work adoption. Results show that, if remote-work was to be adopted to its full potential, according to each city's share of employment in remotely-performable occupations, larger cities would grow in size, welfare, and productivity. This result is the sum of a number of agglomeration forces, linked to the initial consumption and productivity advantages, to the higher frictions (and savings) entailed in their size, and to the higher share of workers in remote-workable occupations that larger cities tend to display. The new spatial equilibrium is found to entail generalised welfare gains that would also benefit smaller and shrinking cities, due to the pro-competitive effect of trade.

Chapter 3 investigates the spatial heterogeneity that factors misallocation reveals in nine EU-member countries (Germany, France, Austria, Italy, Spain, Portugal, Czech Republic, Slovenia and Poland) during the years 2011-2020. Misallocation, meant as the degree of efficiency with which inputs are allocated across firms, is increasingly regarded as the main source of aggregate productivity and income differences across countries. Nevertheless, its within-country spatial and regional dimensions are still largely overlooked, notwithstanding numerous reasons for allocative efficiency to vary across different administrative units. This article aims at filling this gap by firstly performing an exploratory analysis of allocative efficiencies at different levels of territorial aggregation (NUTS0, NUTS1, NUTS2 and NUTS3). Secondly, it provides evidence for the across-regions disparities in allocative efficiency to account for large shares of aggregate misallocation for all the examined European countries (up to 28% at NUTS3 level). Finally, it investigates and finds support for the hypothesis that variations in local institutional quality may help explaining regional differences in allocative efficiencies.

Table of contents

List of tables

List of figures

Acknowledgements

I am deeply grateful to my supervisors, Sandro Montresor and Miquel-Angel Garcia-Lopez for their thorough scientific and emotional support, for their patience and presence throughout this enriching path.
All remaining errors and omissions are my own.
A huge thank you goes to Rosella, Gabriel, David, Riccardo, Maria, Kenneth, Marcela, Alessia, Martin, Yadira, Elisa, Piergiorgio, Maria Teresa, Machus, Marco, Eddi, Marigio and many others, who shared with me their knowledge, experience, passion, good mood and memorable coffee breaks. I also want to thank all those people, that may or may not call themselves artists, who dedicate part of their lives to gifting strangers with a safe escape for their souls during hard times.

Previous versions of the chapters included in this thesis were presented at the following conferences: 5th Geography of Innovation Conference, Stavanger (01/2020); RSA-Regional Studies Association E-Festival (06/2021); 60th ERSA-European Regional Studies Association Congress (08/2021); XLII AISRe-Italian Regional Science Association Congress (09/2021), DRUID21, Copenhagen (10/2021), SEA-Spatial Econometrics Association XVI World Conference, Warsaw (06/2022); AACR-XLVII International Conference on Regional Science, Granada (10/2022); Seminars of the Regional Quantitative Analysis Group (UB-AQR) Universitat de Barcelona (12/2022).

To my family, Luciana, Roberto, Riccardo and Stefano,

my cherished grandmothers and dear aunties,

who are like angels to me.

Introduction

I Overview of the thesis

This doctoral thesis focuses on the spatial distribution of productive factors and provides new original evidence about the role that costs related to distance have for it and for its degree of allocative efficiency, meant as the optimal distribution of scarce resources across users in order to maximize the output of production. The thesis addresses this general topic with respect to the spatial distribution of specific productive factors and analyses three research questions in as many chapters: i) the spatial distribution of risk-capital and the role of proximities in reducing regional equity gaps, in Chapter 1; ii) the distribution of workers across-cities as influenced by the adoption of remote-work arrangements, in Chapter 2; iii) the spatial disparities in firms' ability to efficiently allocate human and physical capital, and the productivity and output losses that these entail, in Chapter 3. By working across these three research questions, the thesis aims at reaching two main objectives. Firstly, it aims at adding to the research at the frontier about the influence that spatial frictions and their mitigation have on the mobility and spatial distribution of productive factors. Secondly, it aims at providing new estimates of the magnitude of the welfare and productivity losses that this interplay determines. This dissertation is motivated by the high policy and research relevance of the spatial distribution of productive factors in presence of distance related costs and agglomeration economies. The strength of agglomeration economies has been steadily increasing in the last century, owing to the reduction in transport and trade costs associated with economic integration and technological advancements (Thisse, 2009). As a result, smaller and peripheral cities tend to face productivity and population decays, struggling in attracting and retaining productive factors. As highlighted by the extant literature, this outcome is not always socially desirable in terms of aggregate welfare and productivity (Desmet and Rossi-Hansberg, 2014a), as the external nature of agglomeration spillovers is likely to lead to inefficient spatial outcomes (Fajgelbaum and Gaubert, 2020; Fujita and Thisse, 2013) which policy makers are asked to address. Can the firms' ability to source productive factors outside main cities be improved?

The topic at stake is also and above all relevant from an academic perspective, as revealed by the intensive research efforts its analysis has attracted across different research streams, which the present work originally proposes to bridge among them. In doing that, the thesis identifies important gaps in the literature, of both a methodological and empirical nature, to which its filling it originally contributes as illustrated in the following sections.

II Background literature

Along its three chapters, the thesis mainly draws on and contributes to three bodies of literature, focusing on, respectively: i) different kinds of proximity; ii) agglomeration, factors distribution, and frictions; iii) factors misallocation.

II.i Different forms of proximity

Spatial proximity is the main requirement for agglomeration economies to be at work. Numerous studies have in fact highlighted and estimated the strong distance-decay experienced by knowledge spillovers (Caragliu and Nijkamp, 2016) and agglomeration economies (Graham *et al.*, 2010) such as the effects of labour pooling and matching on wages and productivity (Dauth *et al.*, 2019). Duranton (1999) argues that, as a perfect mobility of factors, information and goods would make cities cease to exist, the constraining role of spatial frictions in the pre-and-industrial eras became the main enabler of cities' sustained growth in the post-industrial one. The author defines proximity (especially 'personal physical one') as the main engine of the post-industrial growth of cities, which could otherwise 'disappear' thanks to technological improvements reducing the tyranny of distance in production and trade. The author argues that this is rooted in the increase of importance of tacit knowledge in production, and in the low substitutability of face-to-face contacts with telecommunication technology for its transmission. However, spatial proximity is not the only kind of proximity that matters for the spatial distribution of factors. Regional science and regional innovation studies have recently considered other proximity dimensions than the spatial one - such as organizational, social, technological and institutional ones [1] - and other extents of it – like permanent versus temporary. In these streams of research, the analysis of the non-spatial dimensions of proximity has mainly focused on innovation (Doran *et al.*, 2012; Agrawal *et al.*, 2008), learning (Torre, 2008; Marrocu *et al.*, 2014) and on industrial organization topics like M&A deals (Boschma *et al.*, 2016), with few more works considering the effects of tempo-

1 If not otherwise stated, we will refer to Boschma (2005)'s definitions for each proximity dimensions.

rary proximity on entrepreneurship (Gossling and Knoben, 2011), clusters (Ramirez-Pasillas, 2010) and productivity (Mariotti *et al.*, 2015). Conversely, as highlighted by Torre (2019), there is still a gap in the application of such multi-dimensional concept of proximity to the field of economic development. And this is unfortunate, as such an application could help shed a light on economic relations and local-regional divergent development path and policies. Different streams of study have tried to fill this gap so far. A first step in this direction is represented by the work of Martin and Simmie (2008), who reviewed the concept and theories of urban competitiveness to investigate the theoretical and empirical evidence on the role that proximity, mainly geographical and organizational (as in Torre and Rallet, 2005), plays in it. Similarly, endogenous growth theory allowed a shift towards the importance of knowledge creation and diffusion (Lucas, 1988) in explaining, growth, convergence (Romer, 1994) and technological change (Romer, 1990). This shift accompanied the standard focus on spatial proximity towards that on cognitive, organizational and technological proximity that informed the theories on the role of economic complexity on growth and development (Hidalgo and Hausmann, 2009). However, while extensively considered in the analysis of innovation and regional growth, the influence of proximity is likely to go beyond knowledge spillovers, to also affect ownership and investment decisions (Zhao and Jones-Evans, 2017; Boschma *et al.*, 2016; Bathelt and Gluckler, 2003) and residential mobility (Büchel *et al.*, 2020). In the light of that, the complementary (Marrocu *et al.*, 2014) and/or substitutive (Singh, 2005) relations between the tangible (spatial) and non-tangible dimensions of proximity documented with respect to knowledge-spillovers, could apply to other economic processes, like that of the financial capital allocation investigated in the first Chapter of this thesis. A recent stream of literature refers indeed to the multiple dimensions of proximity in accounting for the existence of a big-city or local bias in the access to finance, both for equity and debt capital (Lee and Luca, 2019; Lee and Brown, 2017). This bias contradicts the standard view of financial capital as mostly footloose and claims for further research on the specific frictions and imperfections affecting its spatial flows. Some first bits of evidence in this stream, to which this thesis contributes, showed the relevance of social, organizational and technological proximity in overcoming local-biases in risk-capital flows and predicting investment decisions (Hermann *et al.*, 2016; Berchicci *et al.*, 2011).

II.ii Agglomeration, factors distribution and spatial frictions

The thesis extensively refers to the streams of literature that, by building on international trade and location theories, regional and urban economics, analyses the optimal size of cities, the mechanisms behind factors misallocation and distortions, and the aggregate impact of the trade-off between agglomeration economies and congestion costs. As is well known, agglomeration economies identify the phenomenon that, at least from Marshall (1890, 1919) onward, has been thoroughly studied and documented [2] as the positive externalities that originate from the spatial clustering of firms and workers (Behrens and Robert-Nicoud, 2015b). Urban and regional economists have extensively investigated the causes for cities formation and growth, looking at either market imperfections, heterogeneous space[3] or localized externalities (Ottaviano and Thisse, 2004), *id est* at the interaction among different kinds of increasing returns and of spatial frictions (Combes *et al.*, 2005).

2 See Combes and Gobillon (2015) for an extensive review of the empirical evidence about agglomeration economies.

3 Some examples of first-nature driven endogenous agglomeration models are Fujita and Mori (1996) on the role of natural ports and Krugman (1993) on centrality.

It is here important to notice that, in absence of non-convexities or of spatial frictions [4] (and, if considered, of localised amenities), location would be indifferent for agents. Spatial frictions and increasing returns are thus essential in understanding spatial equilibria, and as such were the ones that allowed to reproduce endogenous cities formation in general equilibrium models. This result was first accomplished by New Economic Geography (NEG), a stream of literature that following the seminal works by Krugman (1980, 1991) and through the development of general equilibrium models of monopolistic competition, evaluates the impact market structure and transport costs on the distribution of economic activities. These models allowed to explain how the exponential drop experienced by shipping costs in the last century unleashed the potential of comparative advantages in presence of scale economies, producing the rise in spatial inequalities observed throughout the last two centuries (Thisse, 2009). The main agglomeration force identified was the so-called home-market-effect (HME), allowing larger markets to disproportionately attract firms in imperfectly competitive industries. However, while able to explain cities formation and the agglomeration of economic activities, given its simplicity in terms of internal and external geography's representation, agents heterogeneity[5] and spatial frictions (reduced to mere overall transport costs), the NEG literature has not been able to offer insights in presence of multiple equilibria nor to be calibrated for empirical applications (Gaspar, 2018) thus lacking context-specific policy implications

4 With zero transport costs and increasing returns, one large firm could potentially and profitably serve the whole demand for a specific good, while for the spatial impossibility theorem, a competitive equilibria involving trade could not rise in an homogeneous space with positive transport costs. See Fujita and Thisse, 2002 for more details.

5 When accounting for heterogeneity in productivity, since higher competition in larger markets acts as a selection channel, NEG was able to prove that the advantage of core regions/cities could be further reinforced by the attraction of most productive agents and firms (Melitz, 2003; Ottaviano, 2011), even if in a more sketched dimensions than its successors, bridging NEG with Urban economics (Melitz and Redding, 2014; Fujita and Thisse, 2013).

(Behrens and Robert-Nicoud, 2011). The recently developed field of Quantitative Spatial Economics (QSE) [6], originally proposes a group of structural models that, even if reducible to the two oldest endogenous location ones (i.e. Helpman, 1998 and McFadden, 1974, as shown by Behrens and Murata, 2021), extended the NEG framework to solve its limited predictive power in terms of city-size and composition, as in factors distribution and sorting. Indeed, QSE models present a number of unprecedented features, such as multiple spatial frictions (e.g. endogenous housing prices, shipping and/or commuting costs), complex and rich internal and external geographies, localized amenities, heterogeneous preferences and agents. It should be noted here that the consideration of spatial frictions beyond shipping costs is important[7] as it permits to reproduce partial agglomerative outcomes (Fajgelbaum and Gaubert, 2020) through their distinct evolution in terms of price and weight in production and trade. Moreover, explaining and predicting cities' size-distribution and competitiveness, that is the attraction of production factors (or why firms and workers decide to locate in a specific city), requires a detailed analysis and quantification of urban externalities, and of their static and dynamic effects (Desmet and Rossi-Hansberg, 2014b). QSE has also built on the work of urban economists about production and demand side urban externalities and about the micro-foundation of agglomeration economies (Duranton and Puga, 2004), that investigate the comparative advantage of larger metropolitan cities with respect to smaller ones. In particular, larger cities enjoy the so-called "urban productivity premium", exhibiting higher GDP per capita than smaller ones (Behrens and Robert-Nicoud, 2015b). The explanations behind

6 Redding and Rossi-Hansberg (2017) offer an extensive review of the field.

7 Indeed, Fujita and Thisse (2013) inverted Krugman (1991)'s results through the inclusion of commuting costs in its core-periphery model, showing that the predicted spatial equilibrium depends heavily on the nature of the included spatial frictions.

urban productivity (and wage) premium are in turn found in the combination of sorting mechanisms – most productive workers choose larger cities as they are more capable to stand the competition and high prices for rents while being attracted by higher wage and education premia – and selection mechanisms – most productive firms can afford high rents in order to enjoy highly productive workers, specialized services, and learning mechanism (Venables, 2011; Eeckhout *et al.*, 2014; Behrens *et al.*, 2014). Highly productive firms and workers can there enjoy better matching for specialized jobs demand and supply, localized learning processes, sharing of inputs (Helsley and Strange, 2002) and of infrastructures[8].

II.iii Factors misallocation

As maintained by Desmet and Rossi-Hansberg (2013), large cities are so for either (or both) first and second nature advantages, *i.e.* having amenities, having been highly productive or run efficiently. In the absence of one of these factors, they would have never grown to the size they are. However, the success of large cities comes at the social and political cost (Rodríguez-Pose, 2017) of the economic decay of disadvantaged areas. In this context, place-based policies[9] often aimed at reducing heterogeneity across cities by targeting declining areas (Neumark and Simpson, 2015; Kline and Moretti, 2014; Glaeser and Gottlieb, 2008). Whether these policies have a viable rationale is however questionable. As also noted by Ottaviano and Thisse (2004), NEG fell short of welfare analysis of the entailed spatial equilibrium, with the exception of punctual contributions[10] such as those of the 'New new Economic

8 See Puga (2010) and Duranton and Puga (2004) for a detailed review of evidences and estimation techniques for agglomeration mechanisms.

9 Examples of these policies include the Structural Fund in the European Union, and the state enterprise zones programs in the United States (Neumark and Simpson, 2015; Ham *et al.*, 2011).

Geography' (Ottaviano, 2011; Baldwin *et al.*, 2003; Charlot *et al.*, 2006). The QSE literature on the optimal size of cities in terms of welfare and productivity, which we have recalled above, is involved in filling this gap. Indeed, it provides evidence about cities being either too small or too large in terms of productive and welfare optimum, depending on the specific distortions and on the determinants of the city-size distribution in place (Desmet and Rossi-Hansberg, 2014a; Behrens and Robert-Nicoud, 2015b; Albouy *et al.*, 2019). As Fajgelbaum and Gaubert (2020) point out, the external nature of agglomeration spillovers is likely to lead to inefficient spatial outcomes, since the impacts of labour supply and demand decisions on city-level efficiency and amenity spillovers are not fully internalized neither by the hiring-firm nor by the worker.

A relevant field of research on these issues, to which the present thesis also refers, is represented by the burgeoning literature on factors misallocation. Factors misallocation occurs whenever productive factors, such as labour, capital, and technology, are not allocated to their most productive uses, leading to lower levels of output. This literature has focused on the effect on aggregate productivity of different spatial and market frictions, like restrictions in the housing markets (Hsieh and Moretti, 2019), sector-specific or place-based policies and taxes (Fajgelbaum *et al.*, 2019; Yang *et al.*, 2017), market segmentation and imperfections (Restuccia and Rogerson, 2017). This body of work has shown how these frictions can affect aggregate output, welfare, and employment (Haltiwanger *et al.*, 2016), firms selection into markets (Aghion *et al.*, 2006) and the propensity and ability to invest in radical innovation (Caggese, 2019). Differences in allocative efficiency have been

10 Combes *et al.* (2005) developed a diagrammatic analytical framework to compare the findings of NEG and urban systems approach to spatial analysis, using it to evaluate a number of policy implications .

held responsible by these studies, in equal measure as the differences in technological adoption, for the persistent disparities in productivity observed across countries (Restuccia and Rogerson, 2017). Several theoretical and empirical reasons lead us to expect that firms' allocative efficiencies also varies across regions, including subnational taxation (Fajgelbaum *et al.*, 2019), local land market regulation (Hsieh and Moretti, 2019), and differences in the quality of local credit markets (Lenzu and Manaresi, 2019) and of local institutions (Lenzu and Manaresi, 2019; Misch and Saborowski, 2020). At the regional level, the literature is still thin and represented by few contributions, signalling the role of agglomeration economies (Fontagné and Santoni, 2019) and of local institutions as drivers of systematic spatial disparities in firms' allocative efficiency (Misch and Saborowski, 2020). However, a full appreciation and measurement of this phenomenon at a within-country level is still missing, a lack that is attributable to the macro-economic field in which the topic originated.

III Research gaps and novelty

In drawing on these three streams of literature, the present thesis contributes to them by identifying and filling some relevant gaps. The theoretical knowledge, the methodological advancements, and the empirical evidence that the thesis obtained in filling these gaps represent its main elements of novelty. A first gap, with respect to the proximity literature under II.i), concerns the role of relational proximity among multiple investors in reducing informational asymmetries and distance-related transaction costs. This role has been largely established by the literature on the geography of Venture Capital investments. However, there is a lack of studies on the relational proximity insisting between investors and target firms. Chapter 1 is, to the best of the author's knowledge, the first study that analyses how relational proximity – the inverse of the professional and ownership network distance insisting between Venture Capital and potential target firms - influences the probability of observing an investment deal. Since the seminal work by Sorenson and Stuart (2001), the multiple extant studies analyzing the role of relational proximity in risk-capital investments mainly looked at the organizational and social ties among the multiple VC firms participating in syndicated investments, finding it can crucially affect the probability that VC invest in spatially (Tykvová and Schertler, 2014; Sorenson and Stuart, 2001), institutionally (Tykvová and Schertler, 2011) or technologically distant firms (Meuleman *et al.*, 2017; Tykvová and Schertler, 2014; Cumming and Dai, 2010). However, the social embeddedness of the target firms was disregarded by the same literature, a gap that the first Chapter of this thesis aims at addressing. This Chapter provides evidence of the significant effect that the relational proximity between VC and target firms has in predicting their successful match.

With respect to the literature about agglomeration, factors distribution, and frictions under II.ii), the thesis identifies two research gaps related to the effect of remote-work adoption on workers' location choices. The first gap pertains to the technical challenge of the treatment of the double-causality inherent in the relationship between residential location and remote-work arrangements. To deal with it, Chapter 2 proposes to exploit an invertible spatial general equilibrium model, to structurally identify workers location preferences given the changes in commuting costs derived from remote-work adoption. The second gap in the analysis of the topic concerns the lack of research on the effect of remote-work on across-cities workers' location decisions, with most studies focusing on within-city frameworks. In that respect, Chapter 2 is the first study that analyses the effect of remote-work on across-cities workers distribution within a spatial general equilibrium framework. With respect to the literature about factor misallocation under II.iii), the thesis identifies a gap in its being largely silent on the within-country spatial characteristics of the phenomenon. To deal with it, Chapter 3 proposes a novel cross-country analysis of regional misallocation performed at different degrees of territorial aggregation. Indeed, Chapter 3 is the first study that proposes a systematic analysis of the across-regions spatial heterogeneity that factors misallocation display within multiple countries. In doing that, the chapter provides novel evidence on the relevance, in terms of aggregate output losses, of the observed spatial disparities in firms' ability to efficiently allocate human and physical capital in production.

IV Outline of the chapters

The thesis is structured in three chapters. While dealing with the same general topic, and having multiple connections among them, the three chapters are self-contained and can be approached in any order.

Chapter 1 aims to investigate the role that different forms of proximity have in the access to Venture Capital (VC) by Innovative Startup Companies (ISC). By combining VC with economic geography literature, it claims that, while tangible (spatial) proximities are relevant for successful VC deals with young innovative firms, different kinds of intangible proximity between them also matter. Intangible proximity includes industrial - the extent to which a VC has already invested in the industry of the ISC - and relational proximity – the inverse of the network distance linking VCs and ISCs through the professional and ownership relations of the firms and their managers, advisors and investors. The hypothesis is that the relevance of intangible proximity could help explaining the absence of a location-mirroring behavior of ISCs looking for risk-capital investments, which in turn concentrates in few main locations. By referring to the population of Italian innovative startups, and by tracing the VC investments occurred in them in the period 2012-2019, the study reveals that tangible proximities account for this matching, showing an expected concave relationship with it. Furthermore, it shows that functional proximity – spatial proximity expressed in travel times to account for the effort that agents must put in place in order to interact – can predict a VC-ISC match more significantly than geographic proximity – the latter being measured as the inverse of geodetic distance, disregarding of transport costs. Industrial proximity between the two actors matters too, with an atypical convex pattern, and makes the role of functional proximity less binding for the matching. The greatest correlation emerges with respect to a relational kind of proximity, due to the closeness between partners in organisational and social terms. Its effect grows exponentially with the level of proximity, but relational proximity does not moderate the impact of functional proximity on the matching. Research and policy implications are drawn accordingly.

Chapter 2 investigates the effects of the adoption of remote-work on the size and competitiveness of US cities. Contributing to the revamp of the debate on the topic stimulated by the Covid-19 pandemic, it first calibrates a Quantitative Spatial Economic model with shipping and commuting costs on US Metropolitan Statistical Areas 2017's data. Then it evaluates the counterfactual changes in population distribution across US cities given remote-work adoption. Results show that, if remote-work was to be adopted to its full potential, accordingly to each city's share of employment in remotely-performable occupations, larger cities would grow in size, welfare, and productivity. This result is the sum of a number of agglomeration forces, linked to the initial consumption and productivity advantages, to the higher urban frictions (and counterfactual savings) entailed in their size, and to the higher share of workers in remote-workable occupations that larger cities tend to display. The counterfactual spatial equilibrium is shown to entail generalised welfare gains resulting from a reduction of firms' markups which, while stronger in larger cities, also positively impacts smaller and declining cities through the pro-competitive effects of trade.

Chapter 3 investigates the spatial heterogeneity that factors misallocation reveals in nine EU-member countries (Germany, France, Austria, Italy, Spain, Portugal, Czech Republic, Slovenia and Poland) during the years 2011-2020. Misallocation, meant as the degree of efficiency with which inputs are allocated across heterogeneously productive firms within each sector, is increasingly regarded as one main source, together with technological adoption, of aggregate productivity and income differences across countries. Nevertheless, its within-country spatial and regional dimensions are still largely overlooked. This article aims at filling this gap by firstly performing an exploratory analysis of allocative efficiencies at different levels of territorial aggregation (NUTS0, NUTS1, NUTS2 and NUTS3). Secondly, it provides evidence for the across-regions disparities in allocative efficiency to account for large shares of aggregate misallocation for all the examined European countries (up to 28% at NUTS3 level). Finally, it investigates and finds support for the hypothesis that variations in local institutional quality may help explaining regional differences in allocative efficiencies.

Despite their being autonomous, the three chapters have some interesting points of contact in dealing with the following transversal issues across them. In particular, all the three chapters include a discussion on the degree of efficiency of the observed spatial distribution of productive factors, and, as far as Chapter 1 and Chapter 2 are concerned, provide punctual evidence on the positive effect that a decrease in spatial frictions may have on it. In Chapter 1, the efficiency of the spatial distribution of risk-capital flows is inferred through the intensity of the observed regional equity gaps. These are measured, as standard in the geography of finance's literature, through the location quotient of VC investment deals given the spatial distribution of potential target firms. Once the heterogeneity in terms of the startups' unobserved characteristics and profitability potential is controlled for, the distance-decay of the probability to attract VC investments is confirmed. However, both the spatial proximity expressed in travel times, and the proximity in professional and investment networks existing between VCs and target firms, are shown to significantly increase the probability of observing risk-capital flows towards secondary regions, reducing the inefficiency implied by regional equity gaps. In Chapter 2 instead, the decrease in average commuting costs deriving from larger remote-work adoption in cities is shown to directly produce welfare and productivity gains. While the productivity gains only occur(through stronger selection) in larger cities gaining population, the consequent larger-cities' drop in markups diffuses in the whole economy through trade, allowing average utility to increase also in shrinking cities. Finally, Chapter 3, while agnostic on the specific role of spatial frictions, provides evidence on the relevance of the spatial dimensions of factors allocative efficiency, finding the latter to vary consistently across regions and that such variation may significantly reduce aggregate productivity.

V Main findings and policy implications

Overall, the thesis provides an across-fields analysis of the role of distance related costs on the spatial distribution of productive factors. It does so with the aim of gaining a better general understanding of this relationship, searching for potential counterweights to agglomeration-related advantages of larger cities with respect to smaller and secondary cities. This aim is grounded on the evidence, provided by the literature and discussed in the three chapters, that observed spatial outcomes are not always efficient. Smaller and secondary cities tend to face productivity and population decays in the face of economic integration, and to struggle in attracting and retaining productive factors. If this outcome were to be found as socially desirable in terms of aggregate welfare and productivity, there would be no purely economic motivations for policy intervention. However, as highlighted in the reference literature above, this is not always the case. The results obtained in the three chapters are quite interesting and rich of policy implications. As for Chapter 1, its main result attains the role of relational proximity, which is found to have a significant positive effect on the probability of a VC investment. The effect of relational proximity is found to be stronger than that of spatial proximity. This result implies that policy makers could improve firms' access to risk-capital in secondary regions by promoting professional and investment networking. Furthermore, given the positive effect of functional proximity, improvement in fast-transport infrastructures across-cities could also reduce regional equity gaps. Chapter 2 finds that, while remote-work is often regarded as a possible way to attracting or retaining population in smaller cities, its main effect could be that of reinforcing agglomeration in larger cities. This results is grounded on one main mechanism: remote-work can reduce the burden of commuting costs, both for remote-workers and for the city-overall, making larger cities that host the largest shares of remotely-performable occupations become more productive. Under the model assumptions implementing remote-work produces a number of socially desirable outcomes, being a generalised drop in average markups that maps into economy-wide welfare gains. However, policy makers should be aware that such outcomes would come at the cost of a shrinking in population and in average productivity in smaller cities.

The main insight of Chapter 3 attains the relevance of across-regions differentials in the ability of firms to efficiently source and allocate human and physical capital. These disparities are found responsible for about a quarter of overall misallocation-related output losses in the manufacturing sector, resulting in lower aggregate productivity in all the considered EU countries. Furthermore, the study shows the magnitude of said territorial inefficiencies to significantly correlate with the quality of local institutions and precisely with the adherence of government spending to the openness, fairness, efficiency, competition and transparency principles. As such, it suggests that disparities in productive allocative efficiency across territories, and the resulting aggregate productivity losses, could be reduced by improving local government spending quality.

VI Limitations and future avenues for research

All the three chapters have been realized with the highest level of accuracy enabled by currently available datasets and existing, though improvable, methodologies. Despite these efforts, the thesis suffers from some limitations, which are detailed in the respective chapters, and here briefly recalled. Chapter 1 controls for one main source of potential endogeneity bias, linked with the unobserved heterogeneity in the quality of startup projects. It does so through the creation of a dyadic sample composed by the sole target-firms that obtained a VC backing and the VC funds that invested in any innovative startup in the same 6 months period. However, a second possible source of bias could be present if more capable entrepreneurs were able to anticipate the importance of spatial and relational proximity on the probability to access Venture Capital investments. In that case, they could define their location, hiring and ownership strategies accordingly. This would render our measures of spatial and relational proximity endogenous, being correlated both with the unobserved quality of the entrepreneurs and the dependent variable under analysis. In absence of a valid instrument for relational proximity, as explained in the relative chapter, we cannot claim causality for the correlation at stake. Furthermore, as the work is based on the Italian case-study characterized by a low-degree of development of its equity market with respect to other European countries, the general validity of the findings should be verified through a cross-country analysis in future research. In Chapter 2, given that the technical challenges of building an invertible model with heterogeneous workers, commuting costs and multiple cities have not been resolved yet, the adopted model treats workers as homogeneous. In doing that, it controls for the effect that the remote-work induced average reduction in commuting costs have on average location preferences, rather than for the specific location choices of remote workers. However, it accounts for cities' heterogenous occupational composition from which it derives the reduction in commuting-times resulting from city-specific potential remote-work adoption. Chapter 3 exploits the so called 'indirect' methodology, which proceeds from a theoretical assumption to quantify misallocation. This approach has the advantage of requiring less data, enabling the analysis of misallocation across multiple countries at different territorial levels. However, the drawback of this approach is the inability to identify causation in the analysis of specific sources of the observed allocative inefficiencies, to which future research could be dedicated.

While previous limitations represent important challenges to be addressed in future research, other avenues for that can be identified by extending the contexts of the thesis. As for Chapter 1, an interesting extension could apply to the result about the strong and significant role that the relational proximity existing between risk-capital investors and target firms has on the investment link. Exploiting the idea of proxying relational proximity with (the inverse of) professional and ownership network distances, future research could explore the characteristics of such networks, and their effect on the success of economic interactions. As far as the Chapter 2 is concerned, the analysis on the effects of remote-work adoption on location choices could be extended by looking at within-city location choices of both firms and workers. Furthermore, if new data or methodologies allowing to account for workers heterogeneity while controlling for the inherent double causality problem were to become available, sorting mechanisms could be included, as these could strengthen the agglomeration-reinforcing effect that the Chapter has found. Last but not least, the novel evidence provided in Chapter 3 on the relevance of regional misallocation for aggregate productivity could open a new stream of research on the topic. The analysis conducted so far could be expanded through the inclusion of the service sector, which previous studies have found to account for higher degrees of misallocation than the manufacturing ones. Furthermore future research should consider the implementation of direct methodologies that allow to measure the impact of specific sources of allocative inefficiencies. All of these, and possibly other ones, are lines of research to which the results obtained in the thesis could confidently be of inspiration on a field of research still relevant and open in many respects.

References

Aghion, P., Fally, T., Scarpetta, S. (2006). Credit Constraints as a Barrier to the Entry and Post-Entry Growth of Firms. *Economic Policy*, 22(52): 731+733-779.

Agrawal, A., Kapur, D., McHale, J., (2008). How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2): 258–269.

Albouy, D., Behrens, K., Robert-Nicoud, F., Seegert, N. (2018). The Optimal Distribution of Population across Cities. *Journal of Urban Economics*, 110: 102–13.

Baldwin, R., Forslid, R., Martin, P., Ottaviano G.I.P., and Robert-Nicoud, F., (2003). *Economic Geography and Public Policy.* Princeton University Press

Bathelt, H., Gluckler, J. (2003). Toward a relational economic geography. *Journal of Economic Geography*, 3(2): 117–144.

Behrens, K., Duranton, G., Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. Journal of Political Economy, 122(3), 507–553.

Behrens, K., and Robert-Nicoud, F., (2011). Tempora Mutantur: In Search of a New Testament for NEG. *Journal of Economic Geography*, 11(2): 215–30.

Behrens, K., Robert-Nicoud, F. (2015a). Agglomeration Theory with Heterogeneous Agents. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), *Handbook of Regional and Urban Economics*, 5: 171-245.

Behrens, K., Robert-Nicoud, F. (2015b). Are cities too small? Equilibrium and optimal urban systems with heterogeneous land. *Working Paper.*

Behrens, K., Mion, G., Murata, Y., Suedekum, J., (2017). Spatial frictions. *Journal* *of Urban Economics*, 97: 40–70.

Behrens, K., Murata, Y., (2021). On Quantitative Spatial Economic models. *Journal of Urban Economics*, 123, 103348.

Berchicci, L., Block, J. H., Sandner, P. G., (2011). The Influence of Geographical Proximity and Industry Similarity in a Business Angel's Investment Choice. *Ssrn.*

Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1): 61–74.

Boschma, R., Marrocu, E., Paci, R. (2016). Symmetric and asymmetric effects of proximities: The case of M&A deals in Italy. *Journal of Economic Geography*, 16(2): 505–535.

Büchel, K., Ehrlich, M. V., Puga, D., Viladecans-Marsal, E. (2019). Calling from the outside: The role of networks in residential mobility. *Journal of Urban Economics*, 1909.

Caggese, A. (2019). Financing constraints, radical versus incremental innovation, and aggregate productivity. *American Economic Journal: Macroeconomics*, 11(2), 275–309.

Caragliu, A., Nijkamp, P. (2016). Space and knowledge spillovers in European regions: The impact of different forms of proximity on spatial knowledge diffusion. *Journal of Economic Geography*, 16(3): 749–774.

Charlot, S., Gaigné, C., Robert-Nicoud, F., Thisse, J. F. (2006). Agglomeration and welfare: The core-periphery model in the light of Bentham, Kaldor, and Rawls. *Journal of Public Economics*, 90(1–2): 325–347.

Combes, P. P., Duranton, G., Overman, H. G. (2005). Agglomeration and the adjustment of the spatial economy. Production.

Combes, P.P., Gobillon, L., (2015). The Empirics of Agglomeration Economies. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), *Handbook of Regional and Urban Economics*, 5, 247–348

Cumming, D., Dai, N. (2010). Local bias in venture capital investments. *Journal of Empirical Finance*, 17(3): 362–380.

Dauth, W., Findeisen, S., Suedekum, J., Moretti, E. (2019). Matching in Cities. *IZA Discussion Paper Series*, 12278: 1–68.

Desmet, K., Rossi-Hansberg, E., (2013). Urban Accounting and Welfare. *American Economic Review*, 103 (6): 2296-2327.

Desmet, K., Rossi-Hansberg, E. (2014a). Analyzing Urban Systems: Have Megacities Become Too Large. The World Bank, *Policy Research Working Paper Series*, 6872: 1-27.

Desmet K., Rossi-Hansberg E. (2014b). Spatial development. *American Economic Review*, 104(4): 1211–43

Doran, J., Jordan, D., O'Leary, E. (2012). The effects of the frequency of spatially proximate and distant interaction on innovation by Irish SMEs. *Entrepreneurship and Regional Development*, 24(7–8), 705–727.

Duranton, G., (1999). Distance , land , and proximity: economic analysis and the evolution of cities. *Environment and Planning A*, 31: 2169–2188.

Duranton, G., Puga, D., (2004). Micro-foundations of urban agglomeration economies. In: Henderson J. V. and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, 1(4): 2063-2117, Elsevier.

Eeckhout, J., Pinheiro, R., Schmidheiny, K., (2014). Spatial Sorting, *Journal of Political Economy*, 122, 554–620.

Fajgelbaum, P. D., Gaubert, C., (2020). Optimal Spatial Policies, Geography, and Sorting. *The Quarterly Journal of Economics*, 135 (2): 959–1036.

Fajgelbaum, P. D., Morales, E., Suarez Serrato, J. C., Zidar, O. (2019). State Taxes and Spatial Misallocation. *Review of Economic Studies*, 86(1): 333–376.

Fontagné, L., Santoni, G. (2019). Agglomeration economies and firm-level labour misallocation. *Journal of Economic Geography*, 19(1): 251–272.

Fujita, M., Mori, T., (1996). The role of ports in the making of major cities: self-agglomeration and hub- effect. *Journal of Development Economics*, 49: 93–120.

Fujita, M., Thisse, J. F. (2002). Economics of Agglomeration. Cities, industrial location, and regional growth. Cambridge University Press, New York.

Fujita, M., Thisse, J.F., (2013). Industrial Agglomeration under Monopolistic Competition. In Fujita, M. (Ed.), *Economics of Agglomeration. Cities, Industrial Location, and Globalization*, 285–345. Cambridge University Press. Second Edition.

Gaspar, J. M. (2018). A prospective review on New Economic Geography. *Annals of Regional Science*, 61(2): 237–272.

Glaeser, E. L., Gottlieb, J. D., (2008). The Economics of Place-Making Policies. *Brookings Papers on Economic Activity*, 2008(1): 155–253.

Gossling, T., Knoben, J., (2011). Proximity and duration in temporary organisations. *International Journal of Human Rights and Constitutional Studies*, 2(3): 207-220.

Graham, D. J., Gibbons, S., Martin, R. (2010). The spatial decay of agglomeration economies: estimates for use in transport appraisal. LSE, Department for Transport.

Haltiwanger, J., Jarmin, R. S., Kulick, R. B., Miranda, J. (2016). High Growth

Young Firms: Contribution to Job, Output and Productivity Growth. In Haltiwanger, J., Hurst, E., Miranda, J., Schoar, A., (Eds.), *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, University of Chicago Press.

Ham, J.C., Swenson, C., İmrohoroğlu, A., Song, H., (2011). Government programs can improve local labour markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Community. *Journal of Public Economics*, 95(7–8): 779-797.

Helsley, R. W., Strange, W.C. (2002). Innovation and input sharing. *Journal of Urban Economics*, 51 (1): 25-45.

Helpman, E., (1998). The size of regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), *Topics in Public Economics. Theoretical and Empirical Analysis*, Cambridge University Press, 33-54.

Hermann, J., Avdeitchikova, S., Hjertstrom, A. (2016). The influence of functional and relational proximity on business angel investments. *International Journal of Entrepreneurship and Small Business*, 29(3): 468-490.

Hidalgo, C.A., Hausmann, R., (2009). The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*, 106 (26): 10570–75.

Hsieh, C. T., Moretti, E., (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2): 1–39.

Hsieh, C. T., and Klenow, P. J., (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, CXXIV(4): 1403–1447

Kline, P., Moretti, E. (2014). People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs. *Annual Review of Economics*, 6(1): 629–662.

Krugman, P., (1980). Scale Economies, Product differentiation, an the pattern of Trade. *American Economic Review*, 70(5): 950-959.

Krugman, P., (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3): 483-499.

Krugman, P., (1993). First nature, second nature, and metropolitan location. *Journal of Regional Science*, 33: 129–144

Lee, N., Brown, R. (2017). Innovation, SMEs and the liability of distance: The demand and supply of bank funding in UK peripheral regions. *Journal of Economic Geography*, 17(1): 233–260.

Lee, N., Luca, D. (2019). The big-city bias in access to finance: Evidence from firm perceptions in almost 100 countries. Journal of *Economic Geography*, 19(1): 199–224.

Lenzu, S., Manaresi, F. (2019). Sources and Implications of Resource Misallocation: New Evidence from Firm-Level Marginal Products and User Costs. *SSRN Electronic Journal*.

Lucas, R.E., (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1): 3–42.

Mariotti, S., Mutinelli, M., Nicolini, M., Piscitello, L., (2015). Productivity Spillovers from Foreign Multinational Enterprises to Domestic Manufacturing Firms: To What Extent Does Spatial Proximity Matter? *Regional Studies*, 49(10): 1639–1653.

Marrocu, E., Paci, R., Usai, S. (2014). The Complementary Effects of Proximity Dimensions on Knowledge Spillovers. Spatial Economic Analysis, 9(1), 9–30.

Marshall, A., (1890). Principles of Economics : An Introductory Volume. Macmillan.

Marshall, A., (1919). Industry and Trade. Macmillan

Martin, R., Simmie, J., (2008). The theoretical bases of urban competitiveness : does proximity matter ? *Revue d'économie régionale et urbaine*, Armand Colin, 0(3): 333-351.

McFadden, D.L., (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In: *Frontiers in Econometrics*, Wiley, New York. Academic Press, 105-142.

Melitz, M.J., (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71: 1695–725.

Melitz, M. J., Redding, S. J. (2014). Heterogeneous Firms and Trade. In Gopinath. G., Helpman, E., Rogoff, K.. (Eds.), *Handbook of International Economics*, 4:1–54.

Meuleman, M., Jääskeläinen, M., Maula, M. V. J., Wright, M. (2017). Venturing into the unknown with strangers: Substitutes of relational embeddedness in cross-border partner selection in venture capital syndicates. Journal of Business Venturing, 32(2), 131–144.

Misch, F., Saborowski, C. (2020). The Drivers and Consequences of Resource Misallocation: Exploiting Variation across Mexican Industries and States. *Economia*, 2:61–96.

Neumark, D., Simpson, H., (2015). Place-Based Policies. In Duranton, G., Henderson J.V, Strange W.C., (Eds), *Handbook of Regional and Urban Economics*, 5: 1197–1287, Elsevier.

Ottaviano, G.I.P., Thisse, J. F., (2004). Agglomeration and economic geography. In Henderson J. V. and Thisse J.F. (Eds.), *Handbook of Urban and Regional Economics* 4: 2563–2608. Elsevier B.V.

Ottaviano, G.I P. (2011). "New" new economic geography: Firm heterogeneity and agglomeration economies. *Journal of Economic Geography*, 11(2): 231–240.

Puga, D. (2010). The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science*, 20: 203-2019.

Ramirez-Pasillas, M. (2010). International trade fairs as amplifiers of permanent and temporary proximities in clusters. *Entrepreneurship and Regional Development*, 22(2): 155-187.

Redding, S. J., and Rossi-Hansberg, E., (2017). Quantitative Spatial Economics. *Annual Review of Economics*, 9(1): 21–58.

Restuccia, D., Rogerson, R. (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives*, 31(3): 151-74.

Rodríguez-Pose, Andrés (2017). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, 11(1): 189-209.1378.

Romer, P.M., (1994). The Origins of Endogenous Growth. *Journal of Economic Perspectives,* 8 (1): 3–22.

Romer, P.M. (1990). Endogenous technological change. *Journal of Political Economy* 98: S71–S102.

Singh, J. (2005). Collabourative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, 51(5): 756-770.

Sorenson, O., Stuart, T., (2001). Syndication Networks and the Spatial Distribution of Venture Capital Investments. *American Journal of Sociology*, 106(6): 1546–1588.

Thisse, J. (2009), How Transport Costs Shape the Spatial Pattern of Economic Activity, OECD/ITF Joint Transport Research Centre Discussion Papers, No.

2009/13, OECD Publishing, Paris.

Tykvová, T., Schertler, A., (2011). Venture capital and internationalization. *International Business Review*, 1(6): 423-439

Tykvová, T., Schertler, A., (2014). Does syndication with local venture capitalists moderate the effects of geographical and institutional distance? *Journal of International Management*, 20(4): 406–420.

Torre, A., Rallet, A., (2005). Proximity and localization. *Regional Studies*, 39(1): 47-60.

Torre, A. (2008). On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42(6): 869–889.

Venables, A.J., (2011). Productivity in Cities: Self-Selection and Sorting. *Journal of Economic Geography*, 11(2): 241–51.

Yang, S., Fu, Y., Lu, Y. (2017). The Spatial Misallocation of Capital and labour in China: A Welfare Analysis Based on Urban Accounting. In: *American Real Estate and Urban Economics Association (Ed.), Agglomeration Economies and Development*, 1–35.

Zhao, T., Jones-Evans, D. (2017). SMEs, banks and the spatial differentiation of access to finance. *Journal of Economic Geography*, 17(4): 791–824.

CHAPTER 1

Tangible and intangible proximities in the access to Venture Capital: evidence from Italian innovative start-ups

1.1 Introduction

Financial capital markets are notably characterised by an uneven geography. Peripheral cities and non-central areas systematically reveal smaller flows of both debt and equity capital than their core counterparts (Martin *et al.*, 2002), and this makes the former places relatively more affected by financial constraints to firms' innovation than the latter (Lee and Luca, 2019; Lee and Brown, 2017; Donati and Sarno, 2015; Lee and Drever, 2014; Cumming and Johan, 2007). The evidence of these regional funding gaps (Martin *et al.*, 2005) has been explained with the joint occurrence of a "local bias" phenomenon, amounting to the tendency of capital investors to invest the majority of their portfolio where they choose to locate (Cumming and Dai, 2010; Lee and Luca, 2019), and the clustering trend of risk-capital investors in major financial cities (Florida and Mellander, 2016; Mason, 2007). In turn, among the causes of this bias, the spatial proximity between investor and investee has received the greatest attention. Given its role in mitigating the information asymmetries (Petersen and Rajan, 2002) and the transaction costs entailed by their financial relationship (Van Osnabrugge, 2000; Zook, 2002), spatial proximity has been established as determinant not only for risk-capital actors and deals, but also for bank investment and lending relationships [11]

However, Fritsch and Schilder (2006) suggested spatial proximity to be less important for VC than for other smart capital suppliers such as banks or Business Angels.

The primacy of spatial proximity in accounting for the access of firms to local sources of financial capital has indeed recently shown some interesting specifications, if not even contradictions. In some dedicated business surveys, for example, VC investors started showing a certain indifference to the investee location and an increasing engagement in deals that are apparently not local (Carlson and Chakrabarti, 2007, Martin *et al.*, 2005). Aligned with this is the evidence emerging from some financial markets, like the VC one, in which investors appear to overcome the boundaries of their location for targeting geographically distant firms, providing they can rely on networking to draw info about them (Sorenson and Stuart, 2001). Furthermore, if spatial proximity was to be a major discriminant for investments to occur, firms seeking equity would locate closer to risk-capital owners and regional equity gaps would not arise. The alleged absence of location-mirroring in VC-ISC financial relationships is thus a phenomenon that requires closer scrutiny. Indeed, it has important implications for the spanning of financial opportunities across places, and for firms' access to finance in areas that miss investors concentration. In looking

11 While Petersen and Rajan (2002) found evidence of a decrease in the relevance of spatial proximity for small businesses lending in the 1973-1993 period, the subsequent literature on the geography of bank investments and lending has found organizational distance - *i.e.* firms or local banks distance to the banks headquarters - to be a strong determinant of the probability to receive loans (Alessandrini *et al.*, 2009, Bragoli *et al.*, 2022) and of the innovation and risk propensity of the investee (Alessandrini *et al.*, 2010). This literature has also investigated the role of other forms of proximity in bank-investments, such as social (Wang *et al.*, 2021) and technological (Antonietti *et al.*, 2015) one, finding their effect to become significant when spatial or organizational proximity increase.

for its determinants, it is important to understand whether, and to which extent, other forms of proximity than the geographical one can account for the matching between investors and investees. Drawing on the geography literature about proximity (Balland *et al.*, 2013), we try to fill this gap and claim that the presence of regional equity biases and the absence of location-mirroring behaviors can be reconciled by looking at a manifold proximity between investor and investee and at the relationship among its different variants. Firstly, we argue that the role of spatial proximity in affecting the firms' access to finance is twofold but indirect, and mainly related to the its ability to facilitate and predict other forms of intangible proximities. As such, we show that it should be evaluated mainly through a functional (Brown and Horton, 1970) rather than a geographical kind of tangible distance between actors, the former measure being more adequate to representing the effort that agents must put in place in order to interact. Secondly, we maintain that the relationship between investor and investee is also affected by an (at least) twofold intangible proximity, accounted by: their sharing experience and familiarity with the industries in which they operate – industrial proximity – and their having investment and professional relationships, which create information and trust effects among them – relational proximity. Thirdly, we posit that these intangible proximities could work in alleviating the binding role of tangible proximities for the financial relationship to take place.

We develop and test these arguments by referring to a specific financial relationship, with respect to which they appear more salient: the relationship between innovative start-up companies (ISCs) and Venture Capital (VC). On the one hand, small, new (and young), innovative firms often lack adequate levels of tangible assets for collaterals, making debt finance more onerous and difficult to access (Pollard, 2003). Furthermore, ISCs tend to exhibit with respect to more mature and less innovative firms, greater informational asymmetry and higher contract incompleteness, due to the deficiency of track records (Zook, 2002). On the other hand, the ISCs access to finance appears possibly more intriguing with respect to VC, representing a specific category of equity investors for young firms, which has been documented to be a key driver of their innovation capacity (Kortum and Lerner, 2000; Peneder, 2010), their growth (Davila *et al.*, 2003; Grilli and Murtinu, 2014; Haltiwanger *et al.*, 2016; Bertoni *et al.*, 2007; Carpenter and Petersen, 2002), and their survival over time (Bonini *et al.*, 2019; Beck and Demirguc-Kunt, 2006). The literature about the role of VC in financing newly established high-tech firms has been growing rapidly in the last years (Colombo *et al.*, 2010; Giraudo *et al.*, 2019; Caviggioli *et al.*, 2020; Alperovych *et al.*, 2020; Colombelli *et al.*, 2020), confirming that VC are possibly more enabling than other financing modes (e.g., crowdfunding and business angels) in their promotion.

In positioning in this literature, we bring three main contributions to it. First of all, as anticipated, we place an original focus on the manifold proximity between VC and ISC, and look at how the tangible and intangible distances between them affect their matching. Given the relevance that different kinds of proximities have been shown to have for the innovation process, and their typical non-linearities and complementarity effects (Boschma, 2005; Davids and Frenken, 2018), this is an unfortunate gap that we aim at filling.

Second, we extend previous research about the relational proximity between VCs and VCs (Sorenson and Stuart, 2001) and rather consider that between VC investors and target companies. In particular, we argue that the network that the two types of players come to determine through their ownership relationship and related professional appointments (on boards and on other settings) can be a leverage for information exchange and trust building, which VC can use to reach more distant targets.

Our third contribution rests in the domain of our empirical application. Unlike the majority of existing studies on the role of geography (as of other proximities) in VC investments, mainly focused on the U.S. (Florida and Mellander, 2016; Carlson and Chakrabarti, 2007; Cumming and Dai, 2010; Sorenson and Stuart, 2001; Zook, 2002), the UK (Lee and Drever, 2014; Mason and Pierrakis, 2013; Martin *et al.*, 2005; Mason and Harrison, 1992; Harrison and Mason, 2002) and the German (Lutz *et al.*, 2013; Bender, 2010; Fritsch and Schilder, 2006; Martin *et al.*, 2005) markets, with few world (Tykvová and Schertler, 2014) and European-wide studies, (Martin *et al.*, 2002), the present one concentrates on the Italian ISCs and their VC investors.

The Italian equity and Venture Capital market has been previously analysed by a number of studies (Grilli, 2019; Vacca, 2013; Bertoni *et al.*, 2011; Bertoni *et al.*, 2007). Nonetheless, to the best of our knowledge, this is the first study to focus on the role of proximit*ies* in the Italian context, which we believe to deserve a specific focus due to a number of motives. First, the Italian equity market is generally considered immature with respect to countries with different models of capitalism (Della Sala, 2004; Vacca, 2013; OECD, 2017; De Socio, 2010; Bertoni *et al.*, 2007), exerting below-average degrees of attractiveness for risk-capital investments with respect to other EU countries (Groh *et al.*, 2010)[12]. This fact makes of our focal relationship a

12 The Italian financial system is actually considered a bank-based one. Nonetheless, access to bank loans by new technology-based firms in Italy was found to be still sparse and to be quite insensitive to demand-side factors (Colombo and Grilli, 2007), reinstating equity and personal funds as the main tools against credit constraints for young and innovative firms in the country.

relatively rare event to observe, and makes us also expect that the market imperfections that render the proximity between investor and investee salient, could be more severe than elsewhere (Whited and Zhao, 2021; Midrigan and Xu, 2014). Second, the country displays below EU-average institutional quality (Adam, 2008), increasing the importance of monitoring and referrals for risk-capital deals[13] imposing in turn an higher weight on transactions costs when evaluating potential risk capital investments. As a result, we expect non-geographical proximities, and especially relational one (Johnson *et al.*, 2002), to express an exceptional role.

Third, the reference to Italy allows us to exploit in the analysis an unambiguous, legal definition of innovative startup companies, which the Italian government has issued in the aftermath of its comprehensive strategy to promote private equity and innovative business creation. Given that investing in such a legally defined kind of ISCs provides VC investors with fiscal exemptions, this automatically restricts our focus to the analysis of subsidized dyadic relationships.[14] However, by retaining all of these start-ups and only them, we do not run the risk to have confounding effects of our focal arguments.[15] It should be noted that, as Finaldi Russo *et al.*, 2016 point out, Italian innovative startup companies differ substantially from other

---

13 In particular the low degree of the judicial system's efficiency and of contract-enforcement with it are deemed to increase moral hazard and strategic behaviors of creditor firms (Schiantarelli *et al.*, 2020; Fabbri, 2010), hindering gross investments (Dejuan and Mora-Sanguinetti, 2019) and credit access both at aggregate (Moro *et al.*, 2018) and local (Giacomelli and Menon, 2016) level.

14 Together with the provision of dedicated tax and financial incentives, the Italian Law has actually instituted and closely monitored the ISC as a novel juridical form, subject to specific innovativeness requirements and to the obligation of producing publicly available data yet unexplored.

15 Available measures of innovativeness tend to exploit data that are generally unavailable for small firms (Iorio and D'Amore, 2017; Battisti and Stoneman, 2019), or that tend to represent sectors unequally (measures based on patents tend to favour the recognition of manufacturing firms, while measures based on intangibles are more likely to identify firms in the service sector (Taques et al, JIK, 2021). Measures based on RD expenditures would rely on data that are not available for Italy, where balance sheets report RD and marketing expenditures jointly.

Italian startups, even from those in high-tech sectors. Italian ISCs tend to focus on early development of breakthrough innovations, and as such to locate in places that exhibit high levels of unrelated variety (Antonietti and Gambarotto, 2020) and of knowledge-spillovers (Ghio *et al.*, 2016; Colombelli, 2016). These stronger localisation constraints, combined with Venture Capital clustering, may render the role of non-spatial dimensions of proximity with VC funds more evident.

Moreover, our work differs from extant ones in terms of its more recent time-span (2014-2019), for its focus on the determinants of investments in innovative startups'.

The rest of the work is structure as follows. Section 1.2 positions the paper in the extant literature and develops our arguments about the relevance of different forms of proximity for the occurrence of VC investments in innovative start-ups. Section 1.3 presents our dataset and the evidence about regional gaps in Italian VC investments, for then illustrating model, methodology and estimation issues of the empirical analysis. Results are discussed in Section 1.4 and Section 1.5 concludes with a discussion of their implications.

1.2 Proximities in the relationship between Venture Capitalists and innovative start-ups

Similarly to other forms of financing, the VC market shows a quite uneven geography, marked by their systematic gaps in outlying regions (Martin *et al.*, 2005; Mason, 2007; Martin *et al.*, 2002). From a theoretical point of view, these regional gaps can be explained by coupling the typical clustering trend of equity and lending actors in major financial cities (Lee and Luca, 2019; Florida and Mellander, 2016;

Mason, 2007) with a local bias in the geographical distribution of their portfolio (Cumming and Dai, 2010; Zook, 2002; Lutz *et al.*, 2013). On the one hand, the monitoring and tutoring activities implemented by VC to address agency and information issues (Van Osnabrugge, 2000) require face-to-face interactions with the investee. These interactions are the more costly and the less effective, the larger their geographical distance, even in the presence of digital forms of communication (Fritsch and Schilder, 2006; Zhao and Jones-Evans, 2017). On the other hand, the average transaction costs that VC encounter in selecting the relevant deals decrease with deal-size, and firms located in smaller peripheral cities tend to be smaller and exhibit lower financial returns than those in large areas (Martin *et al.*, 2005).

The presence of regional equity gaps naturally brings to the front the importance of a tangible kind of proximity between VC and ISC. However, in spite of its alleged importance, recent studies have shown that, when directly interviewed about the relevance of the investee location, VC managers surprisingly appear indifferent to it. Indeed, a perceived spatial non-sensitivity in deals selection has emerged in different VC markets, like those of Germany (Fritsch and Schilder, 2006) and the US (Carlson and Chakrabarti, 2007). This clash between theory and evidence is puzzling and suggests that the pure agency and transaction conceptual framework that has been mainly used so far to interpret the phenomenon, might be in need of further integration with regional and geographical studies (Sorenson and Stuart, 2001). In particular, we maintain that a more geographically sophisticated analysis of the relationship between investor and investee is required, which leads us to recognise the relevance of a manifold notion of proximity between them.

1.2.1 Tangible proximities in VC investments

A first aspect that needs closer scrutiny concerns the tangible kind distance that separates the VC fund from the investee, or its mirror equivalent in terms of *tangible proximity* between them. The extant literature usually refers to this proximity in generic terms, as a factor that facilitates both the pre-investment activities of VC firms – consisting of the identification and the appraisal of investment opportunities – and their post-investment role – amounting to the monitoring of the identified ventures and to the supply of value-added services to them (Sorenson and Stuart, 2001). Indeed, all of these are experiential tasks that involve the acquisition and elaboration of procedural and tacit knowledge, if not even social interactions, which become difficult to implement at a distance.

While the role of tangible proximity in spurring local investing by VC appears quite intuitive and has been ascertained since long (Gupta and Sapienza, 1992), it is rarely considered that its nature is twofold and encompasses two forms of proximity, whose effect is not necessarily equivalent. On the one hand, we have the tangible, or spatial proximity to which evolutionary economic geography and regional studies usually refer with the inverse of the *geographical distance* that separates economic actors as the length of space between them (Boschma, 2005). This is pivotal in this stream of literature, as such a distance is retained to condition the spatial concentration and agglomeration of agents, which permit the knowledge spillovers that conduce innovation, on which it focuses.

On the other hand, when we look at financial relationships like VC investments, the production of innovative knowledge is not the focal outcome of the interaction, which is rather intended to contrast information asymmetries and to facilitate the selection, evaluation and commercial exploitation of the innovative deals. With respect to this kind of interaction, “the effort that it takes to interact”, at the basis of what (Moodysson and Jonsson, 2007, p. 118) have defined *functional distance*, is a different and arguably more relevant form of (inverse) tangible proximity in accounting for VC investments. As for its difference with respect to geographical distance, it is evident that functional proximity additionally accounts for the existence of infrastructures and travelling times, which are arguably pivotal in the interaction required by a VC investment, as they increase the opportunity costs of getting info and monitoring investments. In the light of that, two equally distant places, could be heterogeneously hard to be reached. As for the greater importance of functional proximity or accessibility with respect to the geographical one, this has actually been already documented by previous studies about other forms of equity investments, like business angel investments (Hermann *et al.*, 2016), and its extension to the analysis of VC investments can help us explaining the apparently contradictory evidence on the spatial sensitivity in deals selection.[16]

1.2.2 Intangible proximities in VC investments

While VC and target firms relate between them in the geographical space, their tangible distance is not the only dimension along which they can be retained proximate. This is the main argument that emerges from a quite thick stream of literature on the notion of proximity itself, which recognises its manifold nature with different proximity variants, depending on the specific approach to it (Balland *et al.*, 2013).

16 In business surveys, questions about functional distance, usually posed to managers by using relevant thresholds (e.g. within-two-hours travel distance) are possibly easier to be evaluated than more general questions about geographical distance (e.g. importance of location). This framework effect could also concur to explain the perceived irrelevance of the latter detected by Carlson and Chakrabarti, 2007 and Fritsch and Schilder, 2006.

In general, the main point is that, in spite of the 'over-territorialized' analysis of their relationships over the last two decades (Hess, 2004), economic actors are embedded in different a-spatial contexts, which creates different forms of immaterial proximities between them. Following the seminal works by Polanyi (1944) and Granovetter (1985), the most evident kind of embeddedness is in the social networks that agents come to create by interacting, knowing and trusting each other, and within which they can be more or less socially close, irrespectively from their tangible proximity. Other forms of intangible proximity (like organisational and institutional ones) have been identified in more specific kinds of relationships, like those occurring among agents involved in innovation activities as such, on which evolutionary economic geography has come to focus (Boschma, 2005).

In the kind of financial relationship between VC and ISC that we are investigating, two intangible proximities appear more salient and requires more attention than the one they found so far. Following an 'interactionist' approach to proximity (for which see Balland *et al.*, 2013), these can be considered: a 'similarity' proximity, represented by the industrial closeness between VC and target firm, and a "belonging" proximity, emerging along the business relationships they entertain.

Industrial proximity

An important form of proximity between VC and target firms is determined by the extent to which the former has already invested in the industry of the latter. Through its prior investments in the industry of the target company, the VC fund can in fact get more knowledge, if not even experience, of that industry and increase the chance of success of the prospected deal. The main channels through which an increase of this industrial proximity can make the VC-ISC match more probable are once more connected to both the pre-investment and post-investment phase of the deal (Sorenson and Stuart, 2001). On the one hand, the prior experience a VC has acquired by investing in a certain industry naturally extends the number of contacts with entrepreneurs and third investors of that industry, and this in turn arguably improves the exploration of new investment opportunities. On the other hand, this industry-specific experience can make the VC more confident in its capacity to detect and interpret signs of early-stage problems and to monitor the evolution of the prospected deal in the same industry.

An additional channel through which the VC-ISC industrial proximity can facilitate the matching of a new deal is represented by the synergies it creates between the prospected and the existing backed companies in the VC portfolio. The presence of industry-specific knowledge in fact spurs VC firms to specialise in the industry at stake, and this provides them with coordination economies in the management of their portfolio, which could benefit the new deal too (Norton and Tenenbaum, 1993). However, as successive literature has shown (Buchner *et al.*, 2017; Patzelt *et al.*, 2009), a VC specialisation strategy isn't necessarily superior with respect to a diversification one, which could instead offer knowledge-sharing across funds and higher chances of risk-reduction. In the light of this last consideration, the effectiveness of this channel of industrial proximity is conditional on the actual diversification strategy of the VC. Still, we expect industrial proximity to be positively associated with the probability of observing a VC-ISC match, even if this relationship could be non-monotone given the trade-off among the benefits of specialization and diversification strategies. The consideration of the advantages that industrial proximity offers to the match at stake, leads us to expect that the same proximity could inter-operate with the enabling role that we have recognised to tangible proximities in the previous section. As we said, being close to the target company can facilitate the VC firm in getting and exploiting knowledge that would be difficult to access at a distance. However, the experience that VC acquire through their industrial proximity to the ISC can somehow substitute, though imperfectly, this local knowledge need and possibly enable them to extend the spatial reach of their investments. For example, through the experience entailed by industrial proximity, the VC could increase the number of knowledge sources to be used for a deal, as well as extend and consolidate the synergies of their backed-firms portfolio, and this could compensate the loss of a few or a unique local knowledge source. By developing this argument, we do also expect that the industrial proximity we are referring negatively moderates the positive effect that tangible proximities arguably have on the match between VC and ISC.

Relational proximity

Possibly more important than the "similarity" proximity entailed by VC-investment partners sharing the same industry, is the "belonging" proximity that descends from their being part of common interpersonal networks, or in brief, their relational proximity.

As we have repeatedly noticed, the VC-ISC match we are investigating mainly depends on information transmission and knowledge exchanges among the focal actors and, as diverse streams of sociological literature have widely shown since long (Coleman, 1994; Friedkin, 1998), interpersonal relations are the main driver and structuring factor of information/knowledge circuits. This is particularly so in the VC market, in which public information about investment opportunities and early stage companies is basically missing and in which operators often miss sufficiently large histories of performance on which to base their evaluations. In such a context, trusted information coming from networked actors and verified through multiple networked parties becomes a crucial element, and the same holds true for the personal, investment and professional relationships through which networks come to exist. A social tie (either direct or mediated by a common link with another firm or individual) in fact involves expectations of social obligations (Uzzi, 1996), and is thus considered a trust-based privileged information channel[17]. Access to finance, either as bank loans (Uzzi, 1999) or venture capital, has in fact been found to be positively associated with both direct (Shane and Cable, 2002) and indirect referrals through close contacts (Fried and Hisrich, 1994; Hain *et al.*, 2016). Moreover, the target firm's social capital, and its position and in the VC network can be considered as a signal of experience and reputation (Bollazzi *et al.*, 2019; Hsu, 2007).

Within the networks at stake, the relational proximity between partners indicates the existence and the intensity of the relative ties, which arguably affect the VC-ISC match. Indeed, in the extant literature, this has been postulated and empirically ascertained by mainly looking at the inter-firm relationships through which VC funds come to constitute their community. More precisely, since the seminal work by Sorenson and Stuart, 2001, this has been investigated by looking at the networks that VC firms form through the use of syndicated investing-facilitates[18]: not only do they enable the financial relationship at stake, but they also decrease the space-based constraints posed by tangible proximities.

17 The idea that social networks may facilitate economic transactions by overcoming informational barriers is not new, and have also been applied in the trade literature to show that they can reduce home bias in intra-(Garmendia *et al.*, 2012) and inter-national trade decisions (Combes *et al.*, 2005).

18 As is well-known, this is the case of new ventures that obtain funding from syndicates of investors, that is, from more than one VC firm.

Drawing on this contribution, subsequent studies have found that the social embeddedness of Venture Capitalists, measured through heterogeneous social and organizational ties, crucially affect the unfolding and the performance of the investment (Meuleman *et al.*, 2017; Teten and Farmer, 2010; Milosevic, 2018) as well as the probability that VC invest in spatially (Tykvová and Schertler, 2014; Sorenson and Stuart, 2001), institutionally (Tykvová and Schertler, 2011) or technologically distant firms (Meuleman *et al.*, 2017; Tykvová and Schertler, 2014; Cumming and Dai, 2010). Quite surprisingly, only few studies instead have recently addressed the social links that could exist between VC and target companies (Nigam *et al.*, 2020; Hermann *et al.*, 2016; Fuchs *et al.*, 2021), generally finding that they have a positive effect on the access to financing. Given the role that the interpersonal relations between the two parties of the match could have in facilitating the exchange of information about the deal, and in building up trust relationship that could increase the chance of its success, this is an unfortunate gap that needs to be filled and on which we focus in our empirical application.

As we will see, we put forward an original methodology to proxy the relational proximity between VC and ISC and empirically test if, as we do expect, this proximity facilitates their matching. Furthermore, we will also investigate if, by mimicking what has already been found with respect to the social relations between VCs, the relational proximity between VC and target companies is also capable to extend the geographical coverage of their relationship. In order to do that, we will see whether relational proximity negatively moderates the impact that tangible proximities should have on the VC-ISC match.

1.3 Empirical analysis

1.3.1 Data and descriptive evidence

Our empirical analysis refers to the population of Italian Innovative Startup Companies (ISCs) and to the investments that Venture Capitalists (VC) have made in them over the period 2012-2019. In order to identify a firm as ISC, we refer to a specific form of business that, with the policy support introduced by the Italian Law 221/2012, has been recognised with the following criteria: i) an age of less than five years; ii) at least one of the following requisites: 1) employing at least one-third of workers with a doctoral diploma, or two-thirds with a master diploma, 2) being licensee or depositor of at least one patent or other industrial property rights; 3) investing at least 15% of the value or cost of production in R&D activities. Following this definition, we have collected a panel dataset, observed over the period 2012-2019, constituted by all the 10,213 Italian ISCs that registered as such before June, 6th, 2019. With respect to this firm population, we have merged data contained in the Italian business registry (*Registro Italiano delle Imprese*) with Bureau Van Dijk data and obtained detail information about their localization, ownership structure, investments and other balance sheet data. Among these data, information about all the investors of the identified Italian ISCs have been retrieved at the same date. Quite interestingly, out of the 38,425 detected investors, only 37 are Venture Capital funds: an information that we will carefully retain in the following analysis.

To identify Venture Capital funds we consider all the independent companies[19]

19 We focus on independent VC companies excluding agencies or banks to focus specifically on these capital investors, for two reasons. The first is that Venture Capital represent a specific form of financial investors that developed specifically to invest in high-risk high-potential young and innovative firms. The second is that some studies suggested that the importance of spatial proximity in VC-investments could have been overestimated in the literature (Fritsch and

that mention venture capital investments as their primary activity. We identify 37 VCs that backed at least one ISC in the period 2012-2019. All the VCs are limited liabilities companies. Nine of these funds are not legally based in Italy: however, these foreign funds are responsible for only 10 out of the 160 investments. As shown in Table 1.4, the VC funds in our sample were founded between 1 and 40 years ago, with an average age of 9 which reflects the relatively recent development of the equity sector in the country. To represent their size, we report the number of employees, of managers, of shareholders and of holdings, and the value of their total assets, revenues and share capital. The funds in our sample have in average 8.2 managers, 3.7 employees, 31 holdings and of 25 shareholders. The VC funds exhibit an average of 10 million EUR in total assets and 0.75 million EUR of share capital. In the same Table 1.4, we report VC-specific descriptive statistics for the dyadic sample[20], for the sample with only the observed VC-ISC pairs, and for a sample in which each Venture Capital firm is represented in the year of the last investment. This is done to show the dyadic sample matches well the original sample of the observed deals in terms of VC characteristics. Also in Table 1.4 we portray the numbers and shares of investments of Venture Capital firms by ISC sector (at 4, 3 and 2 digits NACE code) and by ISC-location (city, province and region), showing that VC investments are more concentrated geographically (in average, 75% of VC deals are in the same region, 8.16% in the same province) than industrially (7.36%

Schilder, 2006; Carlson and Chakrabarti, 2007; Martin *et al.*, 2005), and we claim that empirical evidence for the Italian market supports this view. Our objective is thus to provide an explanation for this discrepancy, at least for the specific population of Italian VC-ISC deals, based on the role of the 'intangible' dimensions of proximity.

20 For the way the dyadic sample was constructed (see Section 1.3.2), one could expect larger VCs that invested in more ISCs during the period under analysis could to appear more often in the dyadic sample than in the original sample containing only the 160 observed VC-ISC deals. Through Table 1.4 we show that this is not the case.

of VC investments are in the same sector at 2 digits level). Finally, we show that 90% of Venture Capital firms have their offices in a Metropolitan Area.

A preliminary investigation of these data reveals that ISCs are fairly distributed across the whole Italian territory and especially present in the South, where by contrast VC actors (with at least one Italian deal) do not have any branch (Figure 1.1). A modified version of the Location Quotient (LQ) (see the Appendix to this Chapter for its construction) points to the existence of large regional equity gaps in the Italian market of VC investments in ISCs. However, somehow unexpectedly, VC locations do not tend to mirror that of Italian innovative startups. Regions with higher rates of VC funds location tend to exhibit above-average shares of VC-backed startups. Overall, this evidence adds to that from which we have started this paper, and seems to confirm that geographical proximity could not represent a reliable, or at least, unique predictor of VC investments in Italian ISCs.

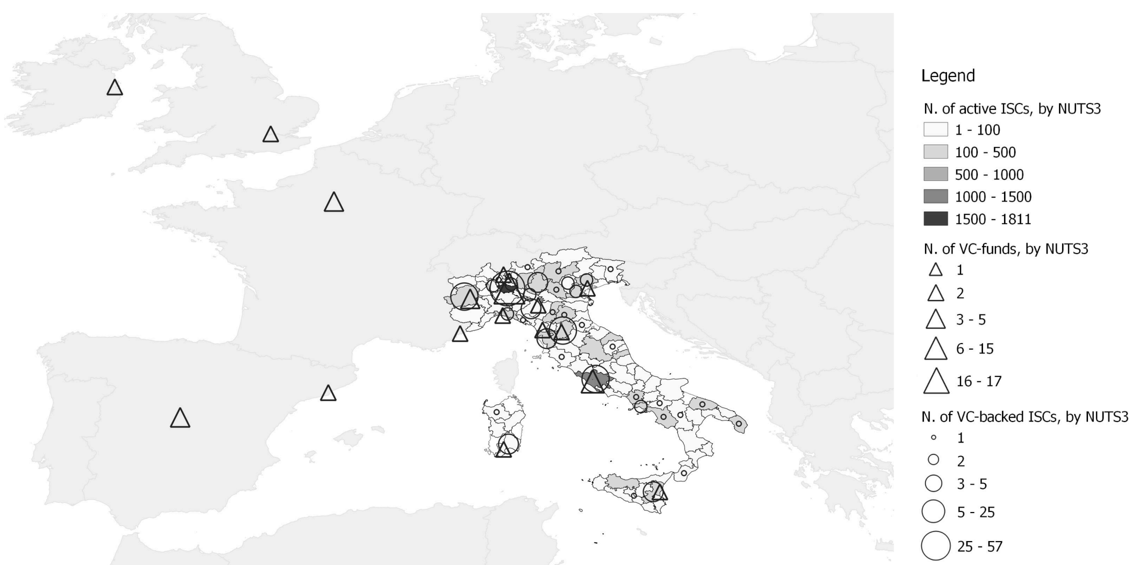


Figure 1.1. Distribution of Italian ISCs, of VCs and VC-backed ISCs in NUTS3 regions

1.3.2 Dependent variable and econometric model

The focal variable of our empirical analysis, Y, is the probability of observing a specific VC-ISC investment pair, of which we aim to investigate the determinants and the role of proximity. In order to do that, we follow an identification strategy that rules out the heterogeneity of firms' financial needs/quality and, following the extant literature (Sorenson and Stuart, 2001; Tykvová and Schertler, 2011), focuses on successfully financed firms and corrects for the entailed selection on the dependent variable.

The set of potential pairs is constructed by considering as bidders all VC funds that completed an investment in an Italian ISC, and as targets any ISC that received a VC investment during a temporal window of 8 months (*id est* within 120 days before or after the original bidder's date of investment): a time-frame consistent with previous evidence on the time evaluation of deals (Petty and Gruber, 2011). Given that 136 startups were backed once or multiple times by 37 different VC funds, for a total of 160 actual investments, their dyadic interaction gave raise to a sample constituted by 8480 dyads built within the above defined time window.

As the proportion of observed pairs represents only 1.89% of the whole dyadic population, lower than the share (i.e. 5%) retained to have a rare event bias (King and Zeng, 2002), we test a number of corrections (see Appendix 1.5) and determine to adopt a Firth (1993) penalized logistic model to estimate the following conditional probability function, Y, to observe the occurrence of a specific VC-ISC investment pair:

$$P(Y = 1|W_{i,j}, X_i, X_j) = \frac{1}{1 + e^{-(W_{i,j}\beta + X_i\gamma + X_j\delta)}} \tag{1.1}$$

where $W_{i,j}$ refers to a set of dyadic proximity variables between j (VC) and i (financed ISC), while X_j and X_i contain investors specific and ISCs control variables, respectively, that will be described in the following Sections.

1.3.3 Proximity variables

(i) Tangible proximity(ies)

Following the arguments we have developed in Section 1.2.1, we build up two sets of variables of tangible proximity between VC and ISC. The first one, *Geographical proximity*, refers to its territorial dimension and measures the inverse of the minimum geodetic distance between the legal and operative offices of the two parties.

The second set of variables aims to capture the functional proximity between VC and ISC. To start with, we define *Functional proximity* with the inverse of the minimum travel time (expressed in hours) separating their respective places, by any means of transport. Alternatively, we consider, in separated alternative specifications, whether such a minimum travel time is by car, within two hours, and within half an hour, by any means of transport.

The set of measurements of tangible proximity is completed by three mutually exclusive dummies, which indicate if at least one among the VC and ISC offices are located in the same city, province, or region.

In order to see whether, by mimicking their role in affecting innovation (Boschma, 2005), the effect of the tangible proximities between VC and ISC on their matching is non-linear, we also plug in the estimates the squared terms of their respective continuous variables (*Geographical proximity* and *Functional proximity*). Indeed, this could serve to see whether an excessive proximity could end up with circumscribing too much the radar of the investment opportunities available to the VC, by making them less attractive than more distant one below a central level of closeness.

Table 1.6 in the Appendix reports the detailed definitions of the previous proximity variables, and the data sources and methodologies used for their calculation. Still in the Appendix, Table 1.5 reports the descriptive statistics for the main proximity measurements. For the dyadic sample used in the analysis, statistics distinguish between potential and successful pairs.

Let us notice that both geographical and functional proximity exhibit higher means for successful than potential investment pairs, supporting the claim that spatial proximity does play a role in predicting the deals.

(ii) Intangible proximities

Consistently with the arguments of Section 1.2.2, our analysis focuses on the role of two kinds of intangible proximities: industrial and relational.

Industrial proximity

In order to capture the industrial proximity between VC and ISC along the dimension we pointed out in Section 1.2.2, accruing to them by sharing the same industry kind of knowledge and experience, we look at the share of previous holdings that each VC fund reveals in the industries in which each partner ISC operates[21]

In particular, in order to investigate the extent to which specific sets of industry knowledge are beneficial for the deal, we build up three proxies of industrial proximity, which compute the previous share for progressively finer levels of industry

21 To compute our measure of industrial proximity we count, in each VC-ISC deal's date, the number of holdings that each VC had in the same 2, 3, or 4-digit industry of each startup and the proportion with respect to the total number of the VC's holdings at that date.

aggregation of the NAICS 2007 classification: 2, 3, and 4 digits. [22].

As shown in Table 1.5 (Appendix), successful VC-ISC pairs tend to exhibit a lower degree of industrial distance, at any industry digit, thus still confirming the role of this variable for VC investments.

Relational proximity

To investigate the role of the relational proximity between VC and target firms we make use of two indicators that refer to the network matrix of the professional and investment links occurred between them before the VC-backing date. We build up this matrix in three steps. In the first step, we identify all the shareholders, holdings, and the name of all the advisors and managers of both the startup and the VC. In the second step, we proceed recursively and collect the same pieces of information for each of the firms or individuals identified in the previous step. More specifically, as shown in Figure 1.2, for shareholding or outward-holding firms we identify the references of advisors, managers, holdings and shareholders; for individuals, such as managers, advisors and individual investors, we detect all previous and contemporaneous professional positions and further investments. We then proceed to the third step, and construct a matrix that report all the undirected links among each VC-backed ISC and each VC firm in our sample.

Using this matrix to proxy the professional and investment relationships established by ISC and VC firms, we first measure the relational proximity between a dyad with the inverse of the minimum number of steps needed to find a link be-

22 In our data, the sectoral classification is provided at 4 digits 2007 NAICS code. In the NAICS, the fourth digit refers to a specific industry group, like 3342 - Communications Equipment Manufacturing, the third indicates the relative sub-sector, like 334 - Computer and Electronic Product Manufacturing, and the first two digits refer to the sector, like 31-33 - Manufacturing.

tween its partners. Furthermore, we measure the intensity of this relationship with the total number of links among the VC and the ISC of each dyad.

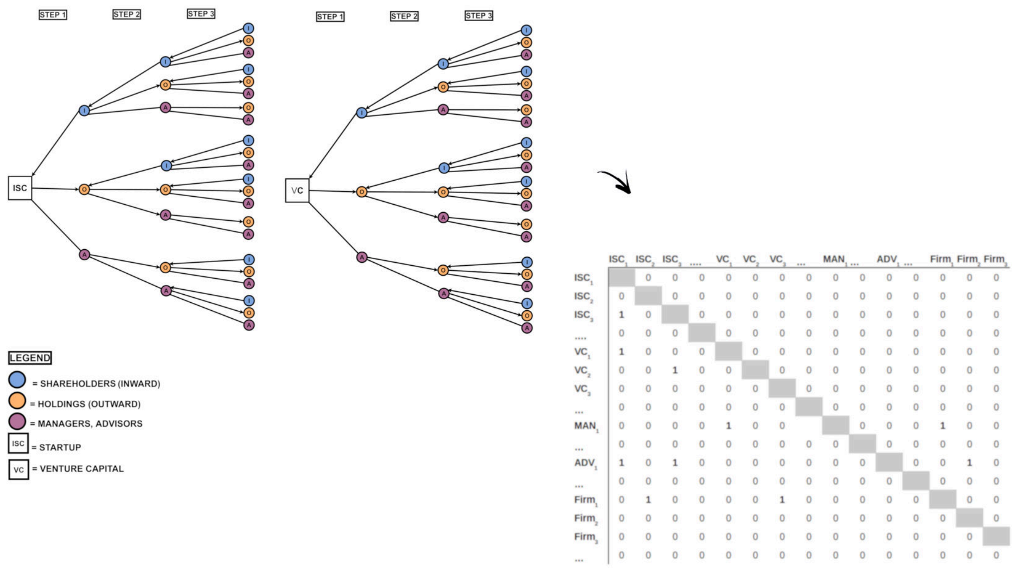


Figure 1.2. Relational network and matrix

Confirming their expected role, both these relational proximity variables exhibit higher means in successful investment-pairs than in potential ones, as shown in Table 1.5 (Appendix).

1.3.4 Control variables

In estimating the role of the previous proximity variables, we should of course retain that firms location choices could correlate both with these variables and with relevant unobservables - such as “the quality of the managerial team” - in turn arguably correlated with the probability that an ISC receives an investment. Furthermore, the model could suffer from another source of endogeneity, as firms location choices could correlate both with tangible and intangible proximities and with relevant unobservables. While this is not sufficient to guarantee causal inference, in order to attenuate the potential bias entailed by these issues, we test numerous ISCs-, VCs-, and location-specific controls. *ISC−specific* controls include: the age of the firm, the number and characteristics of their managers, and the actual fulfilling of each of the innovative requirement to be consider as ISC. We also consider one-year lagged measures of productivity related variables (production costs, costs of research and advertising, per capita value added, value of production, patents rights, labour cost and labour productivity) and of profitability related ones (revenues, debt/equity ratio, return on investments, return on equity; earnings before interest, taxes, depreciation and amortization).[23] As for the *VC−specific* controls, these include size proxies, such as the number of shareholders, managers, employees and companies in the corporate group, along with age, location, and statistics of previous investments. In addition to these characteristics, we also control for whether the ISC had prior VC investments and for if the investment was realised in syndication with other VCs in order to account for the diminished salience of distances in syndicated deals through risk, information and costs sharing (Sorenson and Stuart, 2001)[24]. Finally, we also have a set of *location−specific* controls, retrieved by the Eurostat database at NUTS3 level: population, density, and firms demography by 2-digit NACE code, and the 2000-2018 Italian GDP growth. In order to partially control for the presence

23 Given the presence of missing values for all the above balance-sheet variables, these will only serve to verify the robustness of the identification strategy, and will be omitted in the final models to avoid observations losses.

24 The main focus of our contribution is to looking specifically at the relational proximity that exist between VC-investors and target firms, rather than that between different VCs participating in investments. For the latter has been extensively analysed in numerous previous publications (Tykvová and Schertler, 2014; Catalini and Hui, 2018), and that only 14% of the observed operations were concluded in syndication, we control for it through a dummy variable without further refinements. Anyway, when syndicated investments are excluded from the sample as a robustness check (see Section 1.5) general results in terms of significance of the coefficients hold, but the marginal effect of the relational proximity more than doubles, while those of travel and industrial proximity remain essentially unaltered. If anything, this confirms the relevance of the proposed type of relational proximity.

of specific industrial clusters in the area, the number of active high-growth firms in the province by 2-digit NACE code per year is also considered.This pairs with data on international patent applications to the EPO office, still by two-digit NACE code and location (at NUTS3 level), to account for the innovative capacity of the environment firms operate in. Table 1.3 and 1.4 in the Appendix report the descriptive statistics for the controls. Finally, we retain only those controls that displayed significant coefficients when regressed individually, namely if the investment occurred in syndication, ISCs age, and the lagged GDP at ISC local (NUTS3) level. This last control is important to somehow account for those agglomerative and investor clustering mechanisms for which startups in richer and more successful cities are more likely to attract risk capital investments.

1.4 Results

1.4.1 Baseline model: unpacking spatial proximity

Table 1.1 reports the results of different specifications of a baseline model, where only geographical (Models 1 and 2) and functional proximities (Models 4 - 8) are alternatively considered. Before moving to the illustration of the relative results, let us notice that the retained controls show the expected sign. The fact that the focal ISC has had a prior VC investment increases the probability of its matching with a new one. Consistently with previous studies (Sorenson and Stuart, 2001), syndication provides VC firms with an additional set of information, which increases the chance of a successful matching with a financed ISC. While we mainly observe startup investments, with an average age at finance of 2.8 years as shown in Table 1.3, the startup age is anyhow negatively correlated with the probability of observing a match. Finally, we control for unobserved economic characteristics related to ISCs' location through the (1-year lagged) GDP at NUTS3 level, which exhibit the expected positive and significant sign, and through area (NUTS1 region) fixed effects meant to control for the abiding Italian North-South development divide.

As expected, both geographical and functional proximities, in nearly all the dimensions we have captured it, significantly increase the probability of observing a successful VC-ISC pair[25]. Quite interestingly, when including the squared terms of each of these two distances - in Model 2 and 4, respectively - the effect of spatial proximity appears non-linear. In particular, as Fig. 1.3 (in the Appendix) reveals, after a certain threshold, an increase in the focal proximities reduces, rather than increases, the chance of a successful VC-ISC pair, miming a typical result in innovation studies (Boschma, 2005). In this case, the result suggests that, while facilitating the personal contact between partners, a higher spatial proximity simultaneously reduces the availability of viable investment opportunities and that, after a certain threshold, the latter effect comes to dominate.

While both geographical and functional proximities appear relevant, the lower panel of Table 1.1 shows that being located at a geodetic distance of 200km increases the baseline probability of observing a successful VC-ISC pair by no more than 0.004%. Conversely, being located at a comparable 2hours travel distance has an effect three times higher (up to a positive 0.6% in Model 4), and comparable to the 1.4% average marginal effect of being located *within* a two-hours route (Model 5).

25 The only exception is the co-location within the same province, which is not significant in Model 8. This result could depend on the low number of ISCs and VCs located in the same province but not in the same city, given that most funds and firms locate in metropolitan areas that are in province capitals, and that we exploit a mutually exclusive dummy.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Geographical proximity								
Geographical proximity	0.000***	0.000*						
	(0.000)	(0.069)						
Geographical proximity^2		-0.000						
		(0.836)						
Functional proximity								
Functional proximity			0.195***	0.477***				
(by any means of transport)			(0.000)	(0.001)				
Functional proximity, squared				-0.018**				
(by any means of transport)				(0.046)				
Dummy: Travel time <2hours					1.010***			
(by any means of transport)					(0.000)			
Dummy: Travel time <1/2hours						1.574***		
(by any means of transport)						(0.000)		
Dummy: Minimum travel time is by car							0.881***	
							(0.000)	
Co-location								
Co-location: same city								1.979***
								(0.000)
Co-location: same province (NUTS3)								1.405
								(0.101)
Co-location: same region (NUTS2)								1.451***
								(0.000)
Controls								
Dummy: ISC had prior VC investment	1.035***	1.036***	1.043***	1.038***	1.007***	1.031***	0.976***	1.054***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Dummy: syndicated investment	2.465***	2.463***	2.478***	2.484***	2.511***	2.514***	2.496***	2.479***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ISC's age at finance	-0.122**	-0.125**	-0.111*	-0.096	-0.146**	-0.132**	-0.111*	-0.128**
	(0.043)	(0.020)	(0.066)	(0.117)	(0.016)	(0.027)	(0.062)	(0.033)
L1. GDP, at ISCs NUTS3 (MEUR)	-0.000	-0.000	-0.000	-0.000	-0.000***	-0.000***	0.000	-0.000***
	(0.463)	(0.414)	(0.375)	(0.263)	(0.006)	(0.001)	(0.749)	(0.009)
Area FE	yes	yes	yes	yes	yes	yes	yes	yes
Constant	-3.772***	-3.768***	-3.812***	-3.839***	-4.141***	-3.876***	-4.479***	-4.297***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	8480	8481	8482	8483	8484	8485	8486	8487
VIF	1.251	1.251	1.252	1.252	1.415	1.306	1.334	1.345

Estimated probability. Robust pval in parenthesis (*** p<0.01, ** p<0.05, * p<0.1)

Marginal effects (covariates at median values)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Geographical proximity at 200km	4.21e-05	4.78e-05						
Geographical proximity at 50km	4.21e-05	4.78e-05						
Functional proximity: at 2hours			0.00254	0.00598				
Functional proximity: at 1/2hour			0.00256	0.00605				
Dummy: Travel time <2hours =1					0.0146			
Dummy: Travel time <1/2hours =1						0.0466		
Dummy: Minimum travel time is by car =1							0.0192	
Same city								0.0518
Same province								0.0367
Same region								0.0379

Table 1.1. Baseline model

Finally, the effect is appreciable also with respect to the dummy for the minimum travel route by car (+ 1.9%) and substantially more for being located within half an hour of travelling (+4.7)%.[26]

26 Being co-located in the same city or region has the strongest effects, +5.2% and +3.8% respectively. However, these measures appear highly correlated with relational proximity, suggesting that they could confound with it and cannot be considered as purely *spatial*.

In conclusion, consistently with recent evidence about the managers' perception of their comparative relevance, we find that the geodetic distance between partners is not the most accurate proxy for predicting successful VC deals. Indeed, the relational needs involved in this type of investments are the most favored by the accessibility easiness of the partners, reflected by a functional rather than geographical kind of proximity. However, like the relational one, also the functional proximity between partners only helps up to a certain extent, passing which its limiting the set of viable deal opportunities prevails.

1.4.2 Augmented model: beyond spatial proximity

Table 1.2 reports the results for the model in which the role of tangible proximity is augmented with that of the intangible ones and with their respective interactions. Given the natural correlation between geographical and functional proximities, and the higher explicative role of the latter documented in the previous section, in all the specifications of this augmented model we only retain a functional kind of tangible proximity. More precisely, among the different proxies of this functional proximity, we notice that the within two-hours travel dummy between partners does not appear significantly correlated with any measure of intangible proximity (see Table 1.9 in the Appendix). Furthermore, it exhibits a high predictive power with respect to the outcome variable and consistent marginal effects among different specifications. On this basis, we will stick to this variable of functional proximity in all of the specifications of Table 1.2.

Starting with the role of industrial proximity, Models (1) - (3) show that it significantly increases the chance of a successful VC-ISC pair, but with some important specifications.

First of all, the variable at stake reveals significantly positive only for the low to intermediate degrees of sectoral disaggregation that we have considered: that is, at 2 and 3 digits of the NAICS classification. This interestingly suggests that when the industry-group environment that the partners share is very specific – 4 digits, like an ISC in Basic Chemical Manufacturing targeted by a VC with previous experience of it – the entailed learning return for the VC is possibly too narrow to make the VC retain the deal sufficiently enriching to be concluded. Conversely, at the sector (2digits) and subsector (3 digit) level, the enabling mechanisms of the VC-ISC matching we have envisaged in Section 1.2.2 seem to work.

As a second nuance, Fig. 1.3 shows that the effect of industrial proximity is evidently non-linear and, somehow surprisingly when thinking of its effect on innovation (Noteboom, 2000, Boschma, 2005 and Torre and Rallet, 2005), convex rather than concave. In interpreting this result, we should bear in mind that, in the case of VC investments, the relational disadvantages of a scarce and of an excessive alignment of the partners' industrial experience could be compensated by the diversification and risk-management strategies of funds (see Baldi *et al.*, 2015). On the one hand, deals in industries with zero or very few previous investments could be more attracting in terms of risk and portfolio diversification strategies; on the other hand, deals in sectors where the fund is highly specialized could be more valuable for the lower informational barriers they entail. In concluding the analysis of industrial proximity, Model (7) in Table 1.2 shows that, as we argued in Section 1.2.2, its role in driving the match between ISC and VC interoperates with that of tangible proximity. More precisely, the interaction between industrial proximity (at 3 digits) and the dummy for travel times below 2 hours turns out significantly negative. As we have hypothesised, the experience entailed by industrial proximity could increase the VC knowledge sources to the point of compensating it for the disadvantages of targeting more geographically distant ISCs. In brief, the industrial proximity between VC and target companies is capable to extend the geographical coverage of their relationship. The same is true for the interaction between industrial proximity (at 3 digits) and relational proximity, which again highlights a positive degree of substitutability among these two intangible proximity dimensions.

Coming to the role of relational proximity, the first indicator with which we have tried to capture it is significantly positive (Model 4). This suggests that, as expected, the interlink between VC and ISC created by their investment and/or professional relationships facilitates their matching. The relevance of relational proximity gets confirmed when its intensity is considered (Model 7) and appears a reliable result consistently with previous studies about the role of relational networks in VC deal selection (Catalini and Hui, 2018).

Looking at the marginal effects that relational proximity reveals at different steps of network distance between pairs, two important specifications add to the extant knowledge about it (lower panel of Table 1.2). First, while relevant for concluding a VC deal, the network position of the involved actors has an impact that decays rapidly with the number of their separating steps: being at one step distance increases twofold (+109.7%) the probability of observing a successful pair, while the marginal effect lowers at a +16.5% when the network distance is two steps. Second, unlike that of spatial proximity, the effect of the relational one does not appear bounded and rather shows an exponential impact on the creation of a successful VC-ISC pair. This is an important result, confirmed by the non-linear pattern that we observe in Fig. 1.3 (in the Appendix). Finally, the marginal effects of having

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Functional proximity							
Dummy: travel time<2hours	1.005***	1.007***	1.008***	1.060***	0.844***	1.036***	1.429***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Industrial proximity							
Industrial proximity	1.005***						
2 digits	(0.000)						
Industrial proximity		1.019***				0.498*	2.023***
3 digits		(0.000)				(0.052)	(0.000)
Industrial proximity			0.484				
4 digits			(0.183)				
Relational proximity							
Relational proximity				3.788***		3.535***	4.297***
				(0.000)		(0.000)	(0.000)
N. of ties					1.556***		
(intensity)					(0.000)		
Interactions							
Industrial proximity 3digits * Dummy travel time<2hours							-1.423***
							(0.008)
Relational proximity*Industrial proximity							-1.367*
							(0.069)
Relational proximity * Dummy travel time<2hours							0.152
							(0.848)
Controls							
Dummy: ISC had prior VC investment	1.095***	1.084***	1.027***	1.087***	1.096***	1.116***	1.118***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Dummy: syndicated investment	2.531***	2.525***	2.505***	2.591***	2.527***	2.595***	2.594***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ISC's age at finance	-0.153**	-0.150**	-0.148**	-0.150**	-0.131**	-0.156**	-0.145**
	(0.012)	(0.013)	(0.014)	(0.015)	(0.035)	(0.012)	(0.020)
L1. GDP, at ISCs NUTS3 (MEUR)	-0.000**	-0.000**	-0.000***	-0.000***	-0.000**	-0.000***	-0.000***
	(0.012)	(0.011)	(0.007)	(0.001)	(0.013)	(0.001)	(0.001)
Area FE	yes	yes	yes	yes	yes	yes	yes
Constant	-4.394***	-4.385***	-4.183***	-4.802***	-4.237***	-4.837***	-5.376***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	8,473	8,473	8,473	8,473	8,473	8,473	8,473
VIF	1.377	1.376	1.375	1.375	1.375	1.348	2.771
Estimated probability. Robust pval in parenthesis (*** p<0.01, ** p<0.05, * p<0.1)							
Marginal effects (covariates at median values)							
Dummy: Travel time <2hours =1	0.0116	0.0117	0.0140	0.0131	0.0102	0.0116	0.0166
Industrial proximity at 33%	0.0161	0.0166	0.00788			0.00656	0.0457
Industrial proximity at 50%	0.0191	0.0197	0.00855			0.00714	0.0645
Industrial proximity at 75%	0.0234	0.0242	0.00942			0.00789	0.0967
Relational proximity at 3 steps-distance				0.0867		0.0704	0.100
Relational proximity at 2steps-distance				0.165		0.128	0.209
Relational proximity at 1step-distance				1.097		0.751	1.789
N. of ties, at 1 tie (Relational)					0.0314		
N. of ties, at 2 ties (Relational)					0.0409		
N. of ties, at 3 ties (Relational)					0.0559		

Table 1.2. Augmented model

one and three ties are of 3.1% and 5.6%, respectively, showing that the impact of an additional tie between VC and ISC also increases rapidly.

In concluding the analysis of relational proximity, Model (7) in Table 1.2 reveals that, inconsistently with our expectations, this proximity does not behave like the industrial one in negatively moderating the role of tangible proximity. Indeed, the interaction between the former and the dummy for the functional distance to be lower than 2 hours, is not significant. Unlike what Sorenson and Stuart, 2001 found with respect to the relational proximity between syndicated VC firms, that between VC and ISC is not capable to compensate for the disadvantages that a longer geographical distant between the two could entail. In other words, while getting closer and trusted relationships with other VC firms might render, according to Sorenson and Stuart, 2001, the focal VC more spanning its investments across space, having the same kind of relationships with potential ISC does not render less close investments more palatable to VC firms. While both kinds of relational proximity facilitates the match – as revealed by the relative dummy among the controls of Table 1.2 – their power of widening the geographical scope of the relationship is instead different and limited to the relational proximity among VCs.

As Table 1.9 (in the Appendix) reveals, industrial and relational proximities are significantly correlated, which could diminish the reliability of their estimated coefficients. However, the variance inflation factor reported for each model, provides reassurance regarding this issue. [27]. Furthermore, it is possible to compare the magnitude and significance of their coefficients when simultaneously included (Model 6 and 7 in Table 1.2), which indicate that relational proximity is the one that matters the most in determining the VC-ISC match. The same indicative result would emerge by considering the full Model (7), in which all the proximities at stake (functional, industrial and relational) and all their addressed interactions are considered. In this same model, an interesting substitutive relationship would also

27 In the interests of synthesis we only report each model's mean VIF in the bottom of tables 1.1 and 1.2. However, we controlled the Variance Inflation Factor for each regressor and model, finding a value above 5 only for the interaction term between the travel proximity dummy and relational proximity (VIF of 7.01) in Model 7, accounting for a model Mean VIF of 2.9.

emerge between industrial and relational proximity. The negative coefficient of this interaction term implies that, in the absence of industry-specific knowledge sharing with the ISCs, VCs can rely on their relational proximity to acquire the necessary information and finalize the deal. Conversely, industrial proximity may compensate for lower degrees of relational proximity.

1.4.3 Additional estimates and robustness checks

The results that we have obtained about the role of proximities in driving the probability of a successful VC-ISC pair appear substantially robust to two important checks, which we also report in Appendix 1.5.

The first robustness check concerns the possible presence of spatial autocorrelation in the phenomenon we are investigating. Indeed, this could be suggested by the spatial inequalities in the distribution of startup firms and VC offices we have highlighted in Section 3.4. Excluding, in turn, ISCs and VCs located in agglomerated areas, has the sole effect of increasing the magnitude of the coefficients of the individual and interacted intangible proximity terms. This result provides further evidence on the hypothesis that intangible proximities effectively counteract the barrier of physical distance with respect to Venture Capital funds.The second robustness check that we perform concerns the presence of idiosyncratic VC investments and funds, as results might be affected by second and syndicated VC investments, which the literature has shown to differ from first and solo ones (Berchicci *et al.*, 2011, Catalini and Hui, 2018, Cumming and Dai, 2010). The main results of our analysis do not change when carrying out the previous checks (see Appendix 1.5).

1.5 Conclusions

Being one of the possible sources of regional gaps in innovation, the geography of financial investments in innovative start-ups requires high academic and policy attention. In particular, the existing knowledge about the pervasive diffusion of local biases in VC investments is required to confront with emerging evidence about an apparent non-sensitivity of managers to spatial proximity in selecting the deals.

An important advancement in understanding the geographical distribution of VC investments, and to possibly account for this and possibly other clashes, can be obtained by making the literature on (innovation) financing talk more with that of economic geography: in particular, by drawing from the latter a manifold notion of proximity, which considers its tangible and non-tangible variants, and the possible non-linearity and interactions of their effects.

In this vein, with this chapter we have proposed an original investigation of the role that different dimensions of proximity can have in predicting a successful matching between VC funds and innovative startups seeking for finance. This new theoretical framework has been applied to an original investigation of the Italian VC market: an immature kind of market, whose knowledge is still scanty and in need of more scrutiny when compared to other countries. In contributing to fill this gap, we have exploited a recent legal (i.e. exogenous) identification of innovative start-up companies in Italy, which has allowed us to consider their entire population. By combining different sources, we have obtained a rich dataset, with which to build up already known and novel proxies of their proximity to VC funds and to address the role of these proximities in their successful matching.

Our results have first of all provided updated evidence of the existence and magnitude of regional gaps in the Italian VC market, showing that VC offices are clustered and polarised in some portions of the territory. In investigating the determinants of these gaps, we have first of all found that spatial proximity matters more in functional than in geographical terms, that is, in facilitating the accessibility (by car more by plane) of partners rather than in reducing their distance. Furthermore, the effect is concave and points to the case of an excessive proximity for the deal to occur, in both respects. These results convey a systematic generalisation of what has emerged from specific studies reporting managers' statements about their indifference and preference for a short physical and travel distance, respectively (Fritsch and Schilder, 2006 and Martin *et al.*, 2005). The implications of such a result are also particularly important. This is so both for future research on local biases in innovation financing, which are encouraged to incorporate a more nuanced idea of spatial proximity; and for policy makers, who should consider the development of local transport infrastructures a crucial leverage to promote effective VC deals.

We have also shown that the relationship between VC and innovative start-ups in search of finance is helped by a varied set of intangible proximities too. Among these proximities, relational proximity, in terms of professional and investment networks, emerges as the strongest predictor of the VC-ISC matching. Furthermore, unlike the other proximities, which show a non-monotonic relationship with it, the relational one uniquely exhibits a positive exponential trend with respect to the probability of observing a successful VC-ISC pair. Also this result generalises and integrates previous findings in the financial literature (Catalini and Hui, 2018, Hermann *et al.*, 2016, Sorenson and Stuart, 2001) and has important implications. On the one hand, future research should more closely look at the role of networks in facilitating start-ups in search of financing: in particular, by addressing how strategic holdings in firms directly connected to VC funds could improve their access to risk capital. On the other hand, policy makers should consider that networking incentives could complement or inform public policies aimed at addressing funding gaps and at supporting the development of the VC market.

A last crop of interesting results concerns the specific contingencies under which industrial proximity appear to help the matching between VC and ISCs, as well as the substituting relationship we have detected between industrial and relational proximity.

As usual, our empirical analysis is not free from limitations. Although, as we noticed, the unobserved heterogeneity in the quality of startup projects could be ruled out through the application of a dyadic model, other endogeneity issues could remain. In order to address them, future research should concentrate on the identification of valid instruments for making our focal regressors - both relational and functional proximity - exogenous.A second limitation is represented by the focus of our analysis on a country with low financial-development, like Italy. While suited for the analysis of the role of proximity in mitigating regional equity gaps, such a choice obviously hinders the external validity of the results. A follow-up application in a cross-country framework would thus be required for the sake of generalization of the results that we have obtained.

References

Adam, F., (2008). 'Mapping Social Capital across Europe: Findings, Trends and Methodological Shortcomings of Cross-National Surveys'. *Social Science Information*, 47 (2): 159–86.

Aghion, P., Fally, T., Scarpetta, S. (2006). Credit Constraints as a Barrier to the Entry and Post-Entry Growth of Firms. *Economic Policy*, 22(52): 731+733-779.

Alessandrini, P., Presbitero, A.F., Zazzaro, A. (2009). Banks, distances and firms' financing constraints. *Review of Finance*, 13, 261–307.

Alessandrini, P., Presbitero, A.F., Zazzaro, A. (2010). Bank size or distance: what hampers innovation adoption by SMEs. *Journal of Economic Geography*, 10(6): 845–881.

Alperovych Y., Groha, A., Quas, A., (2020). Bridging the equity gap for young innovative companies: The design of effective government venture capital fund programs. *Research Policy*, 49(10): 104051.

Antonietti, R., Cainelli, G., Ferrari, M., Tomasini, S. (2015). Banks, related variety and firms' investments, *Letters in Spatial and Resource Sciences*, 8(1): 89-99.

Antonietti, R., Gambarotto, F. (2020). The role of industry variety in the creation of innovative start-ups in Italy. *Small Business Economics*, 54: 561-573.

Bacaksiz, E., Koç, S., (2021). Comparison of Different Estimation Approaches in Rare Events Data. *Ege Academic Review*, 21(3): 263–272.

Baldi, F., Baglieri, D., and Corea, F. (2015). Balancing risk and learning opportunities in corporate venture capital investments: Evidence from the biopharmaceutical industry. *Entrepreneurship Research Journal*, 5(3): 221–250.

Balland, P.-A., Boschma, R., Frenken, P., (2013) Proximity and innovation networks: an evolutionary approach. In *Cooke, P. (Ed.). (2013). Re-framing Regional Development. London: Routledge*, 186-200.

Battisti, G., Stoneman, P., (2019). Defining and Measuring the Innovativeness of Firms. *Academy of Management Proceedings*, 1: 15184.

Beck, T., Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking and Finance*, 30(11):2931–2943.

Bender, M. (2010). *Spatial Proximity in Venture Capital Financing. A Theoretical and Empirical Analysis of Germany*. Gabler Research, (Entrepreneurial and Financial Studies), 1-330.

Berchicci, L., Block, J. H., Sandner, P. G. (2011). The Influence of Geographical Proximity and Industry Similarity in a Business Angel's Investment Choice. *Ssrn*.

Bertoni, F., Colombo, M. G., Croce, A., Piva, E. (2007). A review of the venture capital industry in Italy. In G. N. Gregoriou, M. Kooli and R. Kraeussl, (Eds.), *Venture Capital in Europe, pp. 129–141. Elsevier.*

Bertoni, F., Colombo, M.G., Grilli, L., (2011). Venture capital financing and the growth of high-tech start-ups: disentangling treatment from selection effects. *Research Policy*, 40: 1028–1043.

Bollazzi, F., Risalvato, G., Venezia, C. (2019). Asymmetric information and deal selection: evidence from the Italian venture capital market. *International Entrepreneurship and Management Journal*, 15(3): 721–732.

Bonini, S., Perrone, V. E., Zocchi, P. (2019). The performance of angel-backed companies. *Journal of Banking & Finance*, 100: 328-345

Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1): 61–74.

Bragoli, D., Burlina, C., Cortelezzi, F., Marseguerra, G. (2022). Banking proximity and firm performance. The role of small businesses, community banks and the credit cycle, Applied Economics, 54(57): 6605-6624.

Brown, L. A., and Horton, F. E., (1970). Functional distance: an operational approach. *Geographical Analysis*, 2: 76-83.

Buchner, A., Mohamed, A., Schwienbacher, A. (2017). Diversification, risk, and returns in venture capital. *Journal of Business Venturing*, 32(5): 519–535.

Carlson, C., Chakrabarti, P. (2007). Venture Capital Investment in Secondary Cities: Issues and Opportunities for Impact. *Federal Reserve Bank of Boston, Public and Community Affairs Discussion Papers.*

Carpenter, R. E., Petersen, B. C. (2002). Capital market imperfections, high-tech investment, and new equity financing. *Economic Journal*, 112(477): F54-F72.

Carrincazeaux, C., Lung Y., Rallet, A., (2001). Proximity and localisation of corporate R&D activities. *Research Policy*, 30: 777–789

Catalini, C., Hui, X. (2018). Online Syndicates and Startup Investment. *NBER Working Paper*, 24777.

Caviggioli, F., Colombelli, A., De Marco, A., Paolucci, E. (2020). How venture capitalists evaluate young innovative company patent portfolios: empirical evidence from Europe. *International Journal of Entrepreneurial Behavior & Research*, 26(4): 695–721.

Coleman, J. S. (1994). *Foundations of social theory.* Harvard university press.

Colombelli, A. (2016). The impact of local knowledge bases on the creation of innovative start-ups in Italy. *Small Business Economics*, 47: 383–396.

Colombelli, A., Grilli, L., Minola, T., Mrkajic, B. (2020). To what extent do young innovative companies take advantage of policy support to enact innovation appropriation mechanisms?. *Research Policy*, 49(10): 103797.

Colombo, M. G., Grilli, L. (2007). Funding gaps? Access to bank loans by high-tech start-ups. *Small Business Economics*, 29(1–2): 25–46.

Colombo, M. G.T., Mustar, P., Wright, M. (2010). Venture capital and high-tech start-ups. *Venture Capital*, 12(4): 261-266.

Cumming, D., Johan, S. A. B. (2007). Provincial preferences in private equity. *Financial Markets and Portfolio Management*, 20(4): 369–398.

Cumming, D., Dai, N. (2010). Local bias in venture capital investments. *Journal of Empirical Finance*, 17(3): 362–380.

Davids, M., Frenken, K. (2018). Proximity, knowledge base and the innovation process: towards an integrated framework. *Regional Studies*, 52(1): 23–34.

Davila, A., Foster, G., Gupta, M. (2003). Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, 18(6): 689–708.

De Socio A. (2010). La Situazione Economico-Finanziaria delle Imprese Italiane. *Banca d'Italia, Occasional Papers*, 66.

Dejuán, D., Mora-Sanguinetti, J., (2019). 'Quality of Enforcement and Investment Decisions. Firm-Level Evidence from Spain'. Working Paper, 1927, *Banco de España*. https://econpapers.repec.org/paper/bdewpaper/1927.htm

Della Sala, V., (2004). The Italian model of capitalism: on the road between globalization and Europeanization? *Journal of European Public Policy*, 11(6): 1041-

1057.

Donati, C., Sarno, D. (2015). Are firms in “backward” areas of developed regions more financially constrained? The case of Italian SMEs. *Industrial and Corporate Change*, 24(6): 1353–1375.

Elgmati, E., Fiaccone, R., Henderson, R., Matthews, J., (2015) Penalised Logistic Regression and Dynamic Prediction for Discrete-Time Recurrent Event Data. *Lifetime Data Analysis*, 21(4):542–560.

Fabbri, D. (2010), Law Enforcement and Firm Financing: Theory and Evidence, *Journal of the European Economic Association*, 8(4): 776-816.

Finaldi Russo, P., Magri, S., Rampazzi, C. (2016). Innovative start-ups in Italy: their special features and the effects of the 2012 law. *Politica Economica/Journal of Economic Policy*, 32(2): 297–330.

Firth, D., (1993). Bias Reduction of Maximum Likelihood. *Biometrika*, 80(1): 27–38.

Florida, R., Mellander, C. (2016). Rise of the startup city: The changing geography of the venture capital financed innovation. California Management Review, 59(1): 14–38.

Friedkin, N. E. (1998). A structural theory of social influence. *Cambridge University Press*.

Fried, V.H., Hisrich, R.D. (1994). Toward a Model of Venture Capital Investment Decision Making. *Financial Management*, Venture Capital Special Issue, 23(3): 28-37.

Fritsch, M., Schilder, D. (2006). Does venture capital investment really require spatial proximity? An empirical investigation. *Freiberger Arbeitspapiere*, 2006(07).

Fuchs, F., Füss, R., Jenkinson, T., Morkoetter, S. (2021). Winning a deal in private equity: Do educational ties matter? Journal of Corporate Finance, 66: 101740.

Garmendia, A., Llano, C., Minondo, A., Requena, F. (2012). Networks and the disappearance of the intranational home bias. *Economics Letters*, 116(2): 178–182.

Giacomelli, S., Menon, C., (2016). Does Weak Contract Enforcement Affect Firm Size? Evidence from the Neighbour's Court. *Journal of Economic Geography*, 17 (6): 1251–82.

Ghio, N., Guerini, M., Rossi-Lamastra, C. (2016). University knowledge and the creation of innovative start-ups: an analysis of the Italian case. *Small Business Economics*, 47: 293–311.

Giraudo, E., Giudici, G., Grilli, L. (2019). Entrepreneurship policy and the financing of young innovative companies: Evidence from the Italian Startup Act. *Research Policy*, 48(9): 103801.

Granovetter, M., (1985) Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3): 481-510

Grilli, L., Murtinu, S. (2014). Government, venture capital and the growth of European high-tech entrepreneurial firms. *Research Policy*, 43(9): 1523–1543.

Grilli, L., (2019). There must be an angel? Local financial markets, business angels and the financing of innovative start-ups, *Regional Studies*, 53:5, 620-629.

Groh, A. P., von Liechtenstein, H., Lieser, K. (2010). The European Venture Capital and Private Equity country attractiveness indices. *Journal of Corporate Finance*, 16(2): 205–224.

Gupta, A. K., Sapienza, H. J. (1992). Determinants of venture capital firms' preferences regarding the industry diversity and geographic scope of their investments. *Journal of business Venturing*, 7(5): 347-362.

Hain, D., Johan, S., Wang, D. (2016). Determinants of Cross-Border Venture Capital Investments in Emerging and Developed Economies: The Effects of Relational and Institutional Trust. *Journal of Business Ethics*, 138(4): 743–764.

Harrison, R. T., Mason, C. M. (2002). The Geography of Venture Capital Investments in the UK. *Transactions of the Institute of British Geographers*, 27(4): 427-451.

Hermann, J., Avdeitchikova, S., Hjertström, A., (2016). The influence of functional and relational proximity on business angel investments. *International Journal of Entrepreneurship and Small Business*, 29(3): 468-490.

Hess, M. (2004). "Spatial" relationships? Towards a reconceptualization of embeddedness. *Progress in Human Geography*, 28(2): 165–186.

Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, andventure capital funding. *Research Policy*, 36(5): 722-741.

Iorio, R. D'Amore, R., (2017). Are Small Firms Less Innovative than Large Ones? A Statistical Overview on an European Context and the Peculiarities of Italy. *Sociologia Del Lavoro*, 147: 81-107.

Istat, (2014). Labour Market Areas Year 2011. https://www.istat.it/it/files//2014/12/EN_Labour-market-areas_2011.pdf

Johnson, S., J. McMillan and C. Woodruff (2002). Courts and Relational Contracts, *Journal of Law, Economics and Organization*, 18(1): 221-277.

King, G., Zeng. L., (2001). Logistic regression in rare events data. *Political Analysis*, 9: 137–163.

King, G., Zeng, L. (2002). Estimating risk and rate levels, ratios and differences in case-control studies. *Statistics in Medicine*, 21(10): 1409–1427.

Kortum, S., Lerner, J., 2000. Assessing the contribution of venture capital to innovation. *RAND Journal of Economics*, 31(4): 674–692.

Lee, N., Drever, E., (2014). Do SMEs in deprived areas find it harder to access finance? Evidence from the UK Small Business Survey. *Entrepreneurship and Regional Development*, 26(3-4): 337-356.

Lee, N., Brown, R. (2017). Innovation, SMEs and the liability of distance: The demand and supply of bank funding in UK peripheral regions. *Journal of Economic Geography*, 17(1): 233–260.

Lee, N., Luca, D. (2019). The big-city bias in access to finance: Evidence from firm perceptions in almost 100 countries. *Journal of Economic Geography*, 19(1): 199–224

Leitgöb, H., (2020) Analysis of rare events. In: *Atkinson, P.; Delamont, S.; Cernat, A.; Sakshaug, J.W. ; Williams, R.A.. 'SAGE Research Methods Foundations'.* Thousand Oaks, SAGE, 2020.

Lutz, E., Bender, M., Achleitner, A. K., Kaserer, C. (2013). Importance of spatial proximity between venture capital investors and investees in Germany. *Journal of Business Research*, 66(11): 2346–2354.

Martin R., Sunley, P., Turner, D., (2002) Taking risks in regions: the geographical anatomy of Europe's emerging venture capital market. *Journal of Economic Geography*, 2(2): 121–150.

Martin, R., Berndt, C., Klagge, B., Sunley, P., (2005). Spatial proximity effects and regional equity gaps in the venture capital market: Evidence from Germany and the United Kingdom. *Environment and Planning*, 2005, volume 37, pages 1207-1231

Mason, C. M., Harrison, R.T„ (1992) The supply of equity finance in the UK: a strategy for closing the equity gap, *Entrepreneurship and Regional Development*, 4(4): 357-380.

Mason, C., (2007) Venture capital: a geographical perspective. In: H Landström (ed) *Handbook of Research on Venture Capital*, Cheltenham: Edward Elgar, pp 86-112.

Mason, C., Pierrakis, Y. (2013). Venture Capital, the Regions and Public Policy: The United Kingdom since the Post-2000 Technology Crash. *Regional Studies*, 47(7): 1156–1171.

Meuleman, M., Jääskeläinen, M., Maula, M. V. J., Wright, M. (2017). Venturing into the unknown with strangers: Substitutes of relational embeddedness in cross-border partner selection in venture capital syndicates. *Journal of Business Venturing*, 32(2): 131–144.

Midrigan, V., Xu D. Y., (2014). Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review*, 104(2): 422-58.

Milosevic, M. (2018). Skills or networks? Success and fundraising determinants in a low performing venture capital market. *Research Policy*, 47(1): 49–60.

Mistrulli P., E., Vacca, V., P., (2015). Social Capital and the Cost of Credit: Evidence from a Crisis. *Bank of Italy*, Temi di Discussione, 1009.

Moodysson J., Jonsson O., (2007). Knowledge Collabouration and Proximity: The Spatial Organization of Biotech Innovation Projects. *European Urban and Regional Studies*, 14: 115–31.

Moran, P., (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, 10:243-251.

Moro, A., Maresch, D., Ferrando, A., (2018). Creditor Protection, Judicial Enforcement and Credit Access. *The European Journal of Finance* 24(3): 250–281.

Nigam, N., Benetti, C., Johan, S. A. (2020). Digital start-up access to venture capital financing: What signals quality? *Emerging Markets Review*, 45: 100743.

Norton, E., Tenenbaum, B. H. (1993). Specialization versus diversification as a venture capital investment strategy. *Journal of Business venturing*, 8(5): 431-442.

Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management and Governance*, 4(1–2): 69–92.

OECD, (2017). OECD Economic Surveys: Italy 2017, *OECD Publishing*, Paris.

Patzelt, H., zu Knyphausen-Aufseß, D., Fischer, H. T. (2009). Upper echelons and portfolio strategies of venture capital firms. *Journal of Business Venturing*, 24(6): 558–572.

Peneder, M., (2010). The impact of venture capital on innovation behaviour and firm growth. *International Journal of Entrepreneurial Finance*, 12(2),83–107

Petersen, M.A., Rajan, R.G. (2002). Does distance still matter? The information revolution in small business lending. *The Journal of Finance*, 57: 2533–2570.

Petty, J. S., Gruber, M., (2011). In pursuit of the real deal. A longitudinal study of VC decision making. *Journal of Business Venturing*, 26(2): 172–188.

Polanyi K. (1944) *The Great Transformation*. Beacon, Boston.

Pollard, J. S., (2003). Small firm finance and economic geography. *Journal of Economic Geography*, 3(4): 429–452.

Puhr, R., Heinze, G., Nold, M., Lusa, L., Geroldinger, A., (2017) Firth's logistic regression with rare events: accurate effect estimates and predictions? *Statistics in Medicine*, 36(14): 2302-2317.

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, 24: 339–355.

Schiantarelli, F., Stacchini, M., Strahan., P.E., (2020). Bank Quality, Judicial Efficiency, and Loan Repayment Delays in Italy'. *The Journal of Finance*, 75 (4): 2139–2178.

Shane, S,, Cable D., (2002) Network ties, reputation, and the financing of new ventures. *Management Science*, 48(3): 364–381

Sorenson, O., Stuart, T., (2001). Syndication Networks and the Spatial Distribution of Venture Capital Investments. *American Journal of Sociology*, 106(6): 1546–1588.

Teten, D., Farmer, C., (2010). Where are the deals? Private equity and venture capital Funds' best practices in sourcing new investments. *Journal of Private Equity* 14(1): 32–53.

Tykvová, T., Schertler, A., (2011). Venture capital and internationalization. *International Business Review*, 1(6): 423-439.

Tykvová, T., Schertler, A., (2014). Does syndication with local venture capitalists moderate the effects of geographical and institutional distance? *Journal of International Management*, 20(4): 406–420.

Torre, A., Rallet, A. (2005). Proximity and localization. *Regional Studies*, 39(1): 47–59.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*, 61: 674–698.

Uzzi, B., (1999). Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking finance. *American Sociological Review*, 64: 481–505.

Vacca, V., (2013). Financing innovation in Italy: an analysis of venture capital and private equity investments. *Banca d'Italia*, Occasional Paper.

Van Osnabrugge, M., (2000). A comparison of business angel and venture capitalist investment procedures: An agency theory-based analysis. *Venture Capital*, 2(2): 91–109.

Wang, J., Wang, Y., Wu, C., Yang, X., Zhao, L., (2021). Social Proximity, Information, and Incentives in Local Bank Lending. *The Review of Corporate Finance Studies*, cfab011.

Whited, T. M., Zhao, J. (2021). The Misallocation of Finance. *Journal of Finance*, 76(5): 2359–2407.

Yang, S., Keller, F., Zheng, L. (2017). Basics of social network analysis. In: Social network analysis, SAGE Publications.

Zhao, T., Jones-Evans, D., (2017). SMEs, banks and the spatial differentiation of access to finance. *Journal of Economic Geography*, 17(4): 791–824.

Zook, M. A., (2002). Grounded capital: Venture financing and the geography of the internet industry, 1994-2000. *Journal of Economic Geography*, 2(2): 151–177.

Appendix

Location Quotient

Following Martin *et al.* (2005), in order to detect regional concentrations of VC investments that are below or above the average, we calculate a modified version of the location quotient usually applied in the literature on industrial clusters[28] defined as the ratio between: at the numerator, the share of VC-backed startups located in the region over the total of VC-backed Italian innovative startups; at the denominator, the share of all firms created in the region over the total of new firms in the country in the same period of time (2012-2018).[29]

The location quotient (LQ) denotes, at both NUTS3 and NUTS2 level, an over-concentration of investments in six of the twenty NUTS2 Italian regions, with the region of Milan (Lombardy) exhibiting an LQ 2.5 times above one (and reaching 5.3 at NUTS2 level in its capital province). The majority of the other regions instead exhibit an under-concentration of VC-capital investments, especially those located in the South and in the Northeast of the country, where a large number of firms are located but both VC funds and VC-investments are lacking. Overall, the indicator actually points to the existence of large regional equity gaps in the Italian market of VC investments in ISCs.[30]

---

28 The location quotient indicates over-concentration when its value is above 1, and under-concentration for values below 1.

29 Data for firms demography are issued by Infocamere,Labour Market Areas would have possibly been a more suitable geographical level of analysis for this indicator, but NUTS3 regions where the most fine-grained data on firms demography available for the same years of the ISCs' sample.

30 As we will see, to account for such a disparity, which is likely to derive not only by VC location and local bias, but also by the persistence of the Italian North-South development and productivity divide, area fixed effects will be included in the model.

Rare Events correction

Rare events, referred to also as imbalanced dataset or class imbalance in the statistical and data science literature, have been receiving increasing attention in the last two decades, especially since the resurgence of big-data.

When the proportion of ones (or successes) in a population is rare, when using a logistic regression to investigate its determinants its probability may be severely underestimated, and the coefficients biased towards zero. According to King and Zeng, 2001, this is the case especially that this could in fact happen when the proportion of ones over zeros is below 5%. The 'success' observations are the most informative, and yet in logistic regressions they have a smaller contribution to variance. The latter is in fact an inverse function of the odds-ratio $\pi_i(1-\pi)$, which in rare events tend to be larger, since the estimated probability in case of success π_i approaches 0.5. To address the rarity of the event under analysis, we compare different strategies among penalized estimation (Firth, 1993[31]), MLE estimation, and mixed-methods (King and Zeng, 2001; 2002)[32], contributing to the still relatively scarce existence of cross-discipline methodological comparisons including all the above. Through model comparison we select Firth methodology on the original sample, and maintain the methodology for the rest of the chapter, for two reasons.

31 In its seminal 1993 paper, Firth proposed a type of penalized likelihood regression which, by imposing a 'Jeffreys prior' on model coefficients, such that $A(\beta) = \frac{1}{2} logdet(I(\beta))$, to correct for small-sample and rare-event bias in Maximum Likelihood estimates.

32 The methodology proposed by (King and Zeng, 2001) starts from a random selection of one 'potential' pair for each 'observed' one, and progressively increase the proportion of zeros up to when no further efficiency gains in terms of standard errors size is obtained. The process of random selection of zeros and full sampling of ones is an endogenous stratified sampling method that introduces a bias in the logistic model. The bias can be easily solved by weighting the exogenous variables by the true successes proportion in the population (King and Zeng, 2001). The weighting factor take this form $w_i = diag[\hat{\pi}_i(1-\hat{\pi}_i)\omega_i]$, where ω_i is the true success proportion in the population.

Firstly, as shown in Table 1.8, the Firth-penalized logistic method applied on the full sample outperformed King and Zeng's mixed method in terms of minimisation of standard errors for almost all models disregarding the inclusion of controls and main covariates. This result is in line with a number of recent dedicated publications (Leitgob, 2020; Rainey, 2016; Bacaksiz and Koc, 2021; Puhr *et al.*, 2017) who found penalized logistic with Jeffrey's prior to be suitable for bias and SE minimization. However, it must be noted that, while Firth (1993)'s penalization is preferred for providing unbiased parameter estimates, it has been found to be biased towards zero in prediction (Puhr *et al.*, 2017; Elgmati *et al.*, 2015). As such, our marginal effects are likely to underestimated. Secondarily, given the results' high-comparability between traditional MLE on the full sample, the Firth-Penalized ML on the full sample and King and Zeng corrected logistic on the reduced samples, we opt to work with for the original dataset, avoiding data manipulation, increasing the reliability and replicability of our results.

Network distance

The network distance between each VC and each ISC is defined as the *number of steps needed to find an undirected link between them.* Undirected links, which identify "relations that do not distinguish between senders and receivers, like alliance partners" (Yang *et al.*, 2017, p. 11), were calculated in terms of inward (shareholders) and outward-holdings of firms, funds and of their managers and advisors. Previous-to-the-investment professional positions of managers and advisors of all firms involved in the network were also added to the same symmetric matrix, and treated equally and jointly to obtain our minimum-number-of-steps variable for all dyads. To clarify how the matrix is constructed and exploited to measure relational proximity, we propose an example: imagine that an ISC has among its shareholders an individual investor, called Mr. Smith. But Mr. Smith is also an advisor in a VC. In this case, the ISC will have a link with Mr. Smith, who in turn will have a link with the VC, such that the ISC and the VC will be at one step of distance in the contiguity matrix. Relational proximity between the ISC and the VC, measured as the inverse of network distance, will thus be equal to $1^{-1} = 1$.

Imagine now the case in which Mr. Smith is still a shareholder of the ISC, but instead of being an advisor in the VC firm, he is an advisor in a company A, where a manager of the VC used to work (or of which the VC holds some shares). In this case, the network distance between the ISC and the VC will be equal to 2, and relational proximity $2^{-1} = .5$. This relational proximity measure thus borrows the basic methods of social network analysis to represent the process of referral and word-of-mouth that have been previously suggested to be crucial in risky-investments (Teten and Farmer, 2010).

Additional estimates and robustness checks

The results we have obtained about the role of proximities in driving the probability of a successful VC-ISC pair appear substantially robust to two important checks. The first concerns the possible presence of spatial autocorrelation in the phenomenon we are investigating, as suggested by the spatial inequalities in the distribution of startup firms and VC offices we have highlighted in Section 3.4. To check for its actual presence and effects, we have first referred to the Local Market Areas (LMAs hereinafter) of the Italian territory, identified by the national statistical office (Istat) as "sub-regional geographical areas where the bulk of the labour force lives and works, and where establishments can find the largest amount of the labour force necessary to occupy the offered jobs" (Istat, 2014, p.1). With respect to these LMAs, Table 1.10 shows that a significant Moran (1948)'s global index of spatial autocorrelation is revealed only by the number of ISCs. Accordingly, the ISCs located in those LMAs where the local Moran's index (for the number of ISCs) was found significant at 5% level (see Figure 1.4), have been excluded from the original sample before re-estimating the full model. The spatial auto-correlation analysis has been repeated with respect to Italian NUTS3 regions (i.e. Italian provinces). At this level of analysis, no variable appears significantly correlated at the global level, with the exception of the province of Milan, exhibiting significant local autocorrelation throughout the whole set of variables, and of those of Milan and Rome, revealing the same autocorrelation for the number of ISCs. On this basis, the full model has been re-estimated by excluding these two provinces, which are also the ones hosting the largest number of VC funds offices and whose exclusion thus allows us to control for outlier areas.

The model proves robust to the exclusion of ISCs located in both LMAs and NUTS3 regions where the number of ISCs reveals significantly spatially autocorrelated. In particular, when ISCs (Columns 4-5 of Table 1.10) or VCs (Column 6) located in such areas are excluded the marginal effects of both the measures of intangible proximity, *i.e.* industrial (at sub-sector level) and relational, more than double with respect to the same model estimated on the full sample (Model 7 of Table 1.2), suggesting that outside the most advantaged areas, intangible proximity dimensions play a stronger role.

The second robustness check that we perform concerns the presence of idiosyncratic VC investments and funds, as results might be affected by second and syndicated VC investments,[33] which the literature has shown to differ from first and solo ones in terms of signalling effects, information asymmetries, and of strategies to reduce the risks of investing in technologically or physically distant firms (Berchicci *et al.*, 2011, Catalini and Hui, 2018, Cumming and Dai, 2010). In our period of analysis (2014-2019), only 20 investments were done in syndication with another VC partner (all among Italian VCs), while three foreign VCs participated in 1 joint operation. As Table 1.10 shows, the removal of the 23 syndicated investment-pairs from the sample does not change the results substantially.

33 As is well-know, syndication is a common practice in VC investments, denoting the joint presence of multiple investors in providing the funding needed by one company.

Additional tables and figures

Variable	Obs	Mean	Std.Dev.	Min	Max
ISC characteristics					
Age at finance	8480	2.79	1.41	0.00	5.44
N. of managers	8480	4.08	3.18	0.00	15.00
Female managers, share	8278	0.15	0.25	0.00	1.00
N. of employees	6110	3.76	4.67	0.00	28.00
Investment type					
Dummy: ISC had prior VC investment	8480	0.03	0.18	0.00	1.00
Dummy: syndicated investment	8480	0.02	0.15	0.00	1.00
Innovativity requirements (Law 221/2012)					
R&D>15% of prod.costs	8480	0.67	0.47	0.00	1.00
Patents ownership	8480	0.20	0.40	0.00	1.00
2/3 MA degree or 1/3 PhD holders	8480	0.22	0.41	0.00	1.00
Profitability					
Net worth (TEUR)	6113	517.02	1511.90	-5.52	11261.49
Revenues (TEUR)	6183	246.94	540.86	0.00	3299.24
Net profit (TEUR)	6183	-210.62	500.45	-3666.00	205.10
EBITDA (TEUR)	6113	-147.59	340.88	-2692.20	342.55
Location specific					
Population at legal office seat, th., 2012	8480	2316.87	1394.61	182.48	4321.24
Population density at legal office seat, sqkm, 2012	8480	1011.53	821.61	49.50	2622.00
GDP, province, (Million EUR), 2012	8480	91.28	64.36	4.27	156.00
EPO patent appl., per million inhabs, 2012	8480	72.16	43.81	2.72	201.06
N. of active high growth firms, per 4digit NACE, 201	8480	638.10	1080.30	4.00	12791.00
N. of active firms, per 4digit NACE, 2012	8480	21443.80	33501.10	151.00	333068.00
N. of events organized in coworking spaces, 2012	8480	149.33	163.55	0.00	399.00

Table 1.3. ISC and location-specific control variables, descriptive statistics

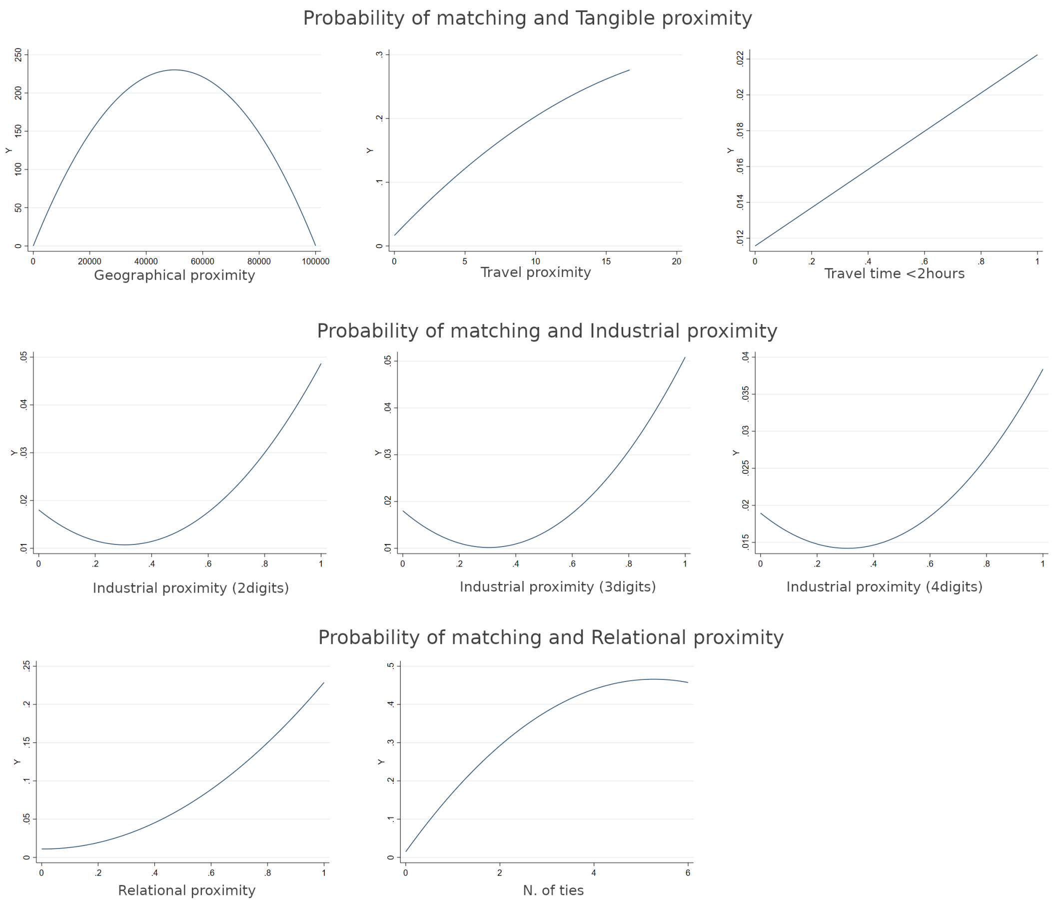
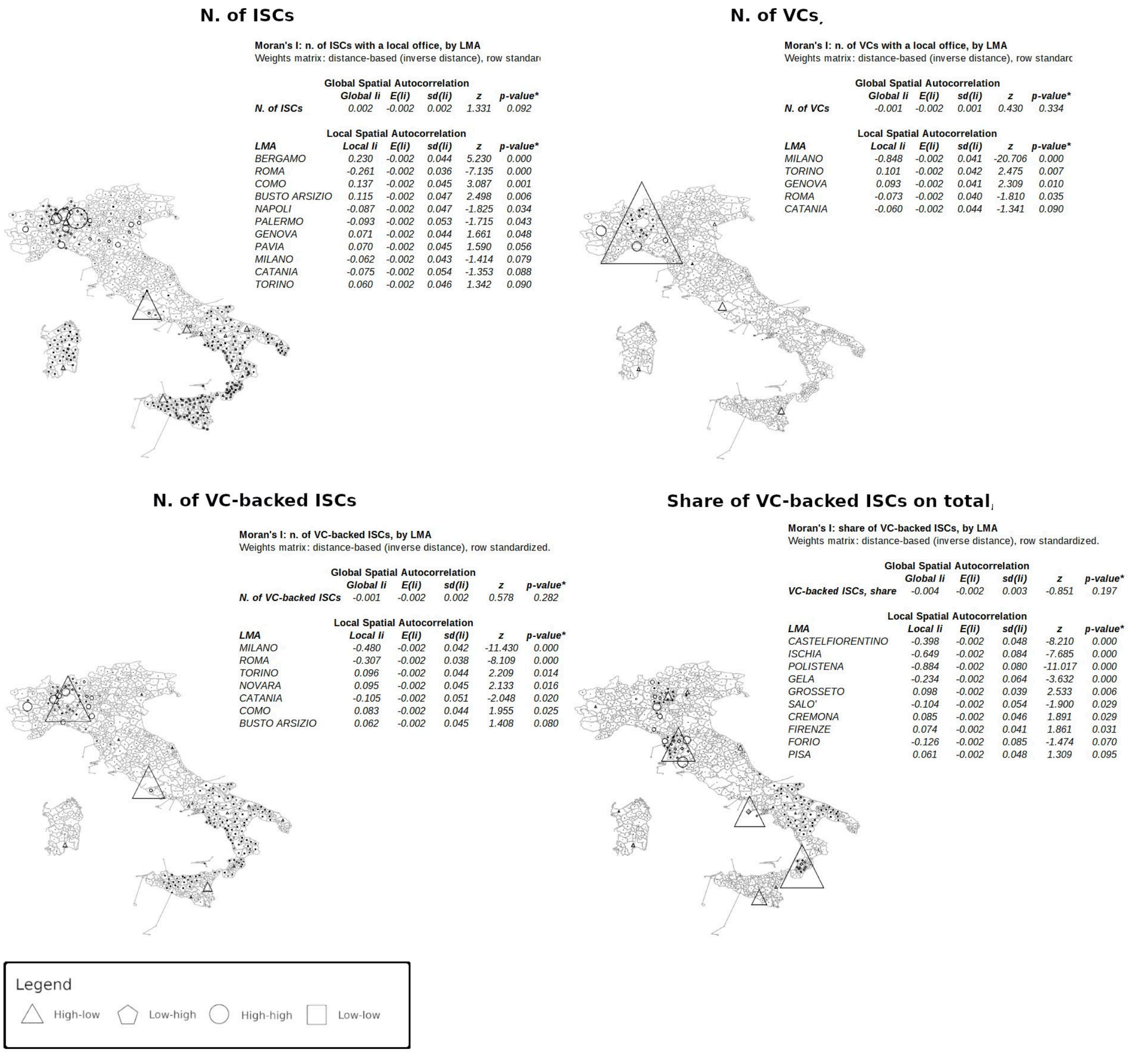


Figure 1.3. Quadratic prediction: dependent VS proximity variables



N. of ISCs — Global Spatial Autocorrelation

	Global Ii	E(Ii)	sd(Ii)	z	p-value*
N. of ISCs	0.002	-0.002	0.002	1.331	0.092

Local Spatial Autocorrelation

LMA	Local Ii	E(Ii)	sd(Ii)	z	p-value*
BERGAMO	0.230	-0.002	0.044	5.230	0.000
ROMA	-0.261	-0.002	0.036	-7.135	0.000
COMO	0.137	-0.002	0.045	3.087	0.001
BUSTO ARSIZIO	0.115	-0.002	0.047	2.498	0.006
NAPOLI	-0.087	-0.002	0.047	-1.825	0.034
PALERMO	-0.093	-0.002	0.053	-1.715	0.043
GENOVA	0.071	-0.002	0.044	1.661	0.048
PAVIA	0.070	-0.002	0.045	1.590	0.056
MILANO	-0.062	-0.002	0.043	-1.414	0.079
CATANIA	-0.075	-0.002	0.054	-1.353	0.088
TORINO	0.060	-0.002	0.046	1.342	0.090

N. of VCs — Global Spatial Autocorrelation

	Global Ii	E(Ii)	sd(Ii)	z	p-value*
N. of VCs	-0.001	-0.002	0.001	0.430	0.334

Local Spatial Autocorrelation

LMA	Local Ii	E(Ii)	sd(Ii)	z	p-value*
MILANO	-0.848	-0.002	0.041	-20.706	0.000
TORINO	0.101	-0.002	0.042	2.475	0.007
GENOVA	0.093	-0.002	0.041	2.309	0.010
ROMA	-0.073	-0.002	0.040	-1.810	0.035
CATANIA	-0.060	-0.002	0.044	-1.341	0.090

N. of VC-backed ISCs — Global Spatial Autocorrelation

	Global Ii	E(Ii)	sd(Ii)	z	p-value*
N. of VC-backed ISCs	-0.001	-0.002	0.002	0.578	0.282

Local Spatial Autocorrelation

LMA	Local Ii	E(Ii)	sd(Ii)	z	p-value*
MILANO	-0.480	-0.002	0.042	-11.430	0.000
ROMA	-0.307	-0.002	0.038	-8.109	0.000
TORINO	0.096	-0.002	0.044	2.209	0.014
NOVARA	0.095	-0.002	0.045	2.133	0.016
CATANIA	-0.105	-0.002	0.051	-2.048	0.020
COMO	0.083	-0.002	0.044	1.955	0.025
BUSTO ARSIZIO	0.062	-0.002	0.045	1.408	0.080

Share of VC-backed ISCs on total — Global Spatial Autocorrelation

	Global Ii	E(Ii)	sd(Ii)	z	p-value*
VC-backed ISCs, share	-0.004	-0.002	0.003	-0.851	0.197

Local Spatial Autocorrelation

LMA	Local Ii	E(Ii)	sd(Ii)	z	p-value*
CASTELFIORENTINO	-0.398	-0.002	0.048	-8.210	0.000
ISCHIA	-0.649	-0.002	0.084	-7.685	0.000
POLISTENA	-0.884	-0.002	0.080	-11.017	0.000
GELA	-0.234	-0.002	0.064	-3.632	0.000
GROSSETO	0.098	-0.002	0.039	2.533	0.006
SALO'	-0.104	-0.002	0.054	-1.900	0.029
CREMONA	0.085	-0.002	0.046	1.891	0.029
FIRENZE	0.074	-0.002	0.041	1.861	0.031
FORIO	-0.126	-0.002	0.085	-1.474	0.070
PISA	0.061	-0.002	0.048	1.309	0.095

Figure 1.4. Spatial Autocorrelation Analysis, at LMA level.

Data source	Variable	Sample	Obs	Mean	Std.Dev.	Min	Max
		VC characteristics					
Orbis, BvD	VC fund, age	Dyadic	8480	8.79	7.22	0.83	39.07
		Observed investments	160	8.96	6.59	0.83	39.07
		Last year observed	37	10.72	8.78	0.83	39.07
	N. of shareholders	Dyadic	8480	25.38	50.21	1	245
		Observed investments	160	25.83	39.19	1	245
		Last year observed	37	21.762	52.534	1	245
	N. of employees	Dyadic	7660	3.68	5.07	0	29
		Observed investments	141	4.31	4.77	0	29
		Last year observed	28	4.750	6.281	0	29
	N. of managers	Dyadic	8480	8.20	9.12	1	38
		Observed investments	160	14.49	14.38	1	38
		Last year observed	37	7.042	7.760	1	38
	N. of holdings	Dyadic	8480	30.88	23.75	4	77
		Observed investments	160	45.77	27.70	4	77
		Last year observed	37	22.14	19.23	4	77
	Total assets (Mio EUR)	Dyadic	7212	10.65	20.58	0.00	154.00
		Observed investments	133	12.62	16.26	0.00	154.00
		Last year observed	25	16.52	32.61	0.00	154.00
	Sale Revenues (Mio EUR)	Dyadic	7068	1.27	2.76	0.00	18.00
		Observed investments	131	1.57	2.60	0.00	18.00
		Last year observed	23	0.02	0.00	0.00	0.02
	Social Capital (Mio EUR)	Dyadic	7068	0.75	1.52	0.01	4.99
		Observed investments	131	0.66	1.42	0.01	4.99
		Last year observed	23	0.98	1.54	0.02	4.99
	Legal office in a metro area, dummy	Dyadic	8480	0.88	0.32	0	1
		Observed investments	160	0.931	0.254	0	1
		Last year observed	37	0.919	0.277	0	1
		VC investments by ISC sector					
Own elaboration on Orbis and AIDA (BvD) data	N. of investments, per NACE 4digits	Dyadic	8480	1.14	2.24	0	16
	N. of investments, per NACE 3digits	Dyadic	8480	1.50	2.94	0	25
	N. of investments, per NACE 2digits	Dyadic	8480	1.47	2.91	0	25
	Sh. of investments, per NACE 4digits	Dyadic	8480	4.54%	12.44%	0.00%	100.00%
	Sh. of investments, per NACE 3digits	Dyadic	8480	7.09%	14.67%	0.00%	100.00%
	Sh. of investments, per NACE 2digits	Dyadic	8480	7.36%	15.26%	0.00%	100.00%
		VC investments by ISC location					
Own elaboration on Orbis and AIDA (BvD) data	N. of investments, per city	Dyadic	8480	0.60	1.41	0	8
	N. of investments, per province	Dyadic	8480	0.63	1.28	0	8
	N. of investments, per region	Dyadic	8480	8.22	13.41	0	68
	Sh. of investments, per city	Dyadic	8480	5.94%	13.77%	0.00%	88.89%
	Sh. of investments, per province	Dyadic	8480	8.16%	18.96%	0.00%	100.00%
	Sh. of investments, per region	Dyadic	8480	75.82%	40.79%	0.00%	100.00%

Table 1.4. VC-specific descriptive statistics

Type	Variable	Y	Obs	Mean	Std.	Min	Max
Geographical	Geographical proximity	0	8320	336.72	5791.76	0.00	100000.00
		1	160	6875.77	25382.09	0.00	100000.00
Functional	Functional proximity	0	8320	0.11	1.06	0.00	16.67
		1	160	1.41	4.38	0.00	16.67
	Dummy: travel time<2hours	0	8320	0.68	0.47	0.00	1.00
		1	160	0.81	0.40	0.00	1.00
	Dummy: travel time<1/2hour	0	8320	0.19	0.39	0.00	1.00
		1	160	0.43	0.50	0.00	1.00
	Dummy: min. travel by car	0	8320	0.41	0.49	0.00	1.00
		1	160	0.61	0.49	0.00	1.00
Co-location	Dummy same city	0	8320	0.19	0.39	0.00	1.00
		1	160	0.45	0.50	0.00	1.00
	Dummy same province	0	8320	0.00	0.06	0.00	1.00
		1	160	0.01	0.08	0.00	1.00
	Dummy same region	0	8320	0.04	0.19	0.00	1.00
		1	160	0.09	0.28	0.00	1.00
	Dummy same area	0	8320	0.12	0.33	0.00	1.00
		1	160	0.08	0.27	0.00	1.00
Relational	Relational proximity	0	8320	0.17	0.09	0.00	1.00
		1	160	0.27	0.28	0.00	1.00
	N. of ties	0	8320	0.02	0.18	0.00	3.00
		1	160	0.35	0.77	0.00	6.00
Industrial	Industrial proximity (2digits)	0	8320	0.18	0.29	0.00	1.00
		1	160	0.26	0.39	0.00	1.00
	Industrial proximity (3digits)	0	8320	0.16	0.28	0.00	1.00
		1	160	0.25	0.39	0.00	1.00
	Industrial proximity (4digits)	0	8320	0.08	0.20	0.00	1.00
		1	160	0.09	0.24	0.00	1.00

Table 1.5. Proximity/distances descriptive statistics by value of the dependent variable.

Definition	Variable	Calculation	Specifics and data sources
Geographical Proximity: inverse of the length of the shortest path between two points, calculated along the ellipsoidal surface of the Earth.	*Geographical proximity*	Inverse of the minimum geodetic distance among all ISC and VC offices (Vincenty's 1975 equation)	Based on VC and ISC offices coordinates, at exact-address precision. **Source:** own calculation on *AIDA BvD* data.
Functional proximity: inverse measure of "the distance separating any two nodes such that it reflects the net effect of nodal properties upon their propensity to interact" (Brown and Horton, 1970, p.76)	*Functional proximity*	Inverse of the minimum travel-time in minutes among all ISC and VC offices	Inverse of travel times (by car, flight, train or ferry or a mix of them) at exact address precision. **Source:** own calculation on data from *hereAPI*, *rome2rio.com* and *openflights.org*
	Dummy: travel time<2hours	Dummy: minimum travel-time is within two hours	Travel time (by car, flight, train or ferry or a mix of them) is within two hours, at exact address precision. **Source:** own calculation on data from *hereAPI*, *rome2rio.com* and *openflights.org*
	Dummy: travel time<1/2hour	Dummy: minimum travel-time is within half an hour	Travel time (by car, flight, train or ferry or a mix of them) is within half an hour, at exact address precision. **Source:** own calculation on data from *hereAPI*, *rome2rio.com* and *openflights.org*
	Dummy: minimum travel time is by car	Dummy: minimum travel-time is by car	Minimum travel time is by car. **Source:** own calculation on webscraped data from *hereAPI*, *rome2rio.com* and *openflights.org*
Co-location: joint presence in the same territory of the ISC and VC offices	*Dummy: same city*	Mutually exclusive dummy variables, indicating if a VC branch is located in the same city, province or region of the ISC offices.	VC and ISC have at least one office in the same municipality. **Source:** AIDA BvD.
	Dummy same province		VC and ISC have at least one office, if not in the same municipality, in the same province. **Source:** AIDA BvD.
	Dummy same region		VC and ISC have at least one office, if not in the same municipality or province, in the same region. **Source:** AIDA BvD.
Relational proximity: a measure encompassing more than one non-tangible dimensions of proximity, such as social, cognitive and organizational ones (Moodysson and Jonsson, 2007)	*Relational proximity*	Inverse of the network distance (Number of steps to find a link among each VC and ISC)	Network distance: calculated considering undirected links among all holdings, investors and managers of ISCs and VCs. **Source:** own calculation on AIDA and ORBIS data.
	N. of ties	Number of ties existing among each VC and ISC	Number of ties: number of common names among managers, holdings, investors in any role in the two firms. **Source:** own calculation on AIDA and ORBIS data.
Industrial proximity: "Shared knowledge base, needed in order to communicate, understand, absorb and process new information successfully" (Boschma, 2005, p. 64)	*Industrial proximity (2 digits)*	Share of VC previous investment in the same ISC division (2digit NAICS code)	Per each VC investment: ratio among the number of investments in any type of firm within a specific NAICS code, on the total number of firms funded from 01/01/2012 up to the date of the investment under analysis. **Source:** own calculation on AIDA and ORBIS BvD data.
	Industrial proximity (3 digits)	Share of VC previous investment in the same ISC group (3digit NAICS code)	
	Industrial proximity (4 digits)	Share of VC previous investment in the same ISC industry (4digit NAICS code)	

Table 1.6. Proximity variables: definitions, calculation and sources

Area	NUTS3 Data		NUTS2 Data									
	Province: NUTS3 (capital regions only)	LQ NUTS3	Region: NUTS2	LQ NUTS2	N. of ISCs		N. of VC offices		N. of VC-backed ISCs		N. of VC-ISC investments	
					N.	Share	N.	Share	N.	Share	N.	Share
	Milano	**5.29**	Lombardia	**2.64**	2587	**25.3%**	17	**58.6%**	57	**41.9%**	69	**43.1%**
	Torino	**1.87**	Piemonte	**1.12**	541	5.3%	2	6.9%	11	8.1%	12	7.5%
	Genova	0.56	Liguria	0.57	191	1.9%	1	3.4%	2	1.5%	3	1.9%
	Aosta	0.00	Valle D'Aosta	0.00	21	0.2%	0	0.0%	0	0.0%	0	0.0%
NorthWest		***1.93***		*1.09*	*3340*	***32.7%***	*20*	***69.0%***	***70***	***51.5%***	***84***	***52.5%***
	Trento	0.96	Trentino	0.45	265	2.6%	0	0.0%	1	0.7%	1	0.6%
	Bologna	0.47	Emilia	0.50	905	8.9%	1	3.4%	5	3.7%	7	4.4%
	Venezia	0.00	Veneto	0.49	866	8.5%	1	3.4%	5	3.7%	1	0.6%
	Trieste	0.00	Friuli	0.48	220	2.2%	0	0.0%	1	0.7%	7	4.4%
NorthEast		*0.36*		*0.48*	*2256*	***22.1%***	*2*	*6.9%*	*12*	*8.8%*	*16*	*10.0%*
	Firenze	**2.40**	Toscana	1.04	438	4.3%	2	6.9%	10	7.4%	13	8.1%
	Roma	**1.90**	Lazio	**1.44**	1139	11.2%	3	10.3%	22	16.2%	25	15.6%
	Ancona	0.00	Marche	0.28	370	3.6%	0	0.0%	1	0.7%	1	0.6%
	Perugia	0.00	Umbria	0.00	193	1.9%	0	0.0%	0	0.0%	0	0.0%
Center		*1.07*		*0.69*	*2140*	*21.0%*	*5*	*17.2%*	***33***	***24.3%***	***39***	***24.4%***
	Cagliari	**2.74**	Sardegna	**1.42**	149	1.5%	1	3.4%	5	3.7%	5	3.1%
	Palermo	0.00	Sicilia	0.59	497	4.9%	1	3.4%	6	4.4%	6	3.8%
Islands		***1.37***		*1.00*	*646*	*6.3%*	*2*	*6.9%*	*11*	*8.1%*	*11*	*6.9%*
	Potenza	**1.32**	Basilicata	0.83	110	1.1%	0	0.0%	1	0.7%	1	0.6%
	Napoli	0.43	Campania	0.43	804	7.9%	0	0.0%	6	4.4%	6	3.8%
	Bari	0.30	Puglia	0.22	401	3.9%	0	0.0%	2	1.5%	2	1.3%
	L'Aquila	0.00	Abruzzo	0.00	218	2.1%	0	0.0%	0	0.0%	0	0.0%
	Catanzaro	0.00	Calabria	0.25	224	2.2%	0	0.0%	1	0.7%	1	0.6%
	Campobasso	0.00	Molise	0.00	74	0.7%	0	0.0%	0	0.0%	0	0.0%
South		*0.34*		*0.29*	*1831*	*17.9%*	*0*	*0.0%*	*10*	*7.4%*	*10*	*6.3%*
Italy, totals		*0.50*		*0.64*	*10213*		*29*		*136*		*160*	

Table 1.7. ISCs, VC, and investments geographical distribution and location quotient.

Model with functional proximity, without controls.

	Sample	Full Sample	Full Sample	Sample 1:7	Sample 1:6	Sample 1:5	Sample 1:4	Sample 1:3	Sample 1:2	Sample 1:1
	Method	Firth ML	logit	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng
Functional proximity										
Dummy: Travel	Coeff.	0.653***	0.665***	0.605***	0.607***	0.632***	0.663***	0.627***	0.613***	0.602**
time <2hours	St.Err.	0.200	0.201	0.21	0.212	0.214	0.217	0.222	0.233	0.262
Constant	Coeff.	-4.432***	-4.448***	-4.398***	-4.400***	-4.417***	-4.439***	-4.415***	-4.406***	-4.401***
	St.Err.	0.179	0.181	0.187	0.189	0.19	0.192	0.197	0.205	0.228
N		8,480	8,480	1,280	1,120	960	800	640	480	320
Marginal effects covariates at median values										
Dummy: Travel time<2hours =1		0.0149	0.0151	0.0136	0.0137	0.0143	0.0152	0.0142	0.0138	0.0135

Model with functional proximities and controls.

	Sample	Full Sample	Full Sample	Sample 1:7	Sample 1:6	Sample 1:5	Sample 1:4	Sample 1:3	Sample 1:2	Sample 1:1
	Method	Firth ML	logit	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng
Functional proximity										
Dummy: Travel	Coeff.	1.010***	1.020***	0.924***	0.886***	0.901***	0.952***	0.816***	0.823***	0.603*
time <2hours	St.Err.	0.24	0.242	0.26	0.261	0.264	0.267	0.276	0.297	0.336
Controls										
Dummy: ISC had prior	Coeff.	1.007***	0.977***	0.897**	0.843**	0.742**	0.835**	0.811**	0.679	0.434
VC investment	St.Err.	0.31	0.316	0.365	0.368	0.372	0.387	0.412	0.443	0.45
Dummy: syndicated	Coeff.	2.511***	2.510***	2.555***	2.427***	2.453***	2.510***	2.641***	2.571***	2.476***
investment	St.Err.	0.232	0.234	0.306	0.309	0.324	0.346	0.398	0.459	0.625
ISC's age at finance	Coeff.	-0.146**	-0.147**	-0.160**	-0.163**	-0.157**	-0.166**	-0.179***	-0.177**	-0.149*
	St.Err.	0.0605	0.0607	0.0661	0.0646	0.0644	0.0657	0.0678	0.0705	0.0805
L1. GDP, at ISCs	Coeff.	-4.68e-06***	-4.62e-06***	-5.20e-06***	-5.39e-06***	-5.95e-06***	-6.45e-06***	-5.40e-06***	-6.51e-06***	-6.58e-06**
	St.Err.	1.72E-06	1.73E-06	1.96E-06	1.97E-06	1.98E-06	2.04E-06	2.1E-06	2.25E-06	2.67E-06
Constant	Coeff.	-4.141***	-4.173***	-3.895***	-3.799***	-3.744***	-3.725***	-3.697***	-3.532***	-3.376***
	St.Err.	0.298	0.3	0.318	0.324	0.331	0.333	0.349	0.38	0.451
Area FE		yes	yes	yes	yes	yes	yes	yes	yes	yes
N		8,473	8,473	1,278	1,118	958	798	638	478	318
Marginal effects (covariates at median values)										
Dummy: Travel time<2hours =1		0.0146	0.0145	0.0139	0.0208	0.0174	0.0237	0.0189	0.0123	0.0119

Model with functional, industrial and relational proximities and controls.

	Sample	Full Sample	Full Sample	Sample 1:7	Sample 1:6	Sample 1:5	Sample 1:4	Sample 1:3	Sample 1:2	Sample 1:1
	Method	Firth ML	logit	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng	KingZeng
Functional proximity										
Dummy: Travel	Coeff.	1.036***	1.051***	0.994***	0.922***	0.947***	0.988***	0.825***	0.827***	0.676*
time <2hours	St.Err.	0.248	0.25	0.276	0.277	0.277	0.284	0.292	0.317	0.36
Industrial proximity										
Industrial proximity	Coeff.	0.498*	0.490*	0.691**	0.772**	0.831***	0.852***	0.878***	0.898**	1.026**
(3digits)	St.Err.	0.256	0.257	0.315	0.313	0.32	0.324	0.335	0.362	0.406
Relational proximity										
Relational proximity	Coeff.	3.535***	3.548***	3.817***	3.953***	3.763***	3.750***	3.405***	3.481***	3.828***
	St.Err.	0.364	0.368	0.596	0.654	0.666	0.708	0.702	0.821	1.128
Controls										
Dummy: ISC had prior	Coeff.	1.116***	1.090***	0.970**	0.924**	0.829**	0.946**	0.937**	0.773	0.657
VC investment	St.Err.	0.315	0.321	0.391	0.401	0.408	0.434	0.468	0.517	0.559
Dummy: syndicated	Coeff.	2.595***	2.599***	2.748***	2.610***	2.642***	2.683***	2.810***	2.721***	2.636***
investment	St.Err.	0.238	0.24	0.313	0.316	0.331	0.357	0.411	0.476	0.658
ISC's age at finance	Coeff.	-0.156**	-0.157**	-0.192***	-0.190***	-0.182***	-0.187***	-0.202***	-0.200***	-0.160*
	St.Err.	0.0621	0.0624	0.0704	0.0693	0.069	0.0701	0.0721	0.074	0.0818
L1. GDP, at ISCs	Coeff.	-5.67e-06***	-5.64e-06***	-6.18e-06***	-6.31e-06***	-6.87e-06***	-7.14e-06***	-5.93e-06***	-6.84e-06***	-6.94e-06**
NUTS3 (MEUR)	St.Err.	1.75E-06	1.76E-06	2.03E-06	2.04E-06	2.06E-06	2.13E-06	2.19E-06	2.35E-06	2.86E-06
Constant	Coeff.	-4.837***	-4.876***	-4.680***	-4.607***	-4.542***	-4.539***	-4.447***	-4.308***	-4.314***
	St.Err.	3.13E-01	3.16E-01	3.55E-01	3.65E-01	3.75E-01	3.79E-01	3.90E-01	4.36E-01	4.84E-01
Area FE		yes	yes	yes	yes	yes	yes	yes	yes	yes
N		8,473	8,473	1,278	1,118	958	798	638	478	318
Marginal effects (covariates at median values)										
Dummy: Travel time<2hours =1		0.0116	0.0115	0.0109	0.0169	0.0136	0.0189	0.0145	0.00912	0.00989
Industrial proximity at 50%		0.00714	0.00687	0.0107	0.0208	0.0181	0.0250	0.0239	0.0155	0.0251
Relational proximity at 2steps-distance		0.128	0.127	0.150	0.270	0.190	0.251	0.186	0.123	0.201

Table 1.8. Comparison of sampling and estimation methods for rare events bias and correction. Models have been ordered from smaller (left) to larger (right) root mean squared error (RMSE).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Dependent variable																		
(1) Y=VC-ISC deal, dummy	1.000																	
Spatial proximity																		
(2) Geographical proximity	0.131	1.000																
geodetic	0.000																	
Functional proximity																		
(3) Travel proximity	0.145	0.920	1.000															
	0.000	0.000																
(4) Dummy	0.036	0.046	0.074	1.000														
Min. travel by car	0.001	0.000	0.000															
(5) Dummy	0.049	0.053	0.079	-0.266	1.000													
Min. travel<2hours	0.000	0.000	0.000	0.000														
Co-location																		
(6) Co-location	0.089	0.138	0.209	0.329	0.285	1.000												
same city	0.000	0.000	0.000	0.000	0.000													
(7) Co-location	0.006	-0.004	-0.005	0.041	0.047	-0.030	1.000											
same province	0.583	0.705	0.647	0.000	0.000	0.006												
(8) Co-location	0.035	-0.014	-0.019	0.100	0.156	-0.099	-0.012	1.000										
same region	0.001	0.211	0.074	0.000	0.000	0.000	0.265											
(9) Co-location	-0.017	-0.025	-0.037	0.092	-0.027	-0.183	-0.022	-0.074	1.000									
same area	0.118	0.020	0.001	0.000	0.015	0.000	0.038	0.000										
Industrial proximity																		
(10) Industrial proximity	0.041	0.042	0.026	0.007	-0.004	0.003	0.047	-0.023	0.101	1.000								
2 digits	0.000	0.000	0.016	0.495	0.707	0.748	0.000	0.032	0.000									
(11) Industrial proximity	0.043	0.047	0.032	0.019	0.004	0.024	0.007	-0.018	0.095	0.946	1.000							
3 digits	0.000	0.000	0.003	0.076	0.694	0.025	0.535	0.105	0.000	0.000								
(12) Industrial proximity	0.012	-0.009	-0.012	0.023	0.002	0.022	0.036	-0.012	0.063	0.571	0.605	1.000						
4 digits	0.267	0.424	0.265	0.038	0.856	0.043	0.001	0.253	0.000	0.000	0.000							
Relational proximity																		
(13) Relational proximity	0.150	0.245	0.246	0.011	0.025	0.083	-0.001	-0.003	-0.017	0.137	0.143	-0.033	1.000					
	0.000	0.000	0.000	0.310	0.023	0.000	0.902	0.758	0.110	0.000	0.000	0.002						
(14) N. of ties	0.209	0.261	0.344	0.035	0.039	0.068	-0.008	0.051	-0.034	0.130	0.141	-0.002	0.603	1.000				
	0.000	0.000	0.000	0.001	0.000	0.000	0.458	0.000	0.002	0.000	0.000	0.835	0.000					
Controls																		
(15) ISC prior VC investment	0.037	-0.003	0.002	0.012	-0.003	0.008	-0.011	-0.034	0.001	-0.056	-0.050	-0.041	-0.010	-0.012	1.000			
Dummy	0.001	0.788	0.853	0.273	0.796	0.487	0.300	0.002	0.898	0.000	0.000	0.000	0.365	0.252				
(16) Syndicated investment	0.142	0.023	0.020	0.014	0.008	0.021	-0.010	0.001	0.015	-0.009	-0.007	0.006	0.009	0.015	0.001	1.000		
Dummy	0.000	0.031	0.071	0.197	0.458	0.049	0.381	0.954	0.173	0.408	0.526	0.565	0.427	0.163	0.940			
(17) ISC's age at finance	-0.015	-0.012	-0.037	0.019	-0.051	-0.049	-0.027	-0.020	-0.023	0.025	0.011	0.027	-0.001	-0.014	-0.010	0.069	1.000	
	0.170	0.274	0.001	0.083	0.000	0.000	0.013	0.065	0.035	0.021	0.297	0.013	0.892	0.199	0.363	0.000		
(18) L1. GDP, at ISCs NUTS3	0.001	0.002	0.042	0.640	-0.337	0.448	0.000	-0.203	-0.015	-0.022	-0.002	0.008	0.001	-0.007	0.019	0.028	-0.015	1.000
	0.891	0.888	0.000	0.000	0.000	0.000	0.970	0.000	0.160	0.047	0.823	0.446	0.898	0.537	0.076	0.009	0.179	

Table 1.9. Pairwise correlation, dependent and proximity variables

Figure 1.5. Spatial Autocorrelation Analysis, at NUTS3 level.

	No Syndication	No Second Investments	No Second-No Syndication	No ISCs located in Milan (NUTS3)	No Spatially Autocorr. SLL, n.ISCs	No Spatially Autocorr. SLL, n.VCs	No Foreign VCs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Functional proximity							
Dummy: travel time<2hours	1.334***	1.329***	1.209***	1.375***	1.496***	1.350***	1.494***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.000)
Industrial proximity							
Industrial proximity	2.188***	2.149***	2.298***	2.312***	2.337***	2.322***	2.346***
3 digits	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Relational proximity							
Relational proximity	4.530***	4.247***	4.382***	4.509***	4.517***	4.496***	4.288***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Interactions							
Industrial proximity 3digits * Dummy travel	-1.326**	-1.275**	-1.124**	-1.201**	-1.229	-1.199**	-1.617***
time<2hours	(0.017)	(0.021)	(0.046)	(0.040)	(0.107)	(0.041)	(0.003)
Industrial proximity 3digits * Relational	-1.616**	-1.797**	-2.080***	-2.307**	-2.138*	-2.307**	-1.533**
Proximity	(0.034)	(0.020)	(0.008)	(0.027)	(0.071)	(0.027)	(0.047)
Relational proximity * Dummy travel	-0.066	0.514	0.447	0.556	0.325	0.571	0.168
time<2hours	(0.936)	(0.552)	(0.615)	(0.590)	(0.802)	(0.580)	(0.835)
Controls							
Dummy: ISC had prior VC investment	1.079***	1.140**		0.962**	1.313**	0.961**	1.064***
	(0.002)	(0.011)		(0.025)	(0.039)	(0.026)	(0.002)
Dummy: syndicated investment		2.560***		2.582***	2.806***	2.583***	2.554***
		(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
ISC's age at finance	-0.175***	-0.150**	-0.182***	-0.137*	-0.092	-0.142*	-0.153**
	(0.008)	(0.019)	(0.007)	(0.062)	(0.360)	(0.060)	(0.017)
L1. GDP, at ISCs NUTS3 (MEUR)	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000***
	(0.007)	(0.000)	(0.002)	(0.004)	(0.143)	(0.005)	(0.001)
Area FE	yes	yes	yes	yes	yes	yes	yes
Constant	-5.313***	-5.283***	-5.202***	-5.506***	-5.263***	-5.471***	-5.399***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	8,269	7,673	7,509	5,612	2,926	5,491	7,700
Proportion of ones	1.58%	1.88%	1.60%	1.84%	1.91%	1.31%	1.95%
Estimated probability. Robust pval in parenthesis (*** p<0.01, ** p<0.05, * p<0.1)							
Marginal effects (covariates at median values)							
Dummy: Travel time <2hours =1	0.0227	0.0239	0.0196	0.0266	0.0489	0.0260	0.0275
Industrial proximity at 33%	0.0768	0.0786	0.0794	0.0958	0.0370	0.0964	0.0937
Industrial proximity at 50%	0.111	0.113	0.117	0.142	0.0550	0.143	0.140
Industrial proximity at 75%	0.173	0.174	0.186	0.225	0.0878	0.228	0.223
Relational proximity at 3 steps-distance	0.162	0.153	0.145	0.182	0.0692	0.181	0.159
Relational proximity at 2steps-distance	0.350	0.315	0.306	0.392	0.149	0.388	0.330
Relational proximity at 1step-distance	3.369	2.631	2.736	3.733	1.427	3.675	2.812

Table 1.10. Robustness Checks: exclusion of syndicated and second investments, of spatially autocorrelated Local Labour Markets by number of ISCs or per number of VCs (as listed in figures Figure 1.4 and Figure 1.5), of firms located in Milan, and of foreign funds.

CHAPTER 2

Effects of remote work on population distribution across cities: a QSE application

2.1 Introduction

Remote work, also referred to as telecommuting, telework or work-from-home,[34] has been slowly but steadily growing in the last twenty years, both in the United States (Wulff and Vernon, 2020; BLS, 2019) and in the European Union (Welz and Wolf, 2010; JRC, 2020) and involving, in 2017, 36 millions of US workers (BLS, 2019). The burst of the Covid-19 pandemic, and the associated need of firms to resort to remote work, has made the phenomenon even more diffused (Barrero *et al.*, 2021; Bick *et al.*, 2021). Moreover, it stimulated a revamp of economic research on it (Kosteas *et al.*, 2022; Aizhan *et al.*, 2022), and on the potential of a rural revival (Gonzalez-Leonardo *et al.*, 2022). There is diffused evidence of outflows from denser cities during the pandemic years (Couture *et al.*, 2021; Whitaker, 2021), but while most of these flows were directed towards suburban areas surrounding dense cities (Ramani and Bloom, 2021) here are exceptions of countries such as Spain or Sweden that showed a slight increase of in-migration towards (and/or a decrease in outmigration from)

34 A standard definition for it is that of "any contractual arrangement allowing to work from home at least occasionally while being paid for it" (BLS, 2019)

rural areas (Rowe *et al.*, 2023; Vogiazides and Kawalerowicz, 2023). Aside from the direction of relocation, and more importantly, its effective threat on urban hierarchy (which, as reported by Gonzalez-Leonardo *et al.*, 2022, have been by far discarded as negligible), one focal matter that is yet to be ascertained is the stickiness that Covid-19 (and remote-work) induced flows will reveal in the post-pandemic years (Glaeser, 2022; Florida *et al.*, 2021; Bick *et al.*, 2021; Althoff *et al.*, 2022).

The post-Covid19 stream of studies adds to more than five decades of wide academic research, policy discussion and initiatives, from which remote-work has generally been considered as a desirable way to decrease cities' congestion, their infrastructural burden, and their level of pollution, as well as to improve their preparedness to disasters (Zhu and Mason, 2014; Donnelly and Proctor-Thomson, 2015). However, this has been accompanied by inconclusive results of the research on the effects of remote work on residential relocation choices and commuting behaviors (de Abreu e Silva and Melo, 2018; Zhu, 2012; Choo *et al.*, 2005; Gubins *et al.*, 2019; Kim *at el.*, 2015). This is of course quite puzzling and, in the extant literature, has been so far mainly attributed to quite standard econometric problems, like the double causality in the relationship between teleworking and residential/commuting decisions, the cross-sectional nature of available data, and the heterogeneity of teleworking arrangements. The studies addressing these issues through general equilibrium models focused mainly on single monocentric-city frameworks and on the impact of remote-work adoption on the urban structure (Behrens *et al.*, 2021; Davis *et al.*, 2021; Monte *et al.*, 2023; Davis *et al.*, 2021). Despite some evidence of across-cities relocation before (Choudhury *et al.*, 2019) and during (Rowe *et al.*, 2023; Monte *et al.*, 2023) Covid-19, between-cities movements of people induced by teleworking remain largely unexplored.

Furthermore, the pandemic and the research stream it prompted showed that the possibility to work remotely is unevenly distributed with respect to skills and occupations. The ability to work remotely has been established to depend on the degree of self-monitoring, on the type of technology involved in daily tasks such as the use of laptop rather than in-site machineries, and on the need of face-to-face activities with clients or colleagues (JRC-Eurofound, 2020; McKinsey, 2021; Dingel and Neiman, 2020). While its impact on inequality and on the urban structure have recently been object of enquiry (Monte *et al.*, 2023) the implications of such differential access to remote-work on have yet to be analysed in a system of cities' framework.

This chapter tries to overcome all of these challenges by exploiting an invertible spatial general equilibrium model on the basis of which it provides estimates of the impact of possible telework adoption on the city's attractiveness for firms and workers. To the best of this author's knowledge, this is the first study to analyse workers relocation as induced by remote-work adoption in a general equilibrium model with multiple cities.

Exploiting the large data availability and the peculiarity of the country in terms of low linguistic and institutional barrier to labour mobility, the model is calibrated with respect to US data, Cities are defined at the Metropolitan Statistical Area level (and excluding Micropolitan Areas), identifying "core areas containing a substantial population nucleus, together with adjacent communities having a high degree of economic and social integration with that core". Each metropolitan statistical area must have at least one urbanized area of 50,000 or more inhabitants.[35]. This defini-

35 The definition of Metropolitan Statistical Areas and the standard for their delineation can be found at https://www.census.gov/programs-surveys/metro-micro/about.html

tion of cities at MSA level is necessary in order to collect all the data required for the model's quantification. However, this also implies that the workers relocation between the core of the metropolitan area and its suburbs will not be captured. Furthermore, as the pandemic crisis is only recently being considered as resolved, neither the persistence nor the sector-specificity of its effects (affected by governments temporary legislation) can yet be assumed from available data. Therefore, the model exploits pre-pandemic data for 2017, as the most suitable year in terms of data coverage, and infers potential future trends through counterfactual levels of local telework adoption. In methodological terms the proposed empirical strategy, based on a counterfactual exercise, makes use of the Quantitative Spatial Economic model developed by Behrens *et al.* (2017). Given their occupational composition, I first estimate each city's potential for remote-work's adoption through the share of employment in teleworkable occupations (Dingel and Neiman, 2020), and infer the average commuting costs variations per each MSA that would derive from such adoption. Then, by exploiting Behrens *et al.* (2017) model and its invertibility property, I am able to quantify the expected counterfactual changes in firms and workers' distribution across cities as prompted by such levels of remote-work adoption (and decreases in commuting costs). The adopted methodology has a number of advantages in addressing the previously acknowledged challenges. Firstly, the invertibility property is crucial to solve the double causality issue in absence of rich panel data or of a suitable natural experiment, allowing for counterfactual exercises. Secondly, to address the discrepancy of previous results due to the heterogeneity of existing teleworking contracts, this study makes use of a unifying definition of regular remote-workers, according to which it is not required to commute to the office more than once per week. In particular, it overcomes the lack of available data on this form of remote-work by producing an upper bound estimates of the number of potential teleworkers per each city, *i.e.* employed in occupations constituted by at least 80% of activities feasible to be performed remotely.[36]

The results of the counterfactual exercise suggest that larger adoption of remote work would reinforce agglomeration in larger cities. Since remote workers exhibit lower commuting-frequency, cities with larger shares of employment in teleworkable occupations will face larger drops in urban frictions, becoming more attractive for firms through savings. The model predicts that cities with larger shares of potential teleworkers will enjoy larger firms attraction and a subsequent rise in wages in the short term (before population adjustments). In the long-term (once workers relocate) firms in larger cities will experience stronger selection and lower average markups population gains as pulled by higher wages and by consumption amenities in the form of more varieties in the consumption good at a lower local average price. Given that larger cities tend to host greater shares of employment in remote-workable occupations, remote work adoption adds up to other agglomeration forces typically entailed in urban size as the well-documented productivity and wage premium (**?**). While welfare gains will also be enjoyed in smaller cities, enabled by the pro-competitive effect of trade which will reduce markups in all cities, all in all, the analysis confutes the idea, also revamped in the aftermath of Covid-19, that remote work will benefit smaller cities in terms of workers attraction.

The rest of the article is organised as follows. Section 2.2 reviews the relevant existing literature, Section 2.3 presents the modeling framework (2.3.1), the counter-

36 A different category of remote work, which could be defined as partial telework, would comprise all workers that could spend at least one full day per week working from home. However, given the greater gap in the analysis of fully-remote work, and that 95% of occupations are found by Dingel and Neiman (2020) to either have null or full shares of activities that could be performed at home, this work will mainly focus on the above defined category.

factual estimation (2.3.2), the data and the quantification method (2.3.3). Section 2.4 will present and discuss the results, and Section 2.5 will conclude.

2.2 Literature review

In the aftermath of the adoption of telecommuting, total-travelled-miles should be expected to shrink, at least in absence of residential relocation. However, the relative empirical evidence shows that the effect of remote-work on individual average travelled miles is ambiguous, in some cases found as positive (Melo and de Abreu e Silva, 2017; Zhu, 2012; de Vos *et al.*, 2018; Kim *at el.*, 2015), and in others negative, but small (Choo *et al.*, 2005), or not significant at all (Gubins *et al.*, 2019).

These discrepancies in ascertaining the effects of remote work are allegedly due to some important empirical problems, of which the extant literature is affected. The most important one is represented by an issue of double-causality, which can be deemed to insist on the relationship between telecommuting and relocation decisions (Ory and Mokhtarian, 2006; Melo and de Abreu e Silva, 2017; de Vos *et al.*, 2018; Muhammad *et al.*, 2007; Choo *et al.*, 2005; Gubins *et al.*, 2019; Ellen and Hempstead, 2002). While telecommuting adoption could spur a residential relocation decision, the distance between the place of residence and the workplace can in turn influence the decision to telecommute. Distinguishing the direction of causality is thus challenging, especially given the cross-sectional and limited nature of existing work-from-home data.

Previous studies on relocation choices have tried to solve this issue by making use of either panel surveys or general equilibrium frameworks. Among the former kind of studies, Choudhury *et al.* (2019) used a 2005-2017 yearly survey to analyse the relocation of US-patent officers across cities of different sizes following the adoption of remote-working contracts. The authors found that the majority of remote workers chose smaller cities (with lower living costs), except early-stage ones who chose to relocate into larger cities. Ory and Mokhtarian (2006) exploited a 10-year retrospective survey to analyse the relocation decisions of teleworkers, by explicitly considering their distance from the workplace. Their main finding is that workers who moved after a telecommuting contract usually sought to locate closer to their office, whereas telecommuting decisions following a moving choice tended to be prompted by a farther-away relocation. These results are partially consistent with those of Nilles (1991), who found that workers living at an above-average distance from their working place were more likely to opt for a teleworking contract. Conversely, in the same study, the relocation choice that followed a telework arrangement is reported to typically happen towards the suburbs. Tayyaran *et al.* (2003) also found some evidence of relocation towards smaller cities or suburbs. However, this appeared linked to specific household characteristics (such as the presence school-aged children in the family), on the accessibility (transport costs) and on the availability of outdoor recreational amenities.

The role of amenities, including endogenous consumption ones (Rappaport, 2008), is in fact increasingly deemed as relevant in explaining residential location choices. Consumption amenities can vary significantly across neighborhoods and cities, explaining within and across-city sorting of workers (Almagro and Dominguez-Iino, 2021; Gaigné *et al.*, 2019; Couture *et al.*, 2019). The crucial role of amenities (Lund and Mokhtarian, 1994), together with that of transport and telecommunication costs (Ota, 2017), also emerges from the stream of studies which use general equilibrium modeling frameworks. Among these studies, Lund and Mokhtarian (1994) found that within-city relocation depends on the position of the office and on available amenities. Ota (2017) suggest the positive correlation between telecommuting adoption and suburbanisation (as in, the relocation towards smaller cities surrounding core ones) to mainly depend on general obstacles to remote-work adoption. These obstacled (such as social systems customs, technological skills and equipment, loss of knowledge spillovers and reputational costs influencing future career development), are also found to be crucial in determining workers teleworking and relocation choices. Nevertheless, the pandemic outbreak, by imposing a sudden and large increase in telework adoption on a worldwide scale, is have significantly reduced them (Barrero *et al.*, 2021; Davis *et al.*, 2021).

Another relevant issue in the extant literature is represented by the heterogeneity of teleworkability across different types of workers. Both stylized facts and systematic data on the topic reveal that in the pre-pandemic period, telecommuters appeared to concentrate in the last quartile of skills, salaries and city sizes distributions. This unevenness was confirmed during Covid-19 times (JRC-Eurofound, 2020; McKinsey, 2021), and linked to the different degree of physical interactions and of autonomy entailed in the job, and to the different technologies used in daily tasks. The pattern at stake was also found by Dingel and Neiman (2020), who classified the feasibility of working from home for all the 1000 U.S. occupations in the O*NET database (by exploiting two surveys in the 24.2 release of the same database); and by Mongey *et al.* (2021), using a variant of the same methodology. Both these works found that teleworkability is highly correlated with income and skill levels. Despite the relevance of this evidence, skill-heterogeneity is hardly retained in the analysis of relocation patterns induced by teleworking. Among the few studies that do so, Ellen and Hempstead (2002) investigated the over-concentration of telecommuters in larger and more densely populated urban area, and found that telecommuters are more likely to live in urban areas than other white-collar workers. While insightful, the study is based on cross-sectional data and thus unable to claim for causality. In another recent article, Behrens *et al.* (2021) developed a model on the effects of telework on within-cities welfare and productivity changes, accounting for the access to remote jobs of skilled and unskilled workers. Their results suggest that allowing for a positive share of teleworkers is consistent with profit maximization at the firm-level; and that the relationship between teleworking and aggregate productivity exhibits a bell-shaped curve. However, as it is based on a single city model, their spatial framework does not allow them to analyse how the size and composition of cities will be affected by telecommuting. Similarly, Monte *et al.* (2023) investigate the effects on productivity, inequality and urban structure through a monocentric general equilbrium model accounting for differences in WFH productivity across occupations and skill level. However, as they aggregate US data as if these all pertained to one unique city, and thus cannot provide any city-specific insight.

A last relevant issue emerging from the literature on the topic, still possibly accounting for the above recalled discrepancies, is represented by the heterogeneity of remote-work arrangements in place (de Vos *et al.*, 2018; Stiles and Smart, 2020; Melo and de Abreu e Silva, 2017). While this is also a relevant issue, the extent to which it has been address is very limited. To the best of this author's knowledge, this has occurred only in the study by Asgari and Jin (2015), who provided a framework to transpose telecommuting behaviors on a daily-frequency level, distinguishing home based workers from the variety of arrangements with above zero weekly hours of work-at-home [37].

In the model quantification of the present article, this complexity is reduced by defining as (potential) remote-workers those employed in occupations with at least 80% of activities that can be performed at home. This definition is quite comprehensive of different types of remote workers, since 95% of all occupations at 6-digits codes are found by Dingel and Neiman (2020) to either have null or full shares of activities that could be performed at home.

2.3 Methodology

The research question of this work is addressed through the quantification of a Quantitative Spatial Economic model (QSE). The term QSE refers to a group of recently developed structural general equilibrium models that, by featuring multiple spatial frictions, localized amenities, and heterogeneous agents while being sufficiently tractable to perform realistic counterfactual exercises (Redding and Rossi-Hansberg, 2017), allow to quantify the effect of specific policies on the endogenous location of firms and workers and on the welfare and productivity entailed in it.

More precisely, this work draws on and extends the model of this kind developed by Behrens *et al.* (2017), which was originally developed to quantify the impact of a decrease of urban and trade frictions on the city-size distribution, the size of individual cities, and their contribution to productivity and competition across and within cities. Behrens *et al.* (2017)'s main results can be summarised as that, if commuting costs were to shrink, larger cities would gain population to the detriment

37 The latter are divided into *non-regular* telecommuters also defined as *potential* remote-workers, and *regular* telecommuters, in turn distinguished in terms of potential daily work-related trips into *primary* (no work-related trips when at home) and *ancillary* (work-related trips may occur when at home).

of smaller ones. Conversely, the horizontal suppression of trade costs would yield the opposite result, benefiting smaller cities. Since telecommuting can be modeled through a reduction of commuting costs, I expect larger remote-work adoption in cities to have similar qualitative effects to those obtained by Behrens *et al.* (2017).

Being the first model that retains pro-competitive effects across-cities enabled by heterogeneously productive firms and trade accounting for commuting costs, and in which production-, consumption- and natural-amenities all participate to the locational choices of population, Behrens *et al.* (2017)'s model is well-suited to the research question of this paper. However, it has some limitations too. First, it assumes identical workers for the sake of tractability, and as a result, it omits any considerations related to sorting and matching. Second, their counterfactual exercise estimates the effect of an equal total suppression of commuting and trade costs across all US cities, and is thus not suitable to detect the effects of differential remote-work adoption potential of cities linked to their specific skill-composition.

These are crucial aspects to retain for the research question at stake, which the new model quantification, as presented in the next sub-sections, will try to address.

2.3.1 The benchmark QSE model

This section, in conjunction with the Appendix in this chapter, aims to provide a summary and some intuitions on the functioning of the benchmark model of this work, as developed by Behrens *et al.* (2017). The model is based on a multiple monocentric-city structure, with endogenous workers' location choices, productivity, and markups. Firms are heterogeneous in productivity, expressed in terms of their marginal labour requirement $m_r(i) \geq 0$. This is drawn from a city-specific continuously differentiable distribution, G_r, and discovered only after making an irreversible entry decision in the city. After having incurred in a fixed entry cost F (expressed in units of labour paid at the market wage, taken as given), firms are selected: they will then produce a variety i of the differentiated good using only labour, provided they can charge prices equal or above their marginal costs in at least one city.

The urban structure is characterised by multiple monocentric cities, whose size is solely determined by the number of workers in city r (L_r): the radius $\bar{x}$ of the circular city r, is thus assumed to be equal to $\bar{x}_r = \sqrt{L_r/\pi}$. Land is used only for housing, while firms are assumed to be located in a dimensionless central business district (CBD). Workers and firms face urban and trade frictions. Urban frictions are represented by iceberg-type commuting costs, negatively affecting the individual city net-(of commuting times) labour supply s_r ($s_r = h_r = \bar{h}_r e^{-\theta_r} x_r$, where $\bar{h}$ represents the sum of the city-specific average weekly working hours h_r plus commuting times) and the aggregate net-labour supply S_r proportionally to the size of the city L_r, and to the parameter $\theta_r > 0$ for commuting technology, such that:

$$S_r = \int_0^{\bar{x}_r} 2\pi x_r s_r(x_r) dx_r = \frac{2\pi \bar{h}_r}{\pi^2} \left[1 - \left(1 + \theta_r \sqrt{L_r/\pi}\right) e^{-\theta_r \sqrt{L_r/\pi}}\right] \tag{2.1}$$

It follows from Eq.2.1 that if all cities are endowed with equal commuting technology (*i.e.* quality of transport infrastructures) $\theta_r > 0$, commuting times would only depend (positively) on the number of workers in each city (L_r).

Since workers are identical, wages net of commuting costs should be equal in every location within each-city. With such condition, one can obtain the equilibrium land rent R_r. Given distributed land and firms ownership, the per-capita expenditure is thus the sum of equilibrium net wages, the individual share of aggregate land rents

(ALR) and of firms' profits.

Market equilibrium in the urban system

By assuming a Pareto distribution for firms' productivity draws, the model can be solved in its equilibrium conditions. In a multiple-cities setting, these equilibrium conditions are i) labour market clearing (Eq.2.25 in the Appendix), ii) trade balance (*i.e.* the total value of exports must equate the total value of imports per each city, as in Eq.2.26 in the Appendix) and iii) zero expected profits (Eq.2.27, in the Appendix). The outcome of the previous conditions can be combined to obtain two equilibrium relationships (Eq.2.2 and Eq.2.3), which depend entirely on the internal productivity cut-off (m_r^d)[38], and two unknowns: the city-specific technological frontier μ_r^{max} and the equilibrium wages $\hat{w}_r$).

$$\frac{h_r}{(m_r^d)^{k+1}} = \sum_s S_s \tau_{rr} \left(\frac{\tau_{rr} \hat{w}_r}{\tau_{sr} \hat{w}_s} \right)^k \frac{1}{\mu_s^{max}} \tag{2.2}$$

$$\mu_r^{max} = \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss} \hat{w}_s}{\tau_{rs} \hat{w}_k} m_s^d \right)^{k+1} \tag{2.3}$$

The previous two equation represent an important aspect of this model, as they constitutes its invertibility: once informed with data about distances, estimated trade elasticity, and labour supply (through data on average hours worked and commuting times), the model can be fully solved and allows to quantify the effect of an exogenous change in one variable. Given the research question at hand, the

38 The internal productivity cut-off, m_r^d, represents the maximum marginal labour requirement for a firm to sell a positive quantity of product at least in the domestic (d) market. This is expected to be higher than external cutoffs m_{rs} required to export from city r to any city s, with the sole exception of cases in which the product of local wages per transport cost is higher than that of external wages per export costs ($w_r t_{rr} \geq w_s t_{sr}$).

variable to modify is the city's average commuting-time, affecting individual labour supply s_r.

Spatial Equilibrium

The static equilibrium of the model is initially solved by assuming population as fixed. Once this has been done, workers are allowed to endogenously relocate following their preferences. To do so, city-specific amenities (A_r) and taste heterogeneity (ξ_r^l) are introduced. In particular, individual ()idiosyncratic taste differences for residential location (ξ_r^l), are assumed to be independent and identically distributed across individuals and cities, according to a double distribution with zero mean and variance $\pi^2\beta^2/\sigma$. The location choice of individuals is assumed to depend on a *linear random utility* à la McFadden (1974): a linear function of the utility (Eq. 2.10 in the Appendix and Eq. 2.5), of observed amenities, A_r^o, (like climate, topography, and water area), and of unobservable ones (A_r^u), and the error term ξ_r^l representing individual taste heterogeneity:

$$V_r^l = U_r + A_r + \xi_r^l \tag{2.4}$$

The indirect utility U_r can be shown to be, in equilibrium, directly proportional to the hours worked, h_r, and inversely to the internal cutoff m_r^d and to local trade costs (t_{rr}),as of inversely proportional to the weighted (per expenditure shares) city-average markup Λ_r ($\Lambda_r = \frac{k3\tau_{rr}m_r^d}{\alpha h_r}$)[39]:

$$U_r = \frac{\alpha}{\tau_r r}\frac{h_r}{m_r^d}\left[\frac{1}{(k+1)(k1+k2)} - 1\right] = \frac{k3}{\Lambda_r}\frac{h_r}{m_r^d}\left[\frac{1}{(k+1)(k1+k2)} - 1\right] \tag{2.5}$$

The spatial equilibrium of the model will thus be defined by matching the choice probability of workers defined by the across-city utility distribution, and the city's share of the economy's total population. In so doing, it can be obtained the unique city size distribution that satisfies the following equation:

$$P_r(V_r^l > \max_{s \neq r} V_s^l) = \frac{exp((U_r + A_r)/\beta)}{\sum_{s=1}^{K} exp((U_s + A_s)/\beta)} = \frac{L_r}{\sum_{s=1}^{K} L_s} \tag{2.6}$$

This can be solved by imposing, without loss of generality[40], a standardization of the utility ($exp(D_r = (U_r + A_r)/\beta)$) taking as numeraire city 1, such that $exp(D_{r=1}) = 1 \Rightarrow D_{r=1} = 0$, which implies $\frac{exp(D_r)}{exp(D_1)} = exp(D_r) = \frac{L_r}{L_1}$, $\forall\ r$, such that D_r is uniquely determined by $D_r = ln(L_r/L_1)$, $\forall\ r$.

Finally, for $\hat{D}_r = (U_r + A_r^o + A_r^u)/\beta$, the location choice parameters, identifying the role of consumption (α_1), natural α_2 and unobserved amenities ϵ_r, can be derived

39 Both U_r and Λ_r are a function of $k1$, $k2$ and $k3$, representing constants that solely depend on the productivity's Pareto distribution shape parameter k. The shape parameter has been chosen accordingly to the original model by Behrens *et al.* (2017), and set to 6.4. However, when testing different values of k (such as k=1.2, or k=2 as in Del Gatto *et al.*, 2006), the results of the counterfactual estimation are both qualitatively and quantitatively comparable. This is consistent with Combes *et al.* (2012)'s findings on the indifference of comparative static results with respect to the choice of productivity distribution parametrization.
Finally, the $k1$, $k2$ and $k3$ constants are obtained as the unique solutions of the integration of the main model equations written as to exploit the Lambert function transformation properties, as shown in Behrens *et al.*, 2017.

40 Given that the model deals with relative location choices, this assumption has no impact on the results.

through the following OLS estimation:

$$\hat{D}_r = \alpha_0 + \alpha_1 \hat{U}_r + \alpha_2 A_r^o + \epsilon_r \tag{2.7}$$

2.3.2 Counterfactual estimation

Using the previous model, the effects of an higher teleworking adoption on the observed spatial equilibrium can be calculated by resorting to a counterfactual estimation. The main aim of this study is indeed to predict how the adoption of remote-work to its full potential, determined by the pre-pandemic occupational structure of cities, will affect city-sizes. In particular, the focus will be on the potential workers relocation across MSAs of different sizes, trying to understand if smaller ones will benefit from such arrangements.

To evaluate the effects of a potential full adoption of remote-work, the total net labour supply must first be allowed to change. This change will be proportional to the local share of potential remote-workers (sh_{FHW}) multiplied per the baseline average hours worked in the city:

$$\widetilde{h_r} = \widetilde{s_r} = h_r(1 - sh_{FHW}) + \bar{h} sh_{FHW} \tag{2.8}$$

In such a way, the city-specific counterfactual individual labour supply $\widetilde{s_r}$ will be determined as the average between the city's gross hours worked (h_r + commuting times) for workers in teleworkable occupations and the net-of commuting hours worked by non-remote workers (h_r), weighted by their relative share.[41]

41 It should be noted that increasing net-working hours and changing the commuting technology parameter θ_r are indifferent in this setting, as one implies the other. Indeed, a lower h_r will map into a lower θ_r. Modifying working hours can thus be chosen for the sake of notation simplicity.

In the first step of the counterfactual exercise, the new spatial equilibrium is first estimated 'before' locational adjustment (short term). The higher labour supply in cities will allow the productivity cutoffs to increase, and more firms to enter the city N_r^{E0} and survive (N_r^{C1}). This will make possible to produce a greater number of varieties in each city with any positive share of FHW workers. The two equilibrium equation (Eq.2.2 and Eq.2.3) obtained in the previous sub-section, must then be solved for the new wages $\widetilde{w_r^0}$ and cutoffs $\widetilde{m_r^{d0}}$, as a function of the baseline city size L_r and of the counterfactual average hours worked $\widetilde{h_r}$. The location choice parameters obtained in the baseline utility estimation of Eq. 47 ($\alpha_0 = -1.330921$, $\alpha_1 = .021724$, $\alpha_2 = .0821857$ and η_r, referring to the constant term, the coefficient of the indirect utility representing consumption amenities, the coefficient of natural amenities, and the predicted error representing taste heterogeneity, respectively), are held as constant. The counterfactual short-term equilibrium wages $\widetilde{w_r^0}$ and cutoffs $\widetilde{m_r^{d0}}$ will yield new utility levels $\widetilde{U_r^0}$ (see Eq.2.10) and markups Λ_r^0.

Allowing the population distribution to adjust to the updated utility distribution ($\frac{exp(\widetilde{U_r^1})}{\sum_{s=1}^{K} exp(\widetilde{U_s^1}}$) as affected by the short-term counterfactual number of firms (as in, varieties) $\widetilde{NE_r^0}$, and the short-term wages $\widetilde{w_r^0}$, will yield a new spatial equilibrium, in turn yielding new wages $\widetilde{w_r^1}$ and cutoffs $\widetilde{m_r^{d1}}$. Iterating this procedure until the convergence of the population distribution, allows to obtain the 'long-term' counterfactual equilibrium results that will be discussed, together with the short term ones, in Section 2.4.

2.3.3 Data and model quantification

The model is quantified at the level of US Metropolitan Statistical Areas (MSA),[42], by exploiting the same data sources of the original Behrens *et al.* (2017)'s model to ensure results comparability. However, in order to have the most updated results as possible, the baseline model is quantified on 2017 data.[43]

As far as data sources are concerned, data on average commuting times and total hours worked by MSA are drawn from American Community Survey 1-year data.[44]

Data on average hours worked comes from the Bureau of Labour Statistics (BLS). BLS data for 2017 are also exploited to obtain aggregate employment data at MSA level[45] in two variants: for all sectors and all occupations (L_r), and in teleworkable 6-digit SOC (Standard Occupational Classification) occupation in all sectors ($L_{r_{FWH}}$). The latter variable, representing the city-specific baseline total employment in teleworkable occupations, exploit Dingel and Neiman (2020) measure of teleworkability,

42 Micropolitan Areas are excluded for both lower data availability and substantial differences in economic characteristics. Furthermore, the MSAs located in non-contiguous states (Alaska, Puertorico and Hawaii) have been excluded. The resulting sample is made up of 373 out of 389 MSAs.

43 The 2017 baseline is preferred to the more recent 2019 available data, for a matter of data consistency, since in that year all estimates issued by the US Bureau of Labour Statistic and by the American Community Survey use the same definition of Metropolitan Areas, *i.e.* that of the O.M.B. bulletin of July 2015.

44 These survey is conducted by exploiting the definition of CBSA areas and boundaries at December, 1st, 2009, for the years 2010, 2011, and 2012, the February 2013 definition for the years 2013, 2014, and 2015, the 2015 July 15th definition for 2016 and 2017 estimates, the 2017 August 15th definition for the year 2018, and the September 14th 2018 for the 2019 estimates. The 2015 definition is used for the baseline estimates in 2017.

45 For the sake of comparability with Behrens *et al.* (2017)'s results, city-population, *i.e.* the number of workers, is expressed in hundreds of thousands of people. This is necessary in order to obtain comparable values of μ_r^{max} ranging [0;200] as those declared in Behrens *et al.* (2017), since all the right-hand side factors of Eq.2.27 are around the unity, so that population L_r is the variable defining the magnitude of μ_r^{max}. This transformation is also consistent with Behrens *et al.* (2017) declared value of θ_r [0;1], which again would otherwise range $[1e^{-06}; 1e^{-05}]$ if population was measured in person units. One must thus conclude that, while not explicitly declared, the original model was quantified with population expressed in hundreds of thousands.

and is obtained by counting as teleworkable those occupation whose activity can be done remotely for a minimum of a 80% of working hours. This measure thus considers as fully remote those jobs that allow to commute to the workplace for 1 out of 5 days per average week.

Trade frictions across ($\tau_{rs} = d_{rs}$) and within-MSAs (τ_{rr}) are quantified by estimating the distance elasticity γ through a log-linear stochastic gravity equation. In this equation, the bilateral trade flows across US states $ln(X_r s)$ (from the 2017 Commodity flow Survey dataset) are regressed with respect to distance (weighted by the shape parameter k), a zero-flow dummy (I^0_{rs}), origin and destination fixed effects (χ^1_r and χ^2_s), and a constant ι. For $k = 6.4$, this procedure produces a value of the distance elasticity of trade $\gamma = .024338$.

$$lnX_{rs} = \iota - k\gamma ln(d_{rs}) + I^0_{rs} +^1_r +^2_s + \epsilon_{rs} \tag{2.9}$$

The total cost of employees is based on the 'compensation of employees' variable provided by the US Bureau of Economic Analysis. In turn, this is constituted by the sum of 'wages and salaries', 'employer contributions for employee pension and insurance funds' and of 'employer contributions for government social insurance'. However, these are used as initial guesses for the equilibrium wages. The latter are found to differ only marginally from the observed wages.

Finally, data on natural amenities are obtained from the US Department of Agriculture (available at county level and aggregated by MSA), while data on bilateral trade flows at state level are obtained from the US Census Commodity Flow Survey (CFS).

Table 2.1 in the Appendix reports the main baseline descriptive statistics, together with the main estimated long-term counterfactual changes.

2.4 Results

Figures 2.1, 2.2 and 2.3 report in different scatterplots the data about: along the vertical axis, the long-term counterfactual percentage change in population; along the horizontal axis, the initial city-population (log of the ratio between initial population and its mean), and the share of workers in potentially fully remote-workable occupations (FHW). The relationship between the initial and counterfactual city-size is clearly positive, meaning that cities with larger shares of initial population (Fig. 2.1) and of potential remote-workers (Fig. 2.2), would attract more workers at the expenses of smaller cities, through larger utility gains (Fig. 2.3). Given the strong correlation existing between the initial city sizes and the share of potential FHW workers (0.4946, pValue=0.000), the two figures (2.2) and 2.2) are almost overlapping. However, this correlation is large enough to make the slope of their linear correlation with population changes to be similar, but small enough to motivate the investigation of city-specific commuting costs changes such as the ones here presented.

The relationship between population (or utility) changes and initial city-size (or initial share of employment in remote-workable occupations), as shown by the quadratic fit in Figures 2.1-2.3, indicating that agglomerative effects are at play, and that the largest cities attract workers disproportionately to their size.

This is a pretty important general result, whose explaining mechanisms can be disentangled using the framework of the proposed QSE model. However, it must pointed out that the computed changes in city-size do not give rise to a change in the rank-size rule, as shown in Fig. 2.4.

A number of mechanisms concur to determine the ascertained 'fortune' of larger cities, with differences between the short and the long run.

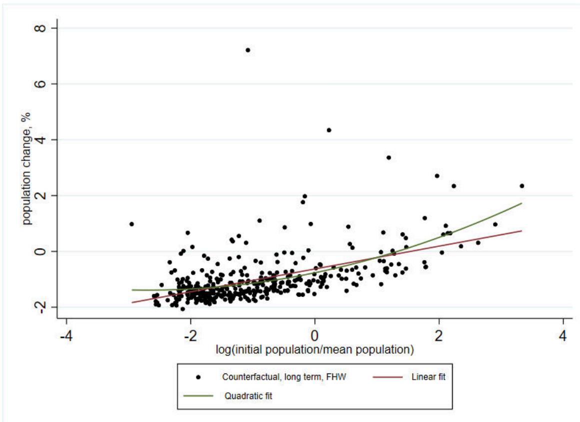


Figure 2.1. Scatterplot, linear and quadratic prediction of counterfactual population change in MSAs, with respect to initial population.

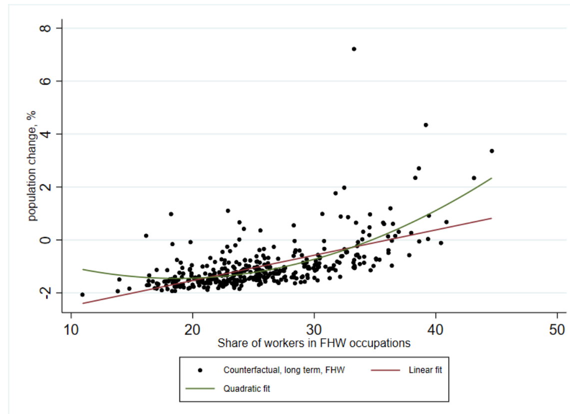


Figure 2.2. Scatterplot, linear and quadratic prediction of counterfactual population change in MSAs, with respect to the initial share of employment in fully remote-workable occupations.

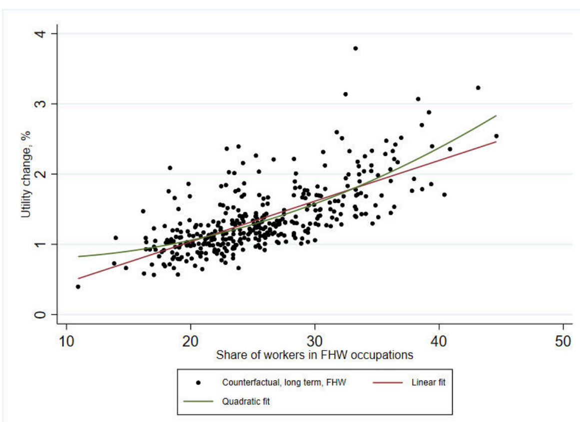


Figure 2.3. Scatterplot, linear and quadratic prediction of counterfactual utility change in MSAs, with respect to initial employment share in fully remote-workable occupations.

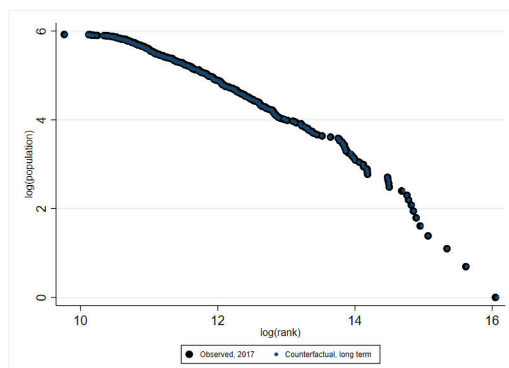


Figure 2.4. Rank-size rule comparison.

Since workers under remote-work contracts need to commute less frequently, all cities with positive shares of telecommuters are expected to face savings in average commuting times. In the short run, when population does not adjust its location in response to the consequences of the increased remote-work allowance, firms will face a larger net supply of labour at the same average cost per worker[46]. These savings will allow for a higher internal productivity cut-off, and thus for more firms to be selected in each market, increasing the number of produced varieties at the local price. The latter will translate on the consumption side into large utility-gains through larger consumption amenities. The higher number of firms in the city will also produce higher wages. Larger cities, while usually endowed with a better commuting technology θ_r, also tend to face higher average commuting times, as these are a function of total population (see Table 2.1). In the light of that, all else being equal, larger cities would face larger increases of net labour supply than their smaller counterparts. The same is true for cities with larger shares of employment in remote-workable occupations: however, as mentioned above, the two features (size and share of potential remote-workers) tend to coincide, reinforcing the mechanism. In the long run, when population is allowed to move, the firms' advantages linked to a higher net labour supply (allowing larger productivity cutoff) will be offset, since the workers inflows in larger cities (attracted by the higher increases in wages and consumption amenities) will increase competition, decreasing markups. In particular, cities that had a larger share of potential teleworkers and largest commuting costs (as it usually happens in large cities), will see the largest increases in size and average productivity.

---

46 Given the model structure, the reduction in average commuting times maps into a larger net labour supply enjoyed by firms. While this might seem an overly simplifying assumption, it may is useful to retain the savings that firms with large shares of remote workers can obtain by cutting office space and utilities. Moreover, since the commuting time savings for workers are not directly represented in the model, the average estimated city-size change ($\Delta L = 0.1\%$, with a maximum of $+7.2\%$) should be considered as a conservative lower bound.

On the contrary, cities losing population (in general, smaller cities, but also MSAs with low shares of employment in remote-workable occupations) will also face a drop in the number of firms and, consequently, in average productivity.

Still in the long run, utility will rise everywhere with respect to the baseline equilibrium (as shown in Figure 2.3), due to the general decrease in markups, that, while originated by the tougher selection in the cities gaining population, will affect also smaller cities through the pro-competitive effect of trade. Indeed, as it can be observed in Figure 2.5, the change in markups is negative also in those MSA that loose population.

In the Appendix to this chapter in Table 2.1 are reported, for each MSA, the initial values (based on 2017 data) and the counterfactual long-term equilibrium changes of the city-size ranking, the total employment, the average weekly commuting hours, the estimated commuting technology parameter and the estimated average markup. Always in Table 2.1, I also report the long-term counterfactual changes in utility, and the baseline values of the GDP, the number of firms, the share of workers in remote-workable occupations and the estimated value of natural and unobserved amenities. There it can be observed how utility (and markups) changes are positive (negative) everywhere, even in cities losing population. It can also be noted that most large cities exhibit above average commuting hours, below average commuting technology parameter (where a lower parameter indicates a better technology), above average shares of potential FHW employment. Moreover, it is interesting to notice how, as previously mentioned, city-size and the share of potential FHW employment are highly correlated but not always consistent, determining, together with other model parameters, the city-specificity of remote-work adoption effects on the city's attractiveness. For example the Riverside-San Bernardino-Ontario MSA of California, was the 17th largest MSA in the U.S. in 2017. Notwithstanding its initial size and the large estimated values of its natural and unobserved amenities[47], full potential remote-work adoption would make the city loose one size-rank position and 0.6% of its total employment in the long run, due to its below-average share of FHW (25.25%). This relationship of course is not as straightforward, as the city's attractiveness depends also on the availability and price of external varieties depending on the trade costs and accessibility of the city. Some smaller MSA such as Midland, Texas, ranked 205 by employment size and below average share of potential FHW employment and with quite low natural and unobserved amenities, would still see its population increase by 0.4% and gain 3 rank positions in the after math of remote-work adoption. Altogether these results, while generalizable to some extent, may help explaining the lack of agreement in previous studies on the effects of remote-work on relocation patterns. By predicting the complex effect that the increase in remote-work adoption could have on cities competitiveness, it can help policy makers to make informed decisions and to be prepared for the expected changes.

In concluding the presentation of the results, it should be noted that these are comparable to those of Behrens *et al.* (2017) in terms of general predictions, particularly with respect to the generalised advantages of larger cities or the negligible changes in the size distribution of cities. However, many of the cities that would display population gains in response to an horizontal cut in commuting costs as portrayed in Behrens *et al.* (2017), are now shown to loose population in the face

47 The unobserved amenities parameter, estimated through the error of equation , accounts for all unobserved city-characteristics (and for the aggregation of heterogeneous individual preferences), including cultural amenities. This class of amenities has been found by the literature to be correlated with city-size, and is deemed to have an impact on cities competitiveness and on the attraction of skilled workers (Falck *et al.*, 2018).

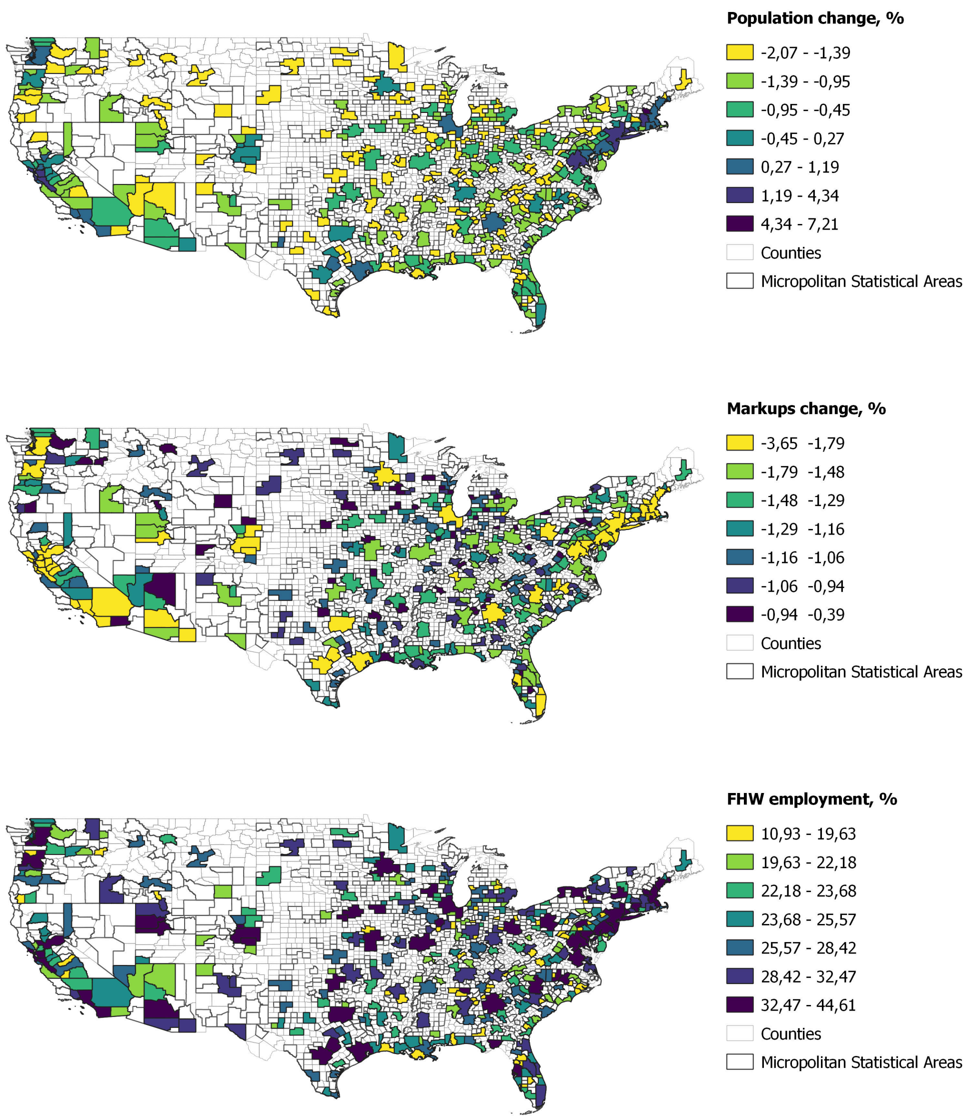


Figure 2.5. Counterfactual changes in population and markups, and share of employment in fully remote-workable occupations.

of a telecommuting-dependent urban frictions shrinkage. This result may be due to a lower exposure to remote-workable occupations notwithstanding a larger initial city-size. Furthermore, the results in terms of the effects of remote-work adoption in terms of population growth mainly benefiting larger cities are consistent with the statistical evidence collected by Aizhan *et al.* (2022), obtained by analysing house-price changes in OECD countries following the Covid-19 pandemic outbreak.

2.5 Conclusions

Inspired by the revamp of the debate about remote work in the aftermath of Covid-19, this study has aimed to provide a quantification of the effects that city-specific potential levels of remote-work adoption could have on workers distribution, welfare and productivity. Indeed, while the literature on the topic is abundant, the reliability of its results is challenged by a set of empirical problems that this article has tried to overcome. To do that, this work has provided a new counterfactual estimate of the QSE model proposed by Behrens *et al.* (2017), in order to evaluate the effect of city-specific changes in commuting cost that realistic levels of remote-work adoption could originate.

The main result of this model, and of the counterfactual exercise it enables, contradicts the general idea, on which the post-pandemic debate has returned to focus, that remote work represents a possible strategy to attract population in smaller cities (defined at the metropolitan area level). Indeed, the obtained empirical evidence shows that larger cities would attract even larger shares of residents in the aftermath of remote-work adoption, growing in terms of size and productivity. The results are in line with recent evidence suggesting a deepening of the urban-rural gap effect of remote-work, that is found to increase workers attraction (Braesemann *et al.*, 2022; Aizhan *et al.*, 2022), urban productivity and average wages (Kyriakoupoulou and Picard, 2022) only in large cities. Furthermore, recent evidence produced exploiting housing-prices for OECD countries, suggest that larger metropolitan areas have seen the largest increases in housing priced and demand since the Covid-19 outbreak. This result appears the outcome of the combination of different agglomeration forces, linked to initial consumption and productivity advantages, to the higher frictions (and savings) entailed in their size, to natural, unobserved and consumption amenities, and to the higher share of workers in remote-workable occupations that larger MSAs tend to display.

The results of the study have some important policy implications. While small and medium-sized cities and cities with low shares of employment in remote-workable occupations will loose population and see their average productivity decrease, the promotion of remote-work could produce welfare and efficiency gains everywhere linked to markups decreases, and productivity gains in larger cities. However, it should also be pointed out that while optimal in terms of utility, aggregate productivity and efficiency, remote-work adoption has the potential to increase the core-periphery structure and the inequality across cities. Indeed, only 36 out of 373 cities will benefit from size and productivity gains, whereas the average MSA will loose 1% of its population. Furthermore, local policy makers could make use of the by-city results here portrayed, in terms for example of city-planning. Indeed, while city-size and the share of potential remote-workers can be considered as good predictors of productivity and population changes, the effect that remote-work adoption to its full potential will have on each MSA will also depend, among other factors, on trade costs and on the changes in the system of cities in the economy.

In concluding, it should be recognised that, in its present form, the proposed modeling strategy suffers from some limitations. The first one refers to the fact that the model does not account for heterogeneous workers. For this reason, it is not suited to analyse neither sorting nor within-city welfare inequality implications, which would be particularly interesting to analyse given the mentioned evidence on the uneven access to remote work. Furthermore, given the endogenous productivity differentials of firms, the above extension would allow to account for matching mechanisms, which could offer further insights on firm and city-level productivity changes. The second limitation is partly related to the first one, and regards heterogeneous residential preferences, that de Vos *et al.* (2018) find as relevant in explaining commuting behaviors and location decisions. These are accounted for in this model, but in a simplistic way (through idiosyncratic preferences), which do not allow for systematic differences in high versus low-skilled individuals. If large shares of remote-workers were to exhibit a stronger-than-average preference for small-sized cities, the results could marginally change, as consumption amenities would reduce their impact[48]. Third, the framework of the presented model would benefit from a more detailed land market, accounting for diminished land consumption by firms and for increases in residential land demand by remote workers. Such a setting would entail a reduced income-effect for teleworkers, and as such relocation could be driven towards cities with more affordable rents. Fourth, the proposed model abstracts from within-city relocation concerns, to focus on the competitive effects that differential levels of telework adoption across cities could give rise to. While this is a limitation, in the light of the reviewed extant studies, the discussion of the obtained

48 It should be noted that average preferences for natural amenities have been accounted for and estimated.

results can be extended to a within-city context with the following line of reasoning. As highlighted in Section 2.2, the reduction in residential location constraints and the savings in commuting costs, as induced by full-day telework agreements, do not necessarily induce a moving choice nor one with a specific predictable direction (towards or farther away from the center). The relocation direction seems, instead, to be driven by heterogeneous characteristics and preferences, by housing and transport prices, and by consumption and natural amenities. In presence of endogenous amenities, such as our and in Almagro and Dominguez-Iino (2021)'s context, sorting in space by income is expected to get reinforced. Income effects such as that of telecommuting savings, are thus likely to lead to movements towards service and amenities, so towards larger cities in our framework and towards the center in a within-city one.

Despite the above listed shortcomings, this study provides an important contribution to the debate on the short and long-term effects of telework, which still lacked of a proper across-cities evaluations. In particular, while the portrayed results are largely qualitatively comparable to those in the reference model by Behrens *et al.* (2017), the estimated change in population could be considered more realistic than the one they obtained by considering the total and equal disappearance of commuting costs in all cities.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6): 2127–2189.

Aizhan T,, Di Lauro, S., Antonelli, G., (2022). Remote Work at the Time of COVID-19 Pandemic and Beyond: A Scoping Review. In: Mondal S.R., Di Virgilio, F., Das, S., (Eds.), *HR Analytics and Digital HR Practices*, 127-169, Springer.

Albouy, D.,Stuart, B. A. (2020). Urban Population and Amenities: the Neoclassical Model of Location. *International Economic Review*, 61(1): 127–158.

Almagro, M. and Dominguez-Iino, T., (2021). Location Sorting and Endogenous Amenities: Evidence from Amsterdam. *Working Paper*, 1-80.

Althoff, L., Eckert, F., Ganapati, S., Walsh. C. (2022). The Geography of Remote Work. *Regional Science and Urban Economics*, 93 (March): 103770.

Asgari, H., Jin, X. (2015). Toward a comprehensive telecommuting analysis framework: Setting the conceptual outline. *Transportation Research Record*, 2496: 1–9.

Athanasiadou, C., Theriou, G., (2021). Telework: Systematic Literature Review and Future Research Agenda. *Heliyon*, 7: e08165.

Autor, D. H., Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labour Market. *American Economic Review*, 103(5): 1553–1597.

Bacolod, M., Blum, B. S., Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65(2): 136–153.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., Yannelis, C. (2020). How does household spending respond to an epidemic? consumption during the 2020

COVID-19 pandemic. *Review of Asset Pricing Studies*, 10(4): 834–862.

Barrero, J. M., Bloom, N., Davis, S. J. (2020). COVID-19 is also a reallocation shock. *Brookings Papers on Economic Activity*, Special Edition, 329–383.

Barrero J. M., Bloom, N., Davis, S. J., (2021). Why Working from Home Will Stick. *NBER Working Paper Series*, 28731.

Behrens, K., Duranton, G., Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553.

Behrens, K., Mion, G., Murata, Y., Suedekum, J., (2017). Spatial frictions. *Journal of Urban Economics*, 97:40–70.

Behrens, K., Kichko, S., Thisse, J-F., (2021). Working from Home: Too Much of a Good Thing? *CESifo* Working Paper Series 8831: 1-46.

Béland, L.-P., Brodeur, A., Wright, T., (2020). The Short-Term Economic Consequences of COVID-19: Exposure to Disease, Remote Work and Government Response. *GLO Discussion Paper Series*, Global labour Organization, 524: 1-90.

Bertschek, I., Polder, M., Schulte, P., (2017). ICT and resilience in times of crisis: Evidence from cross-country micro moments data. *Economics of Innovation and New Technology*, Taylor & Francis Journals, 28(8):759-774.

Bick, A., Blandin, A., Mertens, K., (2021). Work from Home Before and After the COVID-19 Outbreak . Available at SSRN: http://dx.doi.org/10.2139/ssrn.3786142

Bloom N., Liang J., Roberts, J, Ying, Z.J., (2015). Does Working from Home Work? Evidence from a Chinese Experiment, *The Quarterly Journal of Economics*, Oxford University Press, 130(1): 165-218.

Braesemann, F., Stephany, F., Teutloff, O., Kässi, O., Graham, M., Lehdonvirta, V., (2022). The global polarisation of remote work. PLoS ONE17(10): e0274630

BLS - U.S. Bureau of Labour Statistics, (2019). Job Flexibilities and Work Schedules Summary.

Caulfield, B. (2015). Does it pay to work from home? Examining the factors influencing working from home in the Greater Dublin Area. *Case Studies on Transport Policy*, 3(2): 206–214.

Choo S., P. L. Mokhtarian I. Salomon, (2005). Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the U.S. *Transportation*, 32: 37-64.

Choudhury, P., Foroughi, C., Larson, B. (2019). (Live and) Work from Anywhere: Geographic Flexibility and Productivity Effects at the United States Patent Office. *Harward Business review*, Working Paper, 19(054): 1-57.

Chronopoulos, D. K., Lukas, M., Wilson, J. O. . (2020). Consumer Spending Responses to the Covid-19 Pandemic: Great Britain assessment.*Center for Responsible Banking & Finance*, University of St.Andrews, 1–40.

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

Couture, V., Gaubert, C., Handbury, J., Hurst, E. (2019). Income Growth and the Distributional Effects of Urban Spatial Sorting. *SSRN Electronic Journal*, 1-49.

Couture, V., Dingel, J.I., Green, A., Handbury, J., Williams, K.R., (2021). JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19. *Journal of Urban Economics*: 103328

Davis, M.A., Ghent, A.C., Gregory, J.M, (2021). The Work-from-Home Technology

Boon and its Consequences. *NBER Working Papers*, 28461.

de Abreu e Silva, J., Melo, P. C., (2018). Does home-based telework reduce household total travel? A path analysis using single and two worker British households. *Journal of Transport Geography*. Elsevier, 73(C): 148-162.

de Vos, D., Meijers, E., van Ham, M., (2018). Working from home and the willingness to accept a longer commute. *The Annals of Regional Science*, 61:375–398.

De Fraja, G., Matheson, J.,Rockey, J. (2021). Zoomshock: The Geography and Local Labour Market Consequences of Working from Home. *Working Paper*.

Del Gatto, M., MION, G., Ottaviano, G.I.P., (2006). Trade Integration, Firm Selection and the Costs of Non-Europe. *CEPR Discussion Paper*, 5730.

Dingel, J. I. and Neiman, B., (2020). How many jobs can be done at home? *NBER* Working Paper, 26948: 1-19.

Donnelly, N., Proctor-Thomson, S., (2015). Disrupted Work: Home-Based Teleworking (HbTW) in the Aftermath of a Natural Disaster. *New Technology, Work and Employment*, 30.

Eeckhout, J., Pinheiro, R., Schmidheiny, K., (2014). Spatial Sorting, *Journal of Political Economy*, 122:554–620.

Eeckhout, J., Hedtrich, C., Pinheiro, R. (2019). Automation , Spatial Sorting , and Job Polarization. *IZA Conference*. Working Paper.

Ellen, I. G., Hempstead, K. (2002). Telecommuting and the demand for urban living: A preliminary look at white-collar workers. *Urban Studies*, 39(4): 749–766.

Fajgelbaum, P. D., Gaubert, C. (2020). Optimal spatial policies, geography, and sorting. *Quarterly Journal of Economics*, 135(2): 959–1036.

Falck, O., Fritsch, M., Heblich, S., Otto, A., (2018). Music in the Air: Estimating the Social Return to Cultural Amenities. *Journal of Cultural Economics*, 42 (3): 365–91.

Florida, R., Rodríguez-Pose, A., Storper, M., (2021).Critical Commentary: Cities in a post-COVID world. *Urban Studies*, 60(8): 1509-1531.

Gabaix, X. , Ibragimov, R. , (2011). Rank-1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29: 24–39.

Gaigné, C., Koster, H. R. A., Moizeau, F., Thisse, J. F. (2019). Who Lives Where in Cities? Amenities, Commuting and Income Sorting. *Working Paper*, 1-59.

Glaeser, E. L. (2022). Reflections on the Post-Covid City. *Cambridge Journal of Regions, Economy and Society*, 15(3): 747–55.

González-Leonardo, M., Rowe, F., Fresolone-Caparrós, A., (2022). Rural Revival? The Rise in Internal Migration to Rural Areas during the COVID-19 Pandemic. Who Moved and Where? *Journal of Rural Studies*, 96(December): 332–342.

Gschwind, L., Vargas, O. (2019). Telework and its effects in Europe. In *J. C. Messenger, Ed., Telework in the 21th century*. Elgar, 36-75.

Gubins, S., Ommeren, J., Graaff, T., (2019). Does new information technology change commuting behavior?, *The Annals of Regional Science*, Springer. Western Regional Science Association, 62(1): 187-210.

Gupta, A., Peeters, J., Mittal, V., Van Nieuwerburgh, S. (2021). Flattening the Curve: Pandemic-Induced Revaluation of Urban Real Estate. *SSRN Electronic Journal*.

He, S. Y., Hu, L. (2015). Telecommuting, income, and out-of-home activities. *Travel Behaviour and Society*, 2(3): 131–147.

Hook, A., Court, V., Sovacool, B. K., Sorrell, S. (2020). A systematic review of the energy and climate impacts of teleworking. *Environmental Research Letters*, 15(9): 1-30.

JRC, (2020). Telework in the EU before and after the COVID-19: where we were, where we head to. *JRC Science for Policy Brief*.

JRC-Eurofound (2020). Teleworkability and the COVID-19 crisis: a new digital divide? *JRC Working Papers Series on Labour, Education and Technology*, 2020(05): 1-70.

Kim, S. N., Choo, S., Mokhtarian, P. L. (2015). Home-based telecommuting and intra-household interactions in work and non-work travel: A seemingly unrelated censored regression approach. *Transportation Research Part A: Policy and Practice*, 80(2015): 197–214.

Kim, S.,N., (2016). Two traditional questions on the relationships between telecommuting, job and residential location, and household travel: revisited using a path analysis. *Annals of Regional Science*. 56(2): 537–563

Kontou, E., Murray-Tuite, P., Wernstedt, K. (2017): Duration of commute travel changes in the aftermath of Hurricane Sandy using accelerated failure time modeling. In *Transportation Research Part A: Policy and Practice*, Elsevier, 100(C): 170-181.

Kosteas, V. D., Renna, F., Scicchitano, S., (2022). Covid-19 and Working from Home: toward a "new normal"? *GLO Discussion Paper Series*, 1013, Global labour Organization.

Kyriakoupoulou, E., Picard, P. (2022). The Zoom City: Working From Home, Urban Productivity and Land Use, *DEM Discussion Paper Series*, 22-15, Department of Economics at the University of Luxembourg.

Lund, J, Mokhtarian, P.L., (1994). Telecommuting and residential location: Theory and implications for commuter travel in monocentric metropolis. *Transportation Research Record*, 1463: 10–14.

Manson, S., Schroeder, J., Van Riper, D., Ruggles, S., (2019). IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis.

Martin, R., Simmie, J. (2008). The theoretical bases of urban competitiveness : does proximity matter? *Revue d'Économie Régionale & Urbaine*, 3: 333–351.

McFadden, D.L., (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In: *Frontiers in Econometrics*, Wiley, New York. Academic Press, 105-142.

McKinsey Global Institute, (2021). The future of work after COVID-19, *The Post-Pandemic Economy*, 1-149. Retrieved the 1st of March 2021, from www.mckinsey.com

Melo, P. C., de Abreu e Silva, J. (2017). Home telework and household commuting patterns in Great Britain. *Transportation Research*, Part A: Policy and Practice, 103: 1–24.

Messenger, J. C. (2019). Telework in the 21st Century. *Elgar Publishing*, 1–316

Mongey, S., Pilossoph, L. and Weinberg, A. (2021). Which Workers Bear the Burden of Social Distancing? *The Journal of Economic Inequality*, 19: 509–526.

Monte, F., Redding, S. J., Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12): 3855–3890.

Monte, F., Porcher, C., Rossi-Hansberg, E, (2023). Remote Work and City Structure. *Working Paper*.

Monteiro, N.P., Straume, O.R., Valente, M., (2019). Does remote work improve or impair firm labour productivity? Longitudinal evidence from Portugal, *NIPE*

Working Papers, Universidade do Minho, 14/2019.

Muhammad, S., Ottens, H. F. L., Ettema, D., Jong, T. (2007). Telecommuting and residential locational preferences: A case study of the Netherlands. *Journal of Housing and the Built Environment*, 22(4): 339–358.

Nilles, J.M., (1991). Telecommuting and urban sprawl: mitigator or inciter? *Transportation*, 18: 411–432.

O'Brien, W., Aliabadi, M.F., (2020). Does telecommuting save energy? A critical review of quantitative studies and their research methods. *Energy and Buildings*, 225(10): 1-15.

Onken-Menke G., Nüesch, S., Kröll, C. (2018). Are you attracted? Do you remain? Meta-analytic evidence on flexible work practices. *Business Research*. Springer, German Academic Association for Business Research, 11(2): 239-277.

Ory, D. T., and Mokhtarian, P. L., (2006). Which came first, the telecommuting or the residential relocation? An empirical analysis of causality. *Urban Geography*, 27(7): 590–609.

Ota, M. (2017). Telecommuting and Urban Spatial Structure in Metropolitan Areas. In *Japanese Economic Association 2017 Spring Meeting*, 1–17.

Puga, D., (2017). Learning by working in big cities. *Review of Economic Studies*, 84: 106–142.

Ramani, A., Bloom, N., (2021). The Donut Effect of Covid-19 on Cities. *NBER Working Paper*, 28876.

Rappaport, J. (2008). Consumption amenities and city population density. *Regional Science and Urban Economics*, 38(6): 533–552.

Redding and Rossi-Hansberg, (2017). Quantitative Spatial Economics. *Annual Re-*

view of Economics, 9(1):1–58.

Rhee, H-J., (2009). Telecommuting and urban sprawl. *Transportation Research Part D: Transport and Environment*, 14(7): 453-460.

Rowe, F., González-Leonardo, M., Champion, T, (2023). Virtual Special Issue: Internal Migration in Times of COVID-19. *Population, Space and Place*, e2652.

Safirova, E., (2002). Telecommuting, traffic congestion, and agglomeration: a general equilibrium model. *Journal of Urban Economics*, 52(1): 26-52.

Sridhar, K. S., Sridhar, V. (2003). The effect of telecommuting on suburbanisation: Empirical evidence. *Journal of Regional Analysis and Policy*, 33(1): 1–25.

Stiles, J., Smart, M. J. (2020). Working at home and elsewhere: daily work location, telework, and travel among United States knowledge workers. *Transportation*. Springer US.

Su, Y. (2020). Measuring the Value of Urban Consumption Amenities: A Time-Use Approach. *SSRN Electronic Journal*, 1–49.

Tayyaran, M. R., Khan, A. M., Anderson, D. A., (2003). Impact of telecommuting and intelligent transportation systems on residential location choice. *Transportation Planning and Technology*, 26(2): 171-193

Thériault, M., Villeneuve, P., Vandersmissen, M., Rosiers, F. Des. (2005). Homeworking, Telecommuting and Journey to Workplaces: Are Differences among Genders and Professions Varying in Space? *ERSA Conference Papers*, European Regional Science Association, 1–15.

van Wee, B., Geurs, K., Chorus, C. (2013). Information, communication, travel behavior and accessibility. *Journal of Transport and Land Use*, 6(3): 1–16.

Vogiazides, L., Kawalerowicz, J., (2023). Internal Migration in the Time of Covid:

Who Moves out of the Inner City of Stockholm and Where Do They Go? *Population, Space and Place*, 29 (4): e41

Walls, M., Safirova, E. (2004). A Review of the Literature on Telecommuting and Its Implications for Vehicle Travel and Emissions A Review of the Literature on Telecommuting and Its Implications for Vehicle Travel and Emissions. *Resources for the Future*, Discussion Paper 04(44): 1-23.

Whitaker, S.D., (2021). Did the COVID-19 Pandemic Cause An Urban Exodus? *District Data Briefs*, available at https://doi.org/10.26509/frbc-ddb-20210205.

Wulff P.S., Vernon, V., (2020). Telework and Time Use in the United States. *IZA Discussion Paper Series*, 13260.

Welz C., Wolf, F., (2010). Telework in the European Union. Eurofound.

Zhu, P., (2012). Are telecommuting and personal travel complements or substitutes? *The Annals of Regional Science*, Springer. Western Regional Science Association, 48(2): 619-639.

Zhu, P. Mason, S. G., (2014). The impact of telecommuting on personal vehicle usage and environmental sustainability. *International Journal of Environmental Science and Technology*, 11: 2185–2200.

Appendix : Behrens *et al.* 2017 model

Our reference model is based on a multiple monocentric-city structure, with endogenous workers' location choices (and city sizes and their distribution), productivity, and markups. Individuals are identical: they express love for variety, and consume a quantity $q(i)$ of each variety i of a differentiated good, and one unit of homogeneous land which has a distributed ownership (used as numeraire). The maximization of utility in Eq.2.10, entails the first order conditions (Eq.2.11) for the demand of variety i produced in city r and consumed in city r or s, and the indirect utility in city r, (Eq.2.12).

$$\max_{q_{sr}(j), j\in\Omega_{sr}} U_r \equiv \sum_s \int_{\Omega_{s,r}} [1 - e^{\alpha q_{sr}(j)}]dj, \text{ s.t. } \sum_s \int_{\Omega_{s,r}} p_{sr}(j) q_{sr}(j) dj = E_r \tag{2.10}$$

$$q_{rr} = \frac{1}{\alpha} ln \left[\frac{p_r^d}{p_{rr}(i)}\right] \quad ; \quad q_{sr} = \frac{1}{\alpha} ln \left[\frac{p_r^d}{p_{sr}(i)}\right] \tag{2.11}$$

$$V_r = N_r^c - \sum_s \int_{\Omega_{sr}} \frac{p_{sr}(j)}{p_r^d} dj = N_r^c (1 - \frac{\overline{p}_r)}{p_r^d} \tag{2.12}$$

Firms are heterogeneous in productivity, but discover their marginal labour requirement $m_r(i) \geq 0$, drawn form a city-specific continuously differentiable distribution G_r, only after making the irreversible entry decision in the city. So that, after incurring in a fixed entry cost F expressed in units of labour paid at the local (r) market wage w_r, firms are selected. Firms will survive and produce, provided they can charge prices $p_{rs}(i)$ above marginal costs $\tau_{rs} m_r w_r$ in at least one city (or, locally, $p_{rr}(i) \geq \tau_{rr} m_r w_r$).

Profits'(Eq.2.13) maximization F.O.C. (Eq.2.14), interacted with the ones from utility maximization (Eq.2.11), provides the internal (Eq.2.16) and external (Eq.2.15) cutoffs.

$$\pi_r(i) = \sum_s \pi_{rs}(i) = \sum_s L_s q_{rs}(i)[p_{rs}(i) - \tau_{rs} m_r w_r] \tag{2.13}$$

$$ln \frac{p_s^d}{p_{rs}(i)} = \frac{p_{rs}(i) - \tau_{rs} m_r w_r}{p_{rs}(i)} \tag{2.14}$$

$$m_{rs} = \frac{p_s^d}{\tau_{rs} w_r} \tag{2.15}$$

$$m_s^d = \frac{p_s^d}{\tau_{ss} w_s} \tag{2.16}$$

The mass of varieties consumed in city r is then $N_r^p = N_r^E G_r(\max\ m_{rs})$, which is the sum of all firms that are productive enough to sell to market r.

In order to estimate the spatial equilibrium, all firm-level variables needs to be transformed in function of the m (only difference within firms in a city). The Lambert function properties are exploited to this end, by setting: $\psi = e \frac{m}{m_{rs}}$ and $W = \frac{\tau_{rs} m w_r}{p_{rs}(m)}$, and obtaining the following:

$$q_{rs} = \frac{1}{\alpha}(1 - W) \tag{2.17}$$

$$p_{rs} = \frac{\tau_{rs} m w_r}{W} \tag{2.18}$$

$$\pi_{rs}(m) = L_s \frac{\tau_{rs} m w_r}{\alpha}(W^{-1} + W - 2) \tag{2.19}$$

Equations 2.17), 2.18) and 2.19, imply that the markup for a firm located in city r selling in s is:

$$\Lambda_{rs}(m) \equiv \frac{p_{rs}(m)}{\tau_{rs} m_r} = \frac{1}{W} \tag{2.20}$$

We can now characterise the urban structure, composed by monocentric cities of size L_r, with radius $\overline{x}_r = \sqrt{L_r/\pi}$, where land is used only for housing (firms in the CBD), and with urban frictions (commuting, iceberg-type costs θ_r) and trade-frictions $\tau_{rs} > 1$. The total labour supply in city r is thus simply:

$$S_r = \int_0^{\overline{x}_r} 2\pi x_{rsr}(x_r) dx_r \tag{2.21}$$

The equilibrium land rent R_r, is obtained by equating wages net of commuting costs in every location (given homogeneous labour), as in (Eq.2.22), such that the aggregate land rent is found in (Eq.2.23).

$$R_r^*(x_r) = w_r(e^{-\theta_r x_r} - e^{-\theta_r \overline{x}_r})\overline{h}_r \tag{2.22}$$

$$ALR_r = \int_0 \overline{x}_r 2\pi x_r R_r^*(r) dx_r = \frac{2\pi w_r \overline{h}_r}{\theta_r^2}[1 - (1 + \theta_r \overline{x} + \frac{\theta_r^2 L_2}{2\pi})e^{-\theta_r \overline{x}_r}] \tag{2.23}$$

Given distributed land and firm ownership, the per-capita expenditure is thus the sum of equilibrium net wages, the individual share of ALR and of firms' profits, as in Eq.2.24.

$$E_r = w_r \overline{h}_r e^{-\theta_r \overline{x}_r} + ALR_r/L_r + \Pi_r/L_r = w_r h_r + \Pi_r/L_r \tag{2.24}$$

Market equilibrium in the urban system

Choosing a Pareto distribution for firms productivity draws $(G_r(m) = m/m^{max})^k$, with upper bound $m^{max} > 0$ and shape parameter $k \geq 1$, allows to calculate the model equilibrium conditions, which in the multiple-cities case are the zero expected profits, labour market clearing and trade balance (total value of exports must equate the total value of imports per each city).

$$N_r^E \left[\frac{k_1}{\alpha (m_r^{max})^k} \mu_r^{max} + F \right] = S_r \tag{2.25}$$

$$\frac{N_r^E w_r}{(m_r^{max})^k} \sum_{\forall s \neq r} L_s \tau_{rs} \left(\frac{\tau_s w_s}{\tau_{rs} w_r} m_s^d \right)^{k+1} = L_r \sum_{\forall s \neq r} \tau_{sr} \left(\frac{\tau_{rr} w_r}{\tau_{sr} w_s} m_r^d \right)^{k+1} \frac{N_s^E w_s}{(m_s^{max})^k} \tag{2.26}$$

$$\mu_r^{max} = \sum_s L_s \tau_{rs} \left(\frac{\tau_s w_s}{\tau_{rs} w_r} m_s^d \right)^{k+1} \tag{2.27}$$

To reduce the number of unknowns (w_r, N_r^E, n_r^d), equations 2.25 and 2.26) are combined to obtain $N_r^E = \frac{k_2}{k_1+k_2} \frac{S_r}{F}$ as a function of parameters, and the following equation:

$$\frac{h_r}{(m_r^d)^{k+1}} = \sum_s S_s \tau_{rr} \left(\frac{\tau_{rr} w_r}{\tau_{sr} w_s} \right)^k \frac{1}{(\mu_s^{max})} \tag{2.28}$$

Eqs. 2.28 and 2.27 thus constitutes our two equilibrium conditions. Since these are function of only two unknowns, the two vectors of the city-specific equilibrium wages $\hat{w}_r$ and technological frontiers $\hat{\mu}_r$, the model is exactly identified, and can be brought to data, reversed, and exploited for counterfactual estimation.

Spatial Equilibrium

Once solved the static equilibrium, one can turn to the spatial equilibrium, *i.e.* where workers are allowed to endogenously relocate. To do so, city-specific amenities (A_r) and taste heterogeneity (ξ_r^l, i.i.d. across individuals and cities according to a double distribution $[0, \pi^2\beta^2/\sigma]$ are introduced. The location choice of individuals is assumed to follow a *linear random utility* (Eq.2.29), where A_r^o are observed amenities (climate, topography, water area) and A_u, such that the probability of choosing city r can then be expressed in a logistic form (Eq.2.30).

$$V_r^l = U_r + A_r(A_r^o, A_r^u) + \xi_r^l \tag{2.29}$$

$$P_r(V_r^l > \max_{s \neq r} V_s^l) = \frac{exp((U_r + A_r)/\beta)}{\sum_{s=1}^{K} exp((U_s + A_s)/\beta)} \tag{2.30}$$

The city size distribution satisfying $P_r(V_r^l > \max_{s \neq r} V_s^l) = \frac{L_r}{\sum_{s=1}^{K} L_s}, \forall r$ is defined as the spatial equilibrium of the model, which can be easily solved for $\hat{D}_r = U_r + A_r$ (by estimating it by simple OLS as $\hat{D}_r = \alpha_0 + \alpha_1 \hat{U}_r + \alpha_2 A_r^o + \epsilon_r$).

Table 2.1. Descriptive baseline statistics (2017) and main long-term counterfactual results, by Metropolitan Statistical Area.

	Rank		Employment (L_r)		Commuting, hours		Commuting tech (θ_r)		Markups (Λ_r)		Utility (Ur)	GDP, th.$	N. firms	FHW, %	Amenities	
	Baseline	Change	Baseline	Change	Baseline	Change	Baseline	Change	Baseline	Change	Change	Baseline	Baseline	Baseline	A_o	A_u
Mean		*-0.26*	*332,096.1*	*-1.00%*	*2.1*	*-25.56%*	*0.3*	*-13.03%*	*0.06*	*-1.34%*	*1.36%*	*46,736.7*	*21164.946*	*25.56%*	*0.7893*	*0.0007*
MSA																
New York-Newark-Jersey City, NY-NJ-PA	1	0	9,302,700	2.3%	3.11	-38.3%	0.04	-21.42%	0.041	-3.0%	3.07%	1,717,712	628,451	38.31%	0.63	3.77
Los Angeles-Long Beach-Anaheim, CA	2	0	6,047,050	1.0%	2.58	-34.6%	0.04	-18.75%	0.044	-2.3%	2.33%	1,043,735	599,544	34.59%	10.07	3.03
Chicago-Naperville-Elgin, IL-IN-WI	3	0	4,589,680	0.3%	2.55	-34.0%	0.05	-18.17%	0.052	-2.2%	2.25%	679,699	249,637	34.02%	-1.89	3.48
Dallas-Fort Worth-Arlington, TX	4	0	3,485,190	0.2%	2.36	-34.5%	0.05	-18.29%	0.049	-2.0%	2.04%	535,499	177,518	34.53%	0.76	3.03
Washington-Arlington-Alexandria, DC-VA-MD-WV	5	0	3,103,530	2.3%	2.93	-43.1%	0.07	-23.83%	0.044	-3.1%	3.23%	529,991	193,471	43.14%	-0.58	2.81
Houston-The Woodlands-Sugar Land, TX	6	0	2,929,400	0.7%	2.59	-33.4%	0.06	-17.99%	0.045	-2.1%	2.18%	490,074	155,503	33.44%	0.75	2.73
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	7	0	2,813,470	0.7%	2.48	-35.7%	0.06	-19.14%	0.047	-2.2%	2.29%	444,975	163,089	35.67%	-0.67	2.81
Boston-Cambridge-Newton, MA-NH	8	0	2,726,490	0.9%	2.34	-39.4%	0.06	-21.15%	0.045	-2.3%	2.40%	438,684	165,192	39.43%	0.14	2.70
Atlanta-Sandy Springs-Roswell, GA	9	0	2,619,440	0.6%	2.73	-35.7%	0.07	-19.23%	0.051	-2.4%	2.47%	385,542	152,638	35.74%	0.25	2.83
Miami-Fort Lauderdale-West Palm Beach, FL	10	0	2,561,380	0.0%	2.67	-30.8%	0.07	-16.38%	0.055	-2.1%	2.12%	344,882	221,312	30.81%	5.08	2.70
San Francisco-Oakland-Hayward, CA	11	0	2,369,450	2.7%	2.67	-38.6%	0.07	-21.55%	0.034	-2.6%	2.70%	500,710	194,306	38.61%	7.30	1.78
Phoenix-Mesa-Scottsdale, AZ	12	0	1,980,010	-0.6%	2.29	-33.6%	0.07	-17.49%	0.065	-2.0%	2.00%	242,951	99,521	33.56%	4.31	2.66
Detroit-Warren-Dearborn, MI	13	1	1,966,680	-0.6%	2.15	-30.3%	0.06	-15.75%	0.056	-1.7%	1.70%	260,612	97,125	30.31%	-1.67	2.72
Seattle-Tacoma-Bellevue, WA	14	-1	1,944,160	1.2%	2.49	-36.3%	0.07	-19.67%	0.041	-2.3%	2.33%	356,572	130,480	36.27%	4.61	2.02
Minneapolis-St. Paul-Bloomington, MN-WI	15	0	1,932,310	-0.4%	2.02	-36.1%	0.06	-18.81%	0.057	-1.9%	1.90%	260,106	91,813	36.10%	-2.22	2.76
Denver-Aurora-Lakewood, CO	16	0	1,443,130	0.2%	2.32	-36.7%	0.08	-19.39%	0.053	-2.1%	2.17%	208,868	101,685	36.66%	4.20	2.10
Riverside-San Bernardino-Ontario, CA	17	1	1,435,200	-0.6%	3.43	-25.3%	0.12	-13.44%	0.076	-2.2%	2.26%	157,931	120,427	25.25%	6.64	2.39
San Diego-Carlsbad, CA	18	-1	1,433,340	0.5%	2.35	-34.5%	0.08	-18.45%	0.046	-2.1%	2.12%	231,845	109,009	34.55%	9.78	1.69
Baltimore-Columbia-Towson, MD	19	0	1,360,320	0.6%	2.58	-36.5%	0.09	-19.57%	0.050	-2.4%	2.42%	192,178	72,441	36.48%	-0.40	2.17
St. Louis, MO-IL	20	0	1,356,630	-0.8%	2.06	-33.3%	0.07	-17.21%	0.065	-1.7%	1.78%	161,281	85,574	33.28%	-0.55	2.47
Tampa-St. Petersburg-Clearwater, FL	21	0	1,280,170	-0.5%	2.36	-32.7%	0.09	-17.10%	0.062	-2.0%	1.99%	146,349	88,849	32.71%	4.00	2.18
Orlando-Kissimmee-Sanford, FL	22	0	1,209,250	-0.9%	2.29	-29.1%	0.09	-15.02%	0.066	-1.7%	1.72%	132,448	71,416	29.08%	3.68	2.21
Charlotte-Concord-Gastonia, NC-SC	23	0	1,186,840	-0.1%	2.21	-33.6%	0.08	-17.68%	0.050	-1.9%	1.90%	174,029	69,071	33.60%	0.29	2.01
Portland-Vancouver-Hillsboro, OR-WA	24	0	1,157,060	0.0%	2.30	-34.0%	0.09	-17.94%	0.052	-2.0%	2.04%	171,772	87,695	33.95%	2.81	1.92
Pittsburgh, PA	25	0	1,132,950	-0.5%	2.17	-32.6%	0.08	-17.01%	0.057	-1.8%	1.83%	147,368	63,027	32.63%	0.40	2.12
San Jose-Sunnyvale-Santa Clara, CA	26	0	1,089,070	3.4%	2.18	-44.6%	0.09	-24.68%	0.029	-2.5%	2.54%	275,294	73,196	44.61%	5.56	0.71
Cincinnati, OH-KY-IN	27	0	1,056,680	-0.6%	2.01	-32.3%	0.08	-16.76%	0.056	-1.7%	1.68%	138,034	54,180	32.31%	-0.76	2.08
Kansas City, MO-KS	28	0	1,055,320	-0.7%	1.86	-35.4%	0.08	-18.29%	0.061	-1.7%	1.70%	131,092	64,209	35.37%	-1.27	2.19
Columbus, OH	29	0	1,038,240	-0.6%	1.89	-34.1%	0.08	-17.68%	0.057	-1.7%	1.68%	136,296	48,288	34.11%	-1.63	2.11
Indianapolis-Carmel-Anderson, IN	30	1	1,029,390	-0.7%	1.91	-30.4%	0.08	-15.75%	0.053	-1.5%	1.50%	143,874	48,098	30.44%	-2.63	2.05
Cleveland-Elyria, OH	31	-1	1,029,230	-0.6%	1.85	-32.2%	0.08	-16.71%	0.053	-1.5%	1.56%	138,980	55,871	32.22%	-1.43	1.99
San Antonio-New Braunfels, TX	32	1	1,003,370	-0.3%	2.38	-32.5%	0.10	-17.02%	0.057	-1.9%	1.94%	129,298	51,348	32.48%	2.13	1.93
Austin-Round Rock, TX	33	-1	996,540	0.7%	2.30	-40.9%	0.09	-21.77%	0.048	-2.3%	2.36%	148,750	57,775	40.88%	1.62	1.71
Las Vegas-Henderson-Paradise, NV	34	1	962,720	-1.2%	2.08	-25.8%	0.09	-13.11%	0.065	-1.4%	1.38%	112,288	53,942	25.77%	4.86	1.92
Sacramento--Roseville--Arden-Arcade, CA	35	-1	960,180	0.0%	2.33	-36.4%	0.10	-19.20%	0.057	-2.2%	2.21%	126,352	81,972	36.39%	5.41	1.75
Nashville-Davidson--Murfreesboro--Franklin, TN	36	0	940,810	-0.3%	2.27	-30.8%	0.10	-16.15%	0.053	-1.8%	1.79%	133,251	48,899	30.83%	-0.82	1.88
Milwaukee-Waukesha-West Allis, WI	37	0	841,550	-0.9%	1.71	-32.1%	0.08	-16.58%	0.055	-1.4%	1.43%	105,427	45,225	32.08%	-1.70	1.86
Virginia Beach-Norfolk-Newport News, VA-NC	38	0	743,960	-0.6%	2.27	-29.0%	0.11	-15.08%	0.055	-1.6%	1.67%	94,855	41,536	28.98%	0.42	1.65
Salt Lake City, UT	39	0	695,050	-1.0%	1.56	-36.4%	0.08	-18.81%	0.063	-1.5%	1.53%	87,802	45,349	36.38%	2.96	1.63
Jacksonville, FL	40	0	668,140	-0.8%	2.19	-30.1%	0.11	-15.57%	0.063	-1.7%	1.70%	76,650	43,325	30.13%	2.02	1.63
Louisville/Jefferson County, KY-IN	41	1	646,670	-1.2%	1.84	-27.1%	0.10	-13.84%	0.062	-1.3%	1.29%	76,064	35,559	27.15%	-0.98	1.70
Richmond, VA	42	-1	643,860	-0.6%	2.02	-33.4%	0.10	-17.37%	0.057	-1.7%	1.71%	82,739	39,320	33.45%	-1.09	1.60
Memphis, TN-MS-AR	43	0	617,990	-1.2%	1.93	-27.2%	0.10	-13.86%	0.062	-1.3%	1.33%	72,503	27,088	27.17%	-0.79	1.66
Raleigh, NC	44	0	606,510	0.1%	2.26	-36.1%	0.12	-19.06%	0.050	-2.0%	2.07%	83,288	38,793	36.08%	-0.67	1.38
Oklahoma City, OK	45	0	603,780	-0.7%	2.02	-33.7%	0.11	-17.47%	0.059	-1.7%	1.73%	74,884	40,430	33.70%	0.12	1.54
Hartford-West Hartford-East Hartford, CT	46	0	581,750	0.3%	1.94	-38.0%	0.11	-20.06%	0.045	-1.9%	1.93%	90,318	36,342	37.98%	1.48	1.09
Providence-Warwick, RI-MA	47	0	567,620	0.9%	2.97	-32.2%	0.16	-17.60%	0.048	-2.4%	2.51%	82,929	50,743	32.18%	1.28	1.17
New Orleans-Metairie, LA	48	0	552,840	-0.7%	2.15	-25.9%	0.12	-13.41%	0.050	-1.4%	1.41%	79,290	38,844	25.92%	0.27	1.22
Grand Rapids-Wyoming, MI	49	0	551,620	-1.4%	1.66	-26.6%	0.09	-13.53%	0.066	-1.1%	1.15%	60,529	21,987	26.59%	-1.83	1.64
Buffalo-Cheektowaga-Niagara Falls, NY	50	0	547,750	-1.0%	1.77	-33.3%	0.10	-17.17%	0.062	-1.5%	1.50%	60,014	29,440	33.31%	-0.64	1.52
Rochester, NY	51	0	510,010	-1.0%	1.75	-34.9%	0.10	-17.97%	0.064	-1.5%	1.55%	56,550	26,850	34.88%	-0.63	1.49
Birmingham-Hoover, AL	52	0	504,290	-0.7%	2.15	-30.5%	0.13	-15.81%	0.058	-1.7%	1.69%	64,553	29,144	30.48%	0.58	1.32
Omaha-Council Bluffs, NE-IA	53	0	486,650	-0.9%	1.62	-34.1%	0.10	-17.67%	0.055	-1.4%	1.43%	65,053	28,286	34.10%	-1.87	1.33
Albany-Schenectady-Troy, NY	54	0	448,160	-0.6%	1.83	-37.8%	0.11	-19.64%	0.058	-1.7%	1.76%	54,302	23,672	37.82%	-0.24	1.24
Tulsa, OK	55	1	427,880	-0.9%	1.82	-30.4%	0.11	-15.68%	0.055	-1.4%	1.41%	57,747	28,234	30.42%	0.41	1.10
Bridgeport-Stamford-Norwalk, CT	56	-1	415,670	4.3%	2.80	-39.2%	0.18	-22.52%	0.028	-2.8%	2.88%	98,256	35,311	39.18%	2.25	-0.18
Greenville-Anderson-Mauldin, SC	57	0	401,590	-1.3%	1.93	-26.9%	0.13	-13.68%	0.068	-1.3%	1.34%	42,109	20,453	26.94%	1.18	1.23
Baton Rouge, LA	58	0	392,000	-0.8%	2.13	-25.3%	0.14	-13.04%	0.051	-1.3%	1.37%	54,988	23,204	25.34%	-0.75	0.97
Madison, WI	59	0	387,300	-0.9%	1.63	-33.8%	0.11	-17.44%	0.057	-1.4%	1.43%	49,853	18,705	33.75%	-0.82	1.08
Albuquerque, NM	60	0	381,200	-1.1%	1.99	-29.9%	0.14	-15.33%	0.068	-1.5%	1.56%	44,051	22,104	29.94%	3.73	1.08
Knoxville, TN	61	0	380,260	-1.0%	2.06	-29.5%	0.14	-15.13%	0.066	-1.5%	1.56%	41,459	19,283	29.52%	1.00	1.16
Fresno, CA	62	0	372,770	-1.2%	1.99	-27.5%	0.14	-14.01%	0.067	-1.4%	1.43%	42,045	34,853	27.51%	6.03	0.95
Dayton, OH	63	0	371,610	-1.1%	1.73	-30.2%	0.12	-15.46%	0.062	-1.4%	1.37%	41,111	17,724	30.20%	-2.18	1.20
Columbia, SC	64	0	370,160	-1.0%	1.99	-31.6%	0.14	-16.24%	0.066	-1.6%	1.62%	40,884	18,774	31.61%	0.50	1.14
Tucson, AZ	65	1	364,930	-0.9%	2.33	-30.2%	0.16	-15.59%	0.073	-1.8%	1.82%	39,034	18,633	30.21%	4.04	1.08
Greensboro-High Point, NC	66	1	363,510	-1.3%	1.78	-26.4%	0.12	-13.42%	0.060	-1.2%	1.19%	41,543	18,445	26.39%	-0.25	1.07
Des Moines-West Des Moines, IA	67	-2	363,420	-0.5%	1.55	-36.1%	0.11	-18.87%	0.046	-1.4%	1.45%	57,152	21,686	36.11%	-2.03	0.79
Allentown-Bethlehem-Easton, PA-NJ	68	0	358,910	-0.5%	2.45	-28.4%	0.17	-14.88%	0.056	-1.8%	1.80%	43,821	19,476	28.41%	0.30	0.95
Little Rock-North Little Rock-Conway, AR	69	0	347,590	-1.1%	1.83	-31.5%	0.13	-16.16%	0.065	-1.4%	1.47%	38,874	20,966	31.54%	-0.07	1.09
Charleston-North Charleston, SC	70	0	336,560	-0.6%	2.33	-28.3%	0.17	-14.72%	0.057	-1.7%	1.70%	42,004	20,317	28.30%	0.57	0.89
Akron, OH	72	-1	328,230	-0.8%	1.94	-31.9%	0.14	-16.44%	0.061	-1.6%	1.62%	36,518	17,641	31.88%	-2.28	1.05
Harrisburg-Carlisle, PA	73	0	323,720	-1.0%	1.63	-33.2%	0.12	-17.10%	0.060	-1.4%	1.40%	37,182	14,853	33.24%	0.00	0.94
Oxnard-Thousand Oaks-Ventura, CA	74	0	309,860	1.0%	2.89	-30.7%	0.21	-16.81%	0.043	-2.3%	2.31%	50,848	26,664	30.67%	11.17	0.00

MSA	Rank		Employment (L_r)		Commuting, hours		Commuting tech (θ_r)		Markups (Λ_r)		Utility (Ur)	GDP, th.$	N. firms	FHW, %	Amenities	
	Baseline	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Change*	*Baseline*	*Baseline*	*Baseline*	A_o	A_u
Boise City, ID	75	0	308,170	-1.2%	1.92	-31.2%	0.15	-15.93%	0.074	-1.6%	1.59%	33,604	21,710	31.20%	2.29	1.00
Bakersfield, CA	76	0	303,620	-1.2%	2.11	-23.5%	0.16	-11.91%	0.063	-1.3%	1.30%	37,340	18,603	23.55%	4.84	0.73
Syracuse, NY	77	1	302,070	-1.1%	1.68	-31.8%	0.13	-16.27%	0.063	-1.3%	1.35%	33,634	16,791	31.76%	-1.09	0.96
El Paso, TX	78	1	301,590	-1.1%	2.32	-29.5%	0.17	-15.12%	0.076	-1.7%	1.71%	29,033	15,232	29.50%	4.08	0.92
Durham-Chapel Hill, NC	79	-2	298,540	0.0%	1.86	-39.4%	0.14	-20.64%	0.047	-1.8%	1.86%	43,474	14,492	39.36%	0.10	0.54
Toledo, OH	80	0	296,990	-1.4%	1.59	-25.5%	0.12	-12.92%	0.061	-1.1%	1.06%	33,694	14,264	25.50%	-2.46	0.97
Wichita, KS	81	0	296,330	-1.4%	1.60	-28.3%	0.12	-14.42%	0.065	-1.1%	1.16%	33,840	15,547	28.31%	-0.56	0.95
North Port-Sarasota-Bradenton, FL	82	1	289,580	-1.2%	2.18	-24.2%	0.17	-12.27%	0.063	-1.3%	1.36%	31,197	26,443	24.16%	4.71	0.69
Worcester, MA-CT	83	-1	281,770	2.0%	3.69	-32.5%	0.28	-18.45%	0.046	-3.0%	3.14%	43,750	28,079	32.47%	0.50	0.43
Colorado Springs, CO	84	1	274,870	-0.2%	2.46	-33.5%	0.19	-17.65%	0.059	-2.1%	2.11%	32,683	20,497	33.53%	5.39	0.54
New Haven-Milford, CT	85	-1	273,160	1.8%	3.12	-31.8%	0.25	-17.79%	0.040	-2.5%	2.60%	45,252	24,077	31.76%	2.52	0.10
Lexington-Fayette, KY	86	0	272,410	-1.6%	1.59	-23.4%	0.13	-11.80%	0.062	-1.0%	0.96%	29,960	15,491	23.45%	-2.03	0.89
Winston-Salem, NC	87	0	263,840	-1.0%	2.17	-26.1%	0.17	-13.33%	0.061	-1.4%	1.44%	29,749	14,285	26.09%	-0.26	0.76
Jackson, MS	88	0	262,680	-1.0%	1.94	-30.4%	0.15	-15.55%	0.063	-1.5%	1.49%	30,237	14,484	30.39%	-0.66	0.80
Cape Coral-Fort Myers, FL	89	0	258,370	-0.9%	2.66	-22.6%	0.21	-11.67%	0.061	-1.5%	1.55%	27,953	21,727	22.61%	5.23	0.51
Scranton--Wilkes-Barre--Hazleton, PA	90	0	257,730	-1.6%	1.77	-25.5%	0.15	-12.79%	0.076	-1.2%	1.17%	23,688	13,799	25.52%	0.35	0.92
McAllen-Edinburg-Mission, TX	91	0	254,190	-1.5%	2.16	-23.9%	0.17	-11.95%	0.083	-1.3%	1.30%	20,405	12,365	23.92%	0.46	0.98
Ogden-Clearfield, UT	92	0	249,250	-1.0%	2.25	-29.2%	0.19	-15.01%	0.075	-1.7%	1.77%	26,476	16,053	29.18%	3.45	0.75
Stockton-Lodi, CA	93	0	242,740	-0.4%	3.36	-23.1%	0.28	-12.33%	0.062	-2.0%	2.03%	27,089	17,539	23.07%	4.77	0.49
Chattanooga, TN-GA	94	0	241,810	-1.2%	1.88	-26.0%	0.16	-13.16%	0.062	-1.2%	1.24%	27,071	12,151	26.00%	0.28	0.66
Lancaster, PA	95	0	241,190	-1.0%	1.99	-25.1%	0.17	-12.77%	0.056	-1.3%	1.29%	28,938	13,411	25.08%	0.45	0.54
Fayetteville-Springdale-Rogers, AR-MO	96	0	239,920	-1.0%	1.80	-30.3%	0.15	-15.48%	0.060	-1.4%	1.40%	28,504	13,034	30.29%	1.07	0.60
Spokane-Spokane Valley, WA	97	0	232,920	-1.2%	1.90	-28.5%	0.16	-14.44%	0.068	-1.4%	1.39%	25,498	17,375	28.46%	2.38	0.64
Provo-Orem, UT	98	1	229,480	-0.7%	2.10	-35.1%	0.19	-18.18%	0.070	-1.9%	1.98%	25,625	16,173	35.08%	3.03	0.62
Trenton, NJ	99	-1	229,450	-0.1%	1.78	-38.6%	0.16	-20.19%	0.047	-1.8%	1.79%	29,987	10,945	38.64%	-0.80	0.31
Reno, NV	100	0	224,130	-1.2%	1.86	-26.6%	0.16	-13.48%	0.063	-1.3%	1.28%	26,631	14,801	26.64%	6.70	0.35
Huntsville, AL	101	0	222,080	-0.9%	1.79	-33.2%	0.16	-17.07%	0.059	-1.5%	1.53%	25,793	11,010	33.23%	-0.91	0.57
Augusta-Richmond County, GA-SC	102	0	218,410	-1.1%	2.24	-25.3%	0.20	-12.87%	0.064	-1.4%	1.44%	24,395	11,036	25.26%	-0.10	0.62
Lansing-East Lansing, MI	103	0	215,080	-1.1%	1.84	-31.3%	0.17	-15.94%	0.067	-1.5%	1.50%	22,562	9,060	31.25%	-3.34	0.77
Lakeland-Winter Haven, FL	104	0	214,340	-0.9%	2.88	-23.8%	0.25	-12.38%	0.070	-1.7%	1.77%	21,155	13,081	23.85%	3.98	0.52
Ann Arbor, MI	105	0	213,990	-1.2%	1.65	-30.1%	0.15	-15.31%	0.060	-1.3%	1.29%	23,500	8,225	30.05%	-2.19	0.61
Fort Wayne, IN	106	0	213,660	-1.4%	1.72	-26.2%	0.16	-13.17%	0.065	-1.2%	1.17%	22,358	10,069	26.16%	-3.07	0.74
Youngstown-Warren-Boardman, OH-PA	107	0	213,630	-1.4%	2.10	-24.2%	0.19	-12.18%	0.076	-1.3%	1.32%	19,755	13,118	24.25%	-2.28	0.84
Palm Bay-Melbourne-Titusville, FL	108	1	206,760	-0.7%	2.33	-29.4%	0.21	-15.22%	0.063	-1.7%	1.77%	21,849	15,610	29.38%	3.93	0.39
Springfield, MO	109	2	204,320	-1.4%	1.82	-28.1%	0.17	-14.18%	0.075	-1.3%	1.32%	19,528	12,652	28.15%	-0.10	0.70
Portland-South Portland, ME	110	-2	203,740	0.9%	2.73	-32.8%	0.25	-17.81%	0.045	-2.3%	2.33%	32,143	22,178	32.78%	0.96	0.06
Santa Rosa, CA	111	-1	202,410	0.0%	2.55	-28.4%	0.24	-15.07%	0.049	-1.9%	1.89%	28,673	19,803	28.38%	7.93	-0.10
Lafayette, LA	112	0	196,240	-1.1%	2.26	-24.5%	0.21	-12.43%	0.063	-1.4%	1.40%	22,114	14,932	24.45%	-0.34	0.50
Deltona-Daytona Beach-Ormond Beach, FL	113	0	192,010	-1.1%	2.70	-24.1%	0.25	-12.37%	0.077	-1.7%	1.68%	17,199	16,616	24.11%	3.22	0.52
Asheville, NC	114	0	191,430	-1.5%	1.85	-23.1%	0.17	-11.51%	0.065	-1.1%	1.08%	20,179	14,127	23.10%	2.10	0.41
Corpus Christi, TX	115	0	189,030	-1.3%	1.74	-26.8%	0.16	-13.52%	0.059	-1.2%	1.17%	21,547	10,093	26.75%	2.86	0.27
Davenport-Moline-Rock Island, IA-IL	116	1	181,490	-1.4%	1.53	-28.3%	0.15	-14.29%	0.064	-1.1%	1.13%	20,020	9,886	28.30%	-2.69	0.54
Modesto, CA	117	-1	181,330	-0.4%	3.05	-25.5%	0.30	-13.55%	0.061	-2.0%	2.03%	20,660	15,327	25.49%	7.21	0.08
York-Hanover, PA	118	0	180,080	-0.8%	2.72	-24.3%	0.26	-12.64%	0.062	-1.7%	1.71%	19,383	9,121	24.30%	-0.58	0.40
Boulder, CO	119	0	178,460	-0.1%	1.65	-40.4%	0.16	-21.05%	0.046	-1.7%	1.71%	25,274	15,249	40.43%	5.82	-0.24
Lincoln, NE	120	0	175,310	-1.0%	1.61	-33.3%	0.16	-17.02%	0.060	-1.4%	1.39%	20,246	10,805	33.32%	-2.82	0.43
Shreveport-Bossier City, LA	121	0	174,330	-1.2%	1.88	-24.3%	0.19	-12.24%	0.054	-1.1%	1.16%	22,746	11,533	24.26%	0.48	0.18
Green Bay, WI	122	1	173,030	-1.3%	1.52	-29.4%	0.16	-14.91%	0.062	-1.2%	1.17%	19,522	8,154	29.43%	-1.39	0.40
Reading, PA	123	-1	172,860	-0.9%	2.33	-26.2%	0.23	-13.48%	0.061	-1.6%	1.58%	18,679	8,922	26.19%	-0.73	0.36
Mobile, AL	124	0	172,750	-1.3%	1.95	-24.0%	0.20	-12.06%	0.060	-1.2%	1.21%	19,407	10,134	23.97%	1.52	0.25
Salinas, CA	125	0	171,870	-1.1%	2.02	-21.7%	0.20	-10.94%	0.052	-1.1%	1.15%	24,096	13,563	21.70%	9.24	-0.24
Peoria, IL	126	2	170,630	-1.4%	1.58	-27.5%	0.16	-13.86%	0.064	-1.1%	1.13%	18,983	7,996	27.53%	-2.50	0.47
Savannah, GA	127	-1	170,620	-1.2%	2.14	-23.6%	0.21	-11.90%	0.062	-1.3%	1.28%	18,533	9,457	23.57%	0.75	0.29
Pensacola-Ferry Pass-Brent, FL	128	-1	169,930	-0.9%	2.58	-25.5%	0.26	-13.20%	0.065	-1.7%	1.70%	18,016	11,480	25.52%	2.10	0.29
Tallahassee, FL	129	0	168,720	-1.1%	1.97	-33.3%	0.20	-17.03%	0.074	-1.7%	1.70%	16,164	10,312	33.32%	1.84	0.42
Canton-Massillon, OH	130	0	168,230	-1.3%	1.93	-25.5%	0.20	-12.83%	0.068	-1.3%	1.29%	16,771	9,133	25.50%	-1.48	0.47
Salem, OR	131	0	165,560	-1.0%	2.13	-29.2%	0.22	-14.99%	0.069	-1.6%	1.66%	17,333	12,527	29.23%	3.42	0.26
Montgomery, AL	132	0	164,610	-1.3%	1.79	-28.3%	0.18	-14.29%	0.067	-1.3%	1.30%	17,557	8,814	28.32%	0.46	0.35
Beaumont-Port Arthur, TX	133	0	160,260	-0.8%	1.97	-24.7%	0.20	-12.66%	0.045	-1.2%	1.22%	25,022	8,311	24.73%	1.10	-0.19
Salisbury, MD-DE	134	0	157,980	-1.0%	2.22	-23.3%	0.23	-11.89%	0.056	-1.3%	1.34%	19,660	12,021	23.34%	-0.03	0.15
Myrtle Beach-Conway-North Myrtle Beach, SC-NC	135	0	157,450	-1.4%	2.19	-20.5%	0.23	-10.22%	0.063	-1.1%	1.15%	17,216	11,525	20.51%	0.98	0.23
Fort Collins, CO	136	0	154,510	-1.0%	2.11	-27.5%	0.22	-14.04%	0.063	-1.5%	1.49%	17,164	12,074	27.53%	5.62	0.03
Hickory-Lenoir-Morganton, NC	137	1	153,900	-1.7%	1.92	-20.0%	0.20	-9.76%	0.073	-1.0%	0.98%	14,270	7,793	19.99%	1.51	0.33
Evansville, IN-KY	138	-1	153,740	-1.4%	1.69	-23.9%	0.18	-11.96%	0.059	-1.0%	1.05%	18,001	7,536	23.94%	-1.51	0.23
Roanoke, VA	139	0	152,640	-1.5%	1.82	-24.6%	0.19	-12.29%	0.068	-1.1%	1.14%	15,324	9,376	24.64%	0.97	0.28
Visalia-Porterville, CA	140	0	152,430	-1.5%	1.99	-22.6%	0.21	-11.20%	0.075	-1.2%	1.17%	15,360	10,272	22.56%	5.65	0.17
Eugene, OR	141	0	152,110	-1.4%	1.79	-27.6%	0.20	-13.88%	0.073	-1.3%	1.32%	15,802	12,021	27.61%	4.29	0.20
Sioux Falls, SD	142	0	151,480	-1.3%	1.36	-29.4%	0.15	-14.98%	0.055	-1.0%	1.03%	19,700	9,692	29.44%	-3.24	0.20
Gulfport-Biloxi-Pascagoula, MS	143	0	150,080	-1.0%	2.32	-24.4%	0.24	-12.47%	0.059	-1.4%	1.43%	17,430	7,820	24.39%	0.74	0.11
Spartanburg, SC	144	0	149,130	-1.4%	1.82	-23.2%	0.20	-11.57%	0.063	-1.1%	1.09%	15,990	6,669	23.25%	0.01	0.22
Rockford, IL	145	0	143,470	-1.2%	2.10	-23.6%	0.23	-11.89%	0.062	-1.3%	1.28%	15,333	6,803	23.57%	-2.79	0.27
Lubbock, TX	146	1	142,850	-1.6%	1.50	-27.1%	0.16	-13.53%	0.071	-1.0%	1.02%	13,880	8,035	27.07%	1.64	0.23
Naples-Immokalee-Marco Island, FL	147	-1	142,480	-1.1%	2.13	-22.4%	0.23	-11.34%	0.055	-1.2%	1.23%	18,236	13,779	22.42%	5.00	-0.20
Port St. Lucie, FL	148	0	141,190	-0.8%	2.98	-24.3%	0.32	-12.71%	0.068	-1.8%	1.87%	13,933	12,639	24.30%	5.18	0.03

MSA	Rank		Employment (L_r)		Commuting, hours		Commuting tech (θ_r)		Markups (Λ_r)		Utility (Ur)	GDP, th.$	N. firms	FHW, %	Amenities	
	Baseline	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Change*	*Baseline*	*Baseline*	*Baseline*	A_o	A_u
Cedar Rapids, IA	149	0	140,640	-1.0%	1.58	-30.9%	0.18	-15.75%	0.054	-1.3%	1.27%	18,464	8,091	30.92%	-3.30	0.10
Brownsville-Harlingen, TX	150	3	139,690	-1.7%	1.91	-25.6%	0.21	-12.71%	0.088	-1.2%	1.23%	10,197	6,531	25.61%	2.46	0.35
Kalamazoo-Portage, MI	151	0	139,130	-0.9%	2.05	-25.9%	0.23	-13.22%	0.056	-1.4%	1.38%	17,086	6,357	25.90%	-1.33	0.05
Killeen-Temple, TX	152	-2	138,710	-0.4%	2.51	-28.8%	0.27	-15.08%	0.055	-1.8%	1.81%	17,556	7,145	28.79%	1.56	-0.08
Flint, MI	153	1	138,240	-1.1%	2.50	-23.4%	0.28	-11.99%	0.067	-1.5%	1.53%	13,623	6,851	23.44%	-1.90	0.27
Fargo, ND-MN	154	1	137,890	-1.4%	1.35	-30.0%	0.16	-15.18%	0.061	-1.0%	1.06%	16,279	8,530	30.01%	-4.59	0.29
Vallejo-Fairfield, CA	155	-3	136,020	1.1%	3.94	-22.9%	0.43	-13.06%	0.043	-2.3%	2.36%	21,422	11,263	22.90%	5.88	-0.63
South Bend-Mishawaka, IN-MI	156	0	134,230	-1.2%	1.92	-27.4%	0.22	-13.87%	0.063	-1.3%	1.36%	14,225	6,453	27.39%	-2.32	0.21
Gainesville, FL	157	0	131,830	-1.4%	1.82	-25.8%	0.21	-12.93%	0.069	-1.2%	1.22%	12,900	7,406	25.85%	2.09	0.10
Huntington-Ashland, WV-KY-OH	158	0	130,770	-1.3%	1.96	-23.4%	0.23	-11.75%	0.062	-1.2%	1.20%	15,218	8,239	23.45%	-0.36	0.07
Norwich-New London, CT	159	1	127,780	-1.2%	1.93	-22.9%	0.23	-11.50%	0.054	-1.1%	1.16%	15,754	7,470	22.86%	2.43	-0.22
Fayetteville, NC	160	-1	127,420	-0.3%	2.51	-25.9%	0.29	-13.58%	0.049	-1.6%	1.65%	17,263	6,707	25.86%	-0.92	-0.22
Elkhart-Goshen, IN	161	0	127,200	-1.8%	1.16	-19.0%	0.14	-9.62%	0.047	-0.6%	0.57%	17,132	4,682	18.97%	-2.72	-0.18
Duluth, MN-WI	162	0	126,020	-1.5%	1.83	-25.0%	0.22	-12.46%	0.074	-1.2%	1.20%	13,318	7,089	25.01%	-0.81	0.23
Atlantic City-Hammonton, NJ	163	0	125,080	-1.2%	1.98	-25.7%	0.23	-13.01%	0.062	-1.3%	1.32%	13,122	6,474	25.74%	-0.04	0.02
Utica-Rome, NY	164	1	124,540	-1.6%	1.60	-29.2%	0.19	-14.60%	0.079	-1.2%	1.19%	11,071	6,555	29.22%	-1.62	0.31
Burlington-South Burlington, VT	165	-1	124,030	-1.1%	1.66	-31.4%	0.20	-15.94%	0.061	-1.4%	1.38%	14,140	8,365	31.39%	-0.12	0.00
Wilmington, NC	166	0	123,730	-1.1%	1.95	-25.5%	0.23	-12.88%	0.057	-1.2%	1.26%	14,350	9,293	25.47%	0.65	-0.11
Erie, PA	167	1	123,710	-1.7%	1.55	-23.4%	0.19	-11.50%	0.073	-0.9%	0.94%	11,146	6,949	23.39%	-0.57	0.20
Appleton, WI	168	-1	123,460	-1.4%	1.67	-25.9%	0.20	-12.92%	0.063	-1.1%	1.12%	13,120	6,020	25.86%	-2.73	0.14
Kingsport-Bristol-Bristol, TN-VA	169	1	116,510	-1.5%	2.02	-21.4%	0.24	-10.58%	0.067	-1.1%	1.10%	11,967	6,711	21.39%	0.37	0.02
Waco, TX	170	1	116,260	-1.4%	1.68	-26.0%	0.20	-12.96%	0.064	-1.1%	1.09%	12,224	5,563	26.00%	0.34	-0.03
Rochester, MN	171	2	114,770	-1.8%	1.51	-20.1%	0.19	-9.73%	0.066	-0.8%	0.80%	12,568	5,071	20.07%	-3.20	0.13
Amarillo, TX	172	0	114,490	-1.3%	1.72	-26.3%	0.21	-13.15%	0.061	-1.1%	1.13%	13,695	6,722	26.26%	1.50	-0.16
Columbus, GA-AL	173	1	113,410	-1.1%	1.85	-27.6%	0.23	-13.99%	0.057	-1.3%	1.31%	13,821	6,065	27.61%	-0.22	-0.17
Manchester-Nashua, NH	174	-5	112,670	7.2%	4.41	-33.3%	0.52	-21.12%	0.026	-3.7%	3.79%	28,443	12,132	33.27%	0.07	-1.54
Lake Charles, LA	175	2	111,890	-1.7%	1.42	-16.2%	0.18	-7.82%	0.045	-0.6%	0.58%	17,319	5,454	16.24%	0.12	-0.50
College Station-Bryan, TX	176	2	111,780	-1.7%	1.55	-25.5%	0.19	-12.57%	0.075	-1.0%	0.99%	10,198	5,357	25.45%	0.86	0.05
Kennewick-Richland, WA	177	-1	111,340	-1.1%	1.92	-25.4%	0.24	-12.80%	0.058	-1.2%	1.25%	13,868	8,477	25.35%	0.75	-0.21
Charleston, WV	178	2	111,090	-1.4%	1.42	-29.0%	0.18	-14.53%	0.061	-1.1%	1.08%	12,817	6,273	28.98%	-0.71	-0.08
Charlottesville, VA	179	-4	111,050	-0.9%	2.01	-29.4%	0.25	-15.03%	0.057	-1.5%	1.50%	13,414	8,206	29.37%	-0.11	-0.18
Olympia-Tumwater, WA	180	-1	109,400	0.3%	2.64	-36.9%	0.33	-19.72%	0.058	-2.5%	2.52%	12,426	8,359	36.95%	3.32	-0.33
Topeka, KS	181	0	109,400	-1.4%	1.57	-31.3%	0.20	-15.69%	0.074	-1.2%	1.26%	10,655	6,259	31.27%	-1.21	0.10
Fort Smith, AR-OK	182	1	107,840	-1.6%	1.83	-23.2%	0.23	-11.44%	0.077	-1.1%	1.08%	10,061	6,554	23.23%	1.55	0.01
Crestview-Fort Walton Beach-Destin, FL	183	-1	106,850	-0.6%	2.34	-25.8%	0.30	-13.39%	0.052	-1.5%	1.56%	14,404	8,868	25.80%	2.10	-0.43
Springfield, IL	184	0	106,710	-1.4%	1.60	-29.5%	0.21	-14.75%	0.071	-1.2%	1.22%	10,250	4,963	29.46%	-2.62	0.10
St. Cloud, MN	185	0	104,890	-1.5%	1.74	-25.3%	0.23	-12.54%	0.072	-1.1%	1.15%	10,226	5,250	25.26%	-3.00	0.12
Barnstable Town, MA	186	0	103,120	-1.1%	1.88	-26.3%	0.25	-13.33%	0.057	-1.3%	1.30%	11,501	9,468	26.34%	1.52	-0.33
Hagerstown-Martinsburg, MD-WV	187	0	102,640	-0.8%	2.85	-23.6%	0.37	-12.28%	0.067	-1.7%	1.75%	10,232	5,518	23.60%	0.27	-0.12
Tyler, TX	188	0	100,780	-0.7%	1.89	-27.0%	0.24	-13.84%	0.045	-1.3%	1.28%	14,410	6,212	27.03%	0.72	-0.62
Binghamton, NY	189	0	100,550	-1.4%	1.71	-29.0%	0.22	-14.57%	0.070	-1.2%	1.26%	9,626	5,378	29.05%	-0.93	-0.04
Ocala, FL	190	1	100,260	-1.7%	2.37	-17.1%	0.31	-8.32%	0.080	-1.0%	1.05%	8,654	8,231	17.15%	2.59	-0.07
Tuscaloosa, AL	191	-1	100,100	-1.5%	1.94	-19.7%	0.25	-9.66%	0.061	-1.0%	0.98%	11,716	5,177	19.69%	0.60	-0.24
Champaign-Urbana, IL	192	0	99,030	-1.1%	1.68	-29.6%	0.23	-14.99%	0.057	-1.3%	1.29%	12,178	4,733	29.61%	-4.34	-0.13
Macon-Bibb County, GA	193	1	98,750	-1.6%	1.76	-23.4%	0.23	-11.56%	0.069	-1.0%	1.04%	9,811	5,254	23.43%	0.91	-0.15
Lynchburg, VA	194	2	97,940	-1.2%	2.31	-23.9%	0.30	-12.12%	0.069	-1.4%	1.40%	9,873	7,558	23.94%	0.50	-0.14
Laredo, TX	195	2	97,520	-1.5%	1.93	-26.7%	0.25	-13.32%	0.083	-1.3%	1.29%	8,285	5,407	26.70%	1.12	-0.01
Santa Cruz-Watsonville, CA	196	-3	97,370	0.6%	2.99	-28.3%	0.40	-15.43%	0.047	-2.2%	2.22%	13,373	9,487	28.31%	8.49	-0.92
Greeley, CO	197	-2	97,180	-0.2%	3.35	-23.5%	0.43	-12.64%	0.056	-2.0%	2.01%	12,586	7,305	23.51%	1.70	-0.42
Oshkosh-Neenah, WI	198	0	95,240	-1.5%	1.36	-26.0%	0.19	-12.90%	0.058	-0.9%	0.92%	10,511	3,725	25.96%	-1.37	-0.27
Columbia, MO	199	0	92,910	-1.5%	1.45	-27.7%	0.20	-13.76%	0.066	-1.0%	1.04%	9,284	4,992	27.67%	-0.02	-0.22
Iowa City, IA	200	0	92,010	-1.3%	1.68	-25.6%	0.23	-12.78%	0.060	-1.1%	1.12%	10,452	4,968	25.61%	-2.95	-0.21
Lafayette-West Lafayette, IN	201	0	91,980	-1.5%	1.75	-21.8%	0.24	-10.74%	0.060	-1.0%	0.99%	10,415	4,015	21.80%	-3.41	-0.18
Yakima, WA	202	2	89,640	-1.6%	1.63	-23.6%	0.23	-11.61%	0.068	-1.0%	0.99%	9,687	7,803	23.62%	1.48	-0.28
Longview, TX	203	2	89,270	-1.2%	1.96	-24.8%	0.27	-12.44%	0.059	-1.2%	1.22%	10,234	5,706	24.76%	1.10	-0.42
Bloomington, IL	204	-1	89,000	-0.7%	1.52	-35.1%	0.22	-17.93%	0.050	-1.4%	1.39%	12,370	3,767	35.10%	-2.52	-0.48
Midland, TX	205	-3	88,270	0.4%	1.49	-25.6%	0.20	-13.48%	0.021	-0.9%	0.95%	27,819	5,706	25.57%	1.16	-2.40
Waterloo-Cedar Falls, IA	206	1	88,160	-1.7%	1.32	-23.4%	0.19	-11.50%	0.061	-0.8%	0.80%	10,029	4,881	23.41%	-3.69	-0.20
Bremerton-Silverdale, WA	207	-1	86,700	0.4%	3.45	-24.2%	0.47	-13.32%	0.048	-2.1%	2.16%	11,474	6,723	24.21%	2.61	-0.78
Billings, MT	208	1	86,350	-1.6%	1.46	-26.9%	0.21	-13.35%	0.068	-1.0%	1.04%	9,672	7,240	26.95%	2.45	-0.35
Athens-Clarke County, GA	209	-1	86,140	-1.1%	1.84	-26.9%	0.26	-13.55%	0.057	-1.2%	1.25%	9,947	4,731	26.85%	-1.05	-0.40
Bellingham, WA	210	0	85,600	-0.8%	2.17	-25.3%	0.31	-12.95%	0.053	-1.4%	1.41%	11,314	7,340	25.26%	5.26	-0.75
Saginaw, MI	211	2	85,160	-1.8%	1.63	-19.3%	0.24	-9.26%	0.069	-0.8%	0.82%	8,163	3,926	19.27%	-3.33	-0.12
Florence, SC	212	2	84,850	-1.6%	1.90	-20.6%	0.27	-10.12%	0.067	-1.0%	1.01%	8,555	4,226	20.65%	-0.21	-0.27
Sioux City, IA-NE-SD	213	2	84,630	-1.7%	1.41	-22.7%	0.20	-11.09%	0.062	-0.8%	0.83%	9,814	5,449	22.69%	-1.95	-0.29
Gainesville, GA	214	-2	84,620	-1.1%	2.36	-19.8%	0.33	-9.96%	0.054	-1.2%	1.18%	9,696	4,385	19.76%	0.96	-0.57
Medford, OR	215	1	84,240	-1.7%	1.51	-23.5%	0.22	-11.51%	0.072	-0.9%	0.94%	8,590	7,291	23.52%	4.50	-0.41
Houma-Thibodaux, LA	216	1	84,000	-1.5%	2.20	-18.3%	0.31	-8.98%	0.062	-1.0%	1.02%	9,474	5,293	18.27%	0.32	-0.40
Clarksville, TN-KY	217	-6	83,970	-0.3%	3.14	-22.8%	0.44	-12.20%	0.052	-1.8%	1.84%	11,027	4,954	22.84%	-0.17	-0.57
Eau Claire, WI	218	0	81,980	-1.6%	1.62	-24.1%	0.24	-11.86%	0.067	-1.0%	1.01%	8,536	4,361	24.07%	-2.67	-0.22
Panama City, FL	219	0	80,480	-1.3%	2.08	-22.4%	0.30	-11.19%	0.065	-1.2%	1.20%	8,447	5,937	22.40%	2.19	-0.46
Chico, CA	220	0	80,110	-1.3%	1.97	-25.4%	0.29	-12.76%	0.069	-1.3%	1.31%	8,168	8,397	25.39%	5.11	-0.53
Bend-Redmond, OR	221	0	78,520	-1.3%	1.65	-25.9%	0.25	-12.98%	0.057	-1.1%	1.14%	10,150	8,314	25.94%	6.10	-0.78
Joplin, MO	222	0	78,040	-1.7%	1.59	-22.2%	0.24	-10.82%	0.072	-0.9%	0.91%	7,415	4,741	22.22%	-1.32	-0.25

MSA	Rank		Employment (L_r)		Commuting, hours		Commuting tech (θ_r)		Markups (λ_r)		Utility (Ur)	GDP, th.$	N. firms	FHW, %	Amenities	
	Baseline	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Change*	*Baseline*	*Baseline*	*Baseline*	A_o	A_u
Monroe, LA	223	1	77,070	-1.5%	1.72	-23.0%	0.25	-11.32%	0.065	-1.0%	1.00%	8,028	5,123	23.01%	0.34	-0.43
Johnson City, TN	224	-1	77,050	-1.5%	2.11	-22.0%	0.31	-10.90%	0.072	-1.2%	1.19%	7,042	3,890	21.96%	1.51	-0.38
Racine, WI	225	0	74,850	-0.9%	2.54	-23.0%	0.38	-11.84%	0.058	-1.5%	1.52%	8,081	4,350	22.97%	-0.51	-0.54
Greenville, NC	226	0	74,540	-1.3%	1.99	-21.7%	0.30	-10.81%	0.058	-1.1%	1.10%	8,217	3,790	21.71%	-1.74	-0.50
Bowling Green, KY	227	1	74,020	-1.6%	1.83	-21.3%	0.28	-10.40%	0.070	-1.0%	1.00%	7,267	4,469	21.26%	-0.45	-0.36
Jefferson City, MO	228	-1	73,950	-1.4%	1.54	-32.1%	0.24	-16.11%	0.072	-1.3%	1.28%	7,289	3,974	32.13%	0.33	-0.37
La Crosse-Onalaska, WI-MN	229	0	73,800	-1.6%	1.48	-26.0%	0.23	-12.82%	0.067	-1.0%	1.01%	7,508	3,507	25.98%	-1.15	-0.39
Bismarck, ND	230	2	73,020	-1.6%	1.35	-28.9%	0.21	-14.32%	0.068	-1.0%	1.01%	7,998	4,900	28.90%	-1.56	-0.35
Napa, CA	231	-1	72,870	-0.8%	1.96	-21.8%	0.31	-11.10%	0.042	-1.1%	1.12%	11,446	5,798	21.83%	7.53	-1.32
Hilton Head Island-Bluffton-Beaufort, SC	232	-1	72,700	-1.0%	2.38	-20.8%	0.36	-10.58%	0.054	-1.3%	1.27%	9,176	6,127	20.81%	1.09	-0.74
Daphne-Fairhope-Foley, AL	233	0	71,190	-1.4%	2.50	-18.5%	0.39	-9.24%	0.067	-1.2%	1.20%	7,322	6,072	18.54%	1.82	-0.54
Wausau, WI	234	0	70,020	-1.5%	1.47	-24.2%	0.23	-11.91%	0.059	-0.9%	0.92%	8,210	3,460	24.21%	-3.30	-0.48
Warner Robins, GA	235	1	69,530	-1.1%	2.28	-24.2%	0.35	-12.29%	0.063	-1.4%	1.40%	7,163	3,369	24.19%	-0.05	-0.55
Blacksburg-Christiansburg-Radford, VA	236	2	69,370	-1.5%	2.04	-19.6%	0.32	-9.63%	0.065	-1.0%	1.02%	7,218	4,280	19.63%	0.10	-0.53
State College, PA	237	0	69,300	-1.0%	1.78	-26.4%	0.28	-13.33%	0.052	-1.2%	1.21%	9,041	3,570	26.38%	-0.40	-0.77
Merced, CA	238	-3	69,200	-0.5%	3.11	-22.5%	0.49	-11.92%	0.057	-1.8%	1.83%	8,625	6,563	22.46%	4.51	-0.86
Odessa, TX	239	0	68,280	-1.0%	2.01	-21.2%	0.31	-10.68%	0.044	-1.1%	1.07%	9,959	3,912	21.18%	2.50	-1.12
Bloomington, IN	240	0	68,210	-1.3%	1.95	-22.6%	0.31	-11.27%	0.061	-1.1%	1.14%	7,425	3,302	22.61%	-0.23	-0.60
Terre Haute, IN	241	0	67,380	-1.6%	1.84	-21.0%	0.30	-10.28%	0.071	-1.0%	1.00%	6,534	3,573	21.03%	-2.24	-0.37
Dalton, GA	242	0	65,890	-1.8%	1.66	-16.9%	0.27	-7.97%	0.063	-0.7%	0.71%	6,668	2,690	16.88%	0.46	-0.62
Rapid City, SD	243	0	65,680	-1.7%	1.61	-23.6%	0.26	-11.55%	0.073	-1.0%	0.98%	6,881	5,604	23.61%	-0.09	-0.46
Jackson, TN	244	1	65,040	-1.8%	1.44	-21.3%	0.24	-10.25%	0.067	-0.8%	0.79%	6,496	3,001	21.26%	-1.77	-0.48
Idaho Falls, ID	245	-1	64,860	-1.5%	1.58	-27.3%	0.27	-13.55%	0.071	-1.1%	1.15%	7,019	4,684	27.27%	1.04	-0.55
Redding, CA	246	2	64,640	-1.5%	1.72	-24.3%	0.29	-12.03%	0.068	-1.1%	1.09%	7,037	6,895	24.32%	5.69	-0.78
Janesville-Beloit, WI	247	0	64,560	-1.2%	2.13	-23.3%	0.35	-11.75%	0.064	-1.3%	1.29%	6,611	3,375	23.34%	-2.62	-0.50
Morgantown, WV	248	-2	64,520	-1.2%	1.81	-24.0%	0.30	-12.04%	0.054	-1.1%	1.15%	8,146	3,353	23.99%	-0.57	-0.77
Bangor, ME	249	1	64,450	-1.4%	2.05	-25.0%	0.34	-12.52%	0.077	-1.3%	1.34%	6,100	4,876	25.01%	-0.52	-0.41
Abilene, TX	250	-1	64,450	-1.4%	1.71	-25.2%	0.27	-12.52%	0.061	-1.1%	1.08%	7,289	4,246	25.21%	1.31	-0.71
Harrisonburg, VA	251	1	63,440	-1.5%	1.57	-20.8%	0.26	-10.15%	0.051	-0.8%	0.83%	8,130	3,437	20.79%	1.25	-0.94
Dover, DE	252	-1	63,290	-0.7%	2.69	-22.4%	0.44	-11.62%	0.056	-1.5%	1.56%	7,337	4,042	22.37%	-0.07	-0.77
St. George, UT	253	0	63,190	-1.7%	1.66	-24.0%	0.29	-11.77%	0.084	-1.1%	1.07%	5,538	5,638	24.01%	2.57	-0.48
Niles-Benton Harbor, MI	254	0	62,750	-1.5%	1.76	-21.5%	0.30	-10.60%	0.060	-1.0%	0.99%	6,757	3,437	21.55%	-0.30	-0.69
Muskegon, MI	255	2	61,690	-1.7%	2.07	-17.2%	0.35	-8.29%	0.067	-0.9%	0.92%	5,906	2,838	17.17%	-0.40	-0.59
Prescott, AZ	256	0	61,560	-1.3%	2.44	-22.8%	0.41	-11.52%	0.078	-1.4%	1.44%	6,111	5,635	22.82%	5.21	-0.69
Santa Fe, NM	257	-2	61,340	-0.9%	1.88	-30.0%	0.32	-15.32%	0.060	-1.5%	1.48%	7,256	5,568	30.05%	3.02	-0.85
Wheeling, WV-OH	258	0	61,310	-1.1%	1.94	-22.7%	0.33	-11.42%	0.051	-1.1%	1.15%	8,219	3,750	22.66%	-0.05	-0.93
Hattiesburg, MS	259	0	61,210	-1.5%	2.02	-21.3%	0.33	-10.50%	0.067	-1.1%	1.09%	6,165	3,468	21.27%	-0.20	-0.61
Flagstaff, AZ	260	1	60,810	-1.9%	1.59	-21.2%	0.27	-10.20%	0.085	-0.9%	0.87%	5,844	3,363	21.21%	4.93	-0.62
Albany, GA	261	-1	60,720	-1.7%	1.76	-22.6%	0.29	-11.05%	0.076	-1.0%	1.01%	5,482	3,345	22.61%	-0.04	-0.51
Alexandria, LA	262	0	60,460	-1.6%	1.74	-22.4%	0.29	-10.97%	0.069	-1.0%	0.99%	5,992	3,888	22.41%	-0.21	-0.59
El Centro, CA	264	-1	60,240	-1.8%	1.66	-19.8%	0.29	-9.50%	0.074	-0.9%	0.86%	6,029	7,324	19.77%	6.45	-0.80
Burlington, NC	265	-1	60,000	-1.5%	2.31	-18.8%	0.38	-9.28%	0.069	-1.1%	1.11%	5,441	3,439	18.82%	-0.96	-0.57
Pueblo, CO	266	1	59,750	-1.6%	2.06	-21.5%	0.35	-10.55%	0.080	-1.1%	1.13%	5,236	3,398	21.46%	2.11	-0.57
Coeur d'Alene, ID	267	-1	59,700	-1.5%	2.02	-20.8%	0.35	-10.27%	0.070	-1.1%	1.12%	5,931	5,275	20.80%	3.50	-0.74
Grand Junction, CO	268	2	59,580	-1.8%	1.53	-22.1%	0.26	-10.68%	0.077	-0.9%	0.87%	5,579	4,880	22.10%	2.26	-0.61
Missoula, MT	269	-1	59,580	-1.6%	1.46	-28.4%	0.26	-14.04%	0.077	-1.1%	1.09%	5,615	5,247	28.42%	1.74	-0.58
Jackson, MI	270	-1	59,280	-1.2%	2.20	-22.2%	0.38	-11.17%	0.062	-1.3%	1.27%	6,308	2,656	22.18%	-2.45	-0.63
Kingston, NY	271	-6	59,230	-0.3%	3.25	-26.7%	0.53	-14.27%	0.063	-2.2%	2.21%	6,255	5,224	26.68%	0.70	-0.74
Vineland-Bridgeton, NJ	272	-1	58,880	-1.5%	2.01	-19.7%	0.35	-9.68%	0.065	-1.0%	1.03%	5,828	3,102	19.70%	0.38	-0.70
Winchester, VA-WV	273	-1	58,160	-0.9%	3.06	-19.0%	0.52	-9.87%	0.058	-1.5%	1.50%	6,785	3,871	19.02%	0.27	-0.84
Sheboygan, WI	274	-1	57,980	-1.5%	1.52	-22.9%	0.27	-11.24%	0.056	-0.9%	0.90%	6,675	2,813	22.90%	-0.37	-0.85
Altoona, PA	275	0	57,160	-1.8%	1.70	-18.6%	0.30	-8.93%	0.068	-0.8%	0.82%	5,440	3,362	18.65%	-0.86	-0.64
Dubuque, IA	276	0	57,150	-1.8%	1.07	-23.9%	0.19	-11.64%	0.061	-0.7%	0.66%	6,094	3,256	23.85%	-0.79	-0.75
Chambersburg-Waynesboro, PA	277	-3	56,920	-1.3%	2.61	-19.9%	0.45	-10.05%	0.068	-1.3%	1.34%	5,545	3,292	19.91%	-0.57	-0.66
Wichita Falls, TX	278	-1	56,640	-1.5%	1.58	-22.9%	0.27	-11.22%	0.058	-0.9%	0.91%	6,784	3,681	22.88%	-0.07	-0.85
Auburn-Opelika, AL	279	-1	56,410	-1.3%	2.30	-21.4%	0.40	-10.74%	0.067	-1.3%	1.27%	5,484	3,137	21.41%	-0.24	-0.69
Dothan, AL	280	0	56,400	-1.7%	2.02	-19.2%	0.35	-9.33%	0.074	-1.0%	1.00%	5,295	3,775	19.18%	-0.41	-0.59
Battle Creek, MI	281	0	55,980	-1.6%	1.62	-21.9%	0.29	-10.73%	0.060	-0.9%	0.93%	6,154	2,246	21.94%	-2.73	-0.71
Texarkana, TX-AR	282	0	55,800	-1.8%	1.52	-22.5%	0.27	-10.91%	0.071	-0.9%	0.87%	5,402	3,551	22.51%	0.23	-0.66
East Stroudsburg, PA	283	-4	55,630	-0.2%	4.40	-18.3%	0.74	-10.12%	0.057	-2.0%	2.09%	6,340	3,563	18.34%	0.94	-0.92
Rocky Mount, NC	284	-1	55,270	-1.5%	2.02	-20.1%	0.35	-9.92%	0.062	-1.0%	1.03%	5,897	3,049	20.14%	-1.75	-0.73
Jonesboro, AR	285	-1	54,890	-1.6%	1.90	-21.1%	0.33	-10.35%	0.070	-1.0%	1.02%	5,292	3,378	21.13%	-2.25	-0.59
Logan, UT-ID	286	-1	54,570	-1.4%	1.71	-28.5%	0.32	-14.26%	0.077	-1.3%	1.30%	5,141	4,028	28.53%	2.29	-0.70
Elizabethtown-Fort Knox, KY	287	-1	53,810	-0.7%	2.47	-26.2%	0.44	-13.61%	0.058	-1.6%	1.67%	6,275	3,632	26.22%	-0.43	-0.88
Florence-Muscle Shoals, AL	288	-1	53,450	-1.6%	2.18	-18.4%	0.39	-8.95%	0.072	-1.0%	1.03%	5,044	3,383	18.35%	0.81	-0.72
Valdosta, GA	289	-1	53,290	-1.5%	1.93	-23.4%	0.34	-11.61%	0.071	-1.1%	1.14%	5,115	3,204	23.44%	0.49	-0.72
St. Joseph, MO-KS	290	-1	53,270	-1.8%	1.62	-19.0%	0.29	-9.08%	0.065	-0.8%	0.79%	5,626	3,375	18.96%	-1.17	-0.73
Glens Falls, NY	291	-1	52,990	-1.3%	2.19	-23.4%	0.39	-11.74%	0.070	-1.3%	1.30%	5,174	3,694	23.40%	-0.31	-0.71
Mankato-North Mankato, MN	292	1	52,390	-1.6%	1.59	-22.7%	0.30	-11.10%	0.068	-0.9%	0.95%	5,320	2,740	22.73%	-2.52	-0.66
Carbondale-Marion, IL	293	-2	52,370	-1.6%	1.74	-22.2%	0.32	-10.89%	0.067	-1.0%	1.00%	5,288	2,597	22.19%	0.85	-0.81
Decatur, AL	294	-2	52,310	-1.5%	2.27	-18.1%	0.41	-8.90%	0.065	-1.1%	1.06%	5,466	3,234	18.10%	0.79	-0.84
Williamsport, PA	295	0	50,770	-1.7%	1.67	-20.5%	0.31	-9.94%	0.069	-0.9%	0.89%	4,993	3,152	20.54%	0.33	-0.79
Sebastian-Vero Beach, FL	296	-2	50,590	-1.3%	2.15	-21.0%	0.40	-10.52%	0.059	-1.2%	1.17%	5,472	5,025	21.03%	4.72	-1.13

MSA	Rank		Employment (L_r)		Commuting, hours		Commuting tech (θ_r)		Markups (Λ_r)		Utility (Ur)	GDP, th.$	N. firms	FHW, %	Amenities	
	Baseline	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Baseline*	*Change*	*Change*	*Baseline*	*Baseline*	*Baseline*	A_o	A_u
Owensboro, KY	297	-1	50,520	-1.7%	1.60	-19.1%	0.30	-9.19%	0.059	-0.8%	0.79%	5,635	3,367	19.12%	-0.71	-0.90
Lewiston-Auburn, ME	298	0	50,140	-1.5%	1.88	-25.2%	0.35	-12.53%	0.076	-1.2%	1.24%	4,214	3,188	25.19%	-0.32	-0.68
Ithaca, NY	299	-2	50,120	-1.1%	1.47	-34.6%	0.27	-17.45%	0.060	-1.3%	1.30%	5,265	2,672	34.64%	-0.28	-0.92
Mansfield, OH	300	-1	49,970	-1.8%	1.83	-18.5%	0.35	-8.89%	0.072	-0.9%	0.88%	4,470	2,735	18.51%	-2.88	-0.63
Lima, OH	301	-1	49,870	-1.7%	1.53	-17.8%	0.29	-8.49%	0.056	-0.7%	0.71%	5,718	2,487	17.83%	-2.37	-0.92
Lawrence, KS	302	-1	49,520	-1.3%	2.20	-23.7%	0.41	-11.95%	0.069	-1.3%	1.34%	4,558	3,007	23.73%	0.36	-0.82
Grand Forks, ND-MN	303	-1	49,370	-1.8%	1.32	-22.5%	0.25	-10.86%	0.063	-0.8%	0.77%	5,721	3,203	22.48%	-4.29	-0.72
Johnstown, PA	304	-1	49,120	-1.4%	2.14	-25.2%	0.40	-12.69%	0.076	-1.4%	1.40%	4,238	3,513	25.22%	-0.23	-0.71
Columbus, IN	305	0	48,960	-1.8%	1.28	-17.0%	0.24	-7.99%	0.051	-0.6%	0.56%	6,044	1,904	17.03%	-2.38	-1.05
Decatur, IL	306	0	48,870	-1.7%	1.37	-18.6%	0.26	-8.89%	0.051	-0.7%	0.66%	6,224	2,188	18.64%	-2.79	-1.04
Lebanon, PA	307	-3	48,870	-1.2%	2.46	-18.9%	0.46	-9.49%	0.060	-1.2%	1.20%	5,155	2,795	18.91%	-0.66	-0.94
Farmington, NM	308	-1	48,610	-1.4%	1.86	-20.3%	0.36	-9.97%	0.058	-1.0%	0.99%	6,360	2,909	20.26%	2.83	-1.11
Springfield, OH	309	-1	47,910	-1.5%	2.22	-18.4%	0.43	-9.06%	0.068	-1.1%	1.07%	4,531	2,540	18.43%	-2.03	-0.77
Staunton-Waynesboro, VA	310	-1	47,760	-1.4%	2.24	-19.9%	0.42	-9.89%	0.061	-1.1%	1.13%	5,136	3,442	19.89%	-0.12	-0.95
Lake Havasu City-Kingman, AZ	311	-1	47,430	-1.6%	2.40	-20.5%	0.46	-10.13%	0.087	-1.3%	1.27%	4,343	3,692	20.45%	5.84	-0.88
Mount Vernon-Anacortes, WA	312	-1	46,650	-0.8%	2.28	-22.6%	0.44	-11.63%	0.049	-1.3%	1.33%	6,643	4,062	22.64%	4.94	-1.46
Punta Gorda, FL	313	0	46,530	-1.5%	2.58	-16.4%	0.49	-8.03%	0.070	-1.1%	1.09%	4,322	4,593	16.38%	5.10	-1.05
Muncie, IN	314	0	46,240	-1.6%	1.84	-22.6%	0.36	-11.11%	0.074	-1.1%	1.08%	3,961	2,182	22.60%	-2.60	-0.70
Jacksonville, NC	315	-3	46,070	0.2%	3.58	-16.2%	0.66	-8.90%	0.036	-1.5%	1.47%	8,396	3,079	16.17%	0.74	-1.81
San Angelo, TX	316	-1	45,990	-1.5%	1.67	-23.7%	0.32	-11.65%	0.062	-1.0%	0.99%	5,142	3,012	23.68%	1.59	-1.05
Wenatchee, WA	317	-1	45,710	-1.6%	1.76	-20.8%	0.34	-10.15%	0.064	-0.9%	0.94%	5,281	4,674	20.81%	1.12	-1.00
Cleveland, TN	318	0	45,690	-1.6%	2.21	-17.1%	0.42	-8.26%	0.069	-1.0%	0.97%	4,343	2,106	17.05%	0.88	-0.92
Fond du Lac, WI	319	-2	45,600	-1.4%	1.88	-21.7%	0.37	-10.74%	0.060	-1.0%	1.06%	5,052	2,331	21.69%	-1.92	-0.96
Cheyenne, WY	320	-1	45,310	-1.7%	1.31	-22.8%	0.25	-11.06%	0.054	-0.7%	0.74%	5,754	3,618	22.80%	3.05	-1.29
Sherman-Denison, TX	321	-1	45,050	-1.3%	2.42	-21.8%	0.46	-11.00%	0.066	-1.3%	1.32%	4,407	3,076	21.82%	0.78	-0.97
Kankakee, IL	322	0	44,000	-1.4%	2.13	-20.2%	0.42	-10.02%	0.064	-1.1%	1.12%	4,510	2,265	20.20%	-3.30	-0.86
Ames, IA	323	-2	43,960	-1.3%	1.60	-25.5%	0.32	-12.69%	0.054	-1.0%	1.06%	5,245	2,573	25.48%	-3.54	-1.03
New Bern, NC	324	-1	43,930	-1.3%	2.41	-19.2%	0.47	-9.62%	0.060	-1.2%	1.18%	4,988	2,846	19.21%	-0.18	-1.06
Madera, CA	325	-1	43,730	-1.1%	2.75	-17.0%	0.55	-8.62%	0.056	-1.2%	1.23%	5,550	4,197	17.04%	6.00	-1.39
Lawton, OK	326	0	43,190	-1.3%	1.64	-27.1%	0.33	-13.52%	0.057	-1.1%	1.13%	5,155	2,487	27.07%	1.77	-1.21
Anniston-Oxford-Jacksonville, AL	327	0	43,110	-1.7%	2.20	-16.4%	0.44	-7.89%	0.071	-0.9%	0.93%	3,937	2,514	16.40%	0.22	-0.92
Albany, OR	328	1	42,780	-1.5%	2.30	-20.4%	0.47	-10.13%	0.074	-1.2%	1.24%	4,180	3,459	20.36%	3.65	-1.02
Yuba City, CA	329	-4	42,770	0.7%	3.84	-23.8%	0.75	-13.35%	0.049	-2.3%	2.39%	6,024	4,830	23.85%	3.38	-0.85
Morristown, TN	330	0	42,710	-1.6%	2.17	-16.7%	0.43	-8.11%	0.061	-0.9%	0.93%	4,413	2,071	16.72%	1.44	-1.13
Hammond, LA	331	-3	42,700	-1.0%	3.33	-19.9%	0.64	-10.37%	0.069	-1.7%	1.68%	4,040	2,911	19.93%	-0.28	-0.94
Beckley, WV	332	-1	42,670	-1.5%	2.12	-18.5%	0.43	-9.09%	0.066	-1.0%	1.04%	4,478	3,160	18.54%	-0.13	-0.98
Bloomsburg-Berwick, PA	333	0	42,640	-1.8%	1.49	-19.7%	0.30	-9.38%	0.068	-0.8%	0.76%	4,049	1,996	19.65%	-0.02	-0.96
Cape Girardeau, MO-IL	334	-2	42,620	-1.5%	1.89	-21.9%	0.38	-10.80%	0.068	-1.1%	1.07%	4,315	3,153	21.91%	0.62	-0.99
Hanford-Corcoran, CA	335	-1	42,140	-1.1%	2.58	-19.2%	0.52	-9.74%	0.056	-1.3%	1.29%	5,191	3,863	19.15%	3.48	-1.32
Brunswick, GA	336	-1	41,140	-1.3%	2.32	-22.2%	0.46	-11.18%	0.068	-1.3%	1.31%	4,045	3,044	22.24%	1.03	-1.04
Pittsfield, MA	337	-1	39,900	0.0%	2.38	-23.8%	0.50	-12.64%	0.038	-1.5%	1.48%	7,119	5,166	23.83%	0.81	-1.84
Victoria, TX	338	-1	39,880	-1.2%	1.87	-22.8%	0.38	-11.38%	0.053	-1.1%	1.07%	5,076	2,662	22.82%	0.37	-1.32
Watertown-Fort Drum, NY	339	-1	39,490	-1.0%	1.70	-24.6%	0.35	-12.40%	0.045	-1.1%	1.07%	5,942	2,857	24.64%	-1.27	-1.50
Kokomo, IN	340	0	39,430	-2.1%	1.40	-10.9%	0.30	-4.69%	0.053	-0.4%	0.40%	4,624	1,683	10.93%	-3.61	-1.17
Goldsboro, NC	341	-2	39,150	-1.3%	2.47	-16.4%	0.51	-8.12%	0.055	-1.0%	1.03%	4,492	2,278	16.42%	-1.41	-1.22
Ocean City, NJ	342	-1	38,760	-1.1%	1.98	-19.8%	0.42	-9.91%	0.045	-1.0%	1.02%	5,247	3,874	19.79%	0.07	-1.54
Grand Island, NE	343	0	38,670	-1.6%	1.57	-21.9%	0.34	-10.68%	0.061	-0.9%	0.90%	4,487	3,071	21.90%	-1.62	-1.11
Monroe, MI	344	-2	38,290	-0.1%	3.61	-19.8%	0.75	-10.79%	0.049	-1.8%	1.86%	5,049	2,162	19.82%	-1.43	-1.38
Manhattan, KS	345	-1	37,660	-1.4%	1.84	-24.8%	0.39	-12.33%	0.067	-1.2%	1.17%	3,822	2,336	24.77%	-0.10	-1.10
Casper, WY	346	-1	37,590	-1.5%	1.39	-20.3%	0.29	-9.86%	0.043	-0.7%	0.69%	6,294	3,530	20.35%	2.49	-1.77
Weirton-Steubenville, WV-OH	347	-1	37,250	-1.1%	2.68	-17.9%	0.58	-9.06%	0.058	-1.2%	1.26%	4,223	2,557	17.91%	-0.43	-1.25
Cumberland, MD-WV	348	0	37,220	-1.7%	1.97	-18.8%	0.43	-9.07%	0.075	-1.0%	0.96%	3,286	2,112	18.75%	1.01	-1.05
Parkersburg-Vienna, WV	349	0	37,160	-1.6%	1.61	-20.6%	0.36	-10.05%	0.057	-0.9%	0.88%	4,278	2,352	20.64%	-0.25	-1.27
Longview, WA	350	-3	37,100	-1.3%	2.22	-19.2%	0.48	-9.58%	0.056	-1.1%	1.10%	4,477	3,004	19.22%	4.54	-1.50
Rome, GA	351	-1	37,070	-1.6%	1.99	-17.7%	0.42	-8.56%	0.059	-0.9%	0.89%	3,949	1,965	17.67%	0.33	-1.26
Sumter, SC	352	-1	37,000	-1.4%	2.23	-19.1%	0.48	-9.46%	0.062	-1.1%	1.10%	3,872	2,037	19.11%	0.45	-1.22
Corvallis, OR	353	-1	36,240	-1.0%	1.84	-25.3%	0.41	-12.81%	0.052	-1.2%	1.24%	4,668	2,703	25.30%	3.10	-1.55
Gadsden, AL	354	-1	35,630	-1.9%	2.04	-13.8%	0.45	-6.38%	0.076	-0.7%	0.73%	2,993	2,065	13.84%	0.96	-1.07
Great Falls, MT	355	-1	35,570	-1.8%	1.38	-23.4%	0.31	-11.32%	0.074	-0.8%	0.85%	3,487	2,814	23.42%	2.20	-1.15
Hot Springs, AR	356	-1	35,400	-1.7%	1.97	-19.1%	0.43	-9.25%	0.071	-0.9%	0.95%	3,215	3,120	19.07%	1.64	-1.18
Elmira, NY	357	0	34,860	-1.9%	1.51	-17.9%	0.34	-8.46%	0.064	-0.7%	0.69%	3,388	1,947	17.93%	-1.13	-1.18
Midland, MI	358	-2	34,730	-0.7%	1.82	-21.4%	0.41	-10.92%	0.035	-1.0%	1.01%	6,355	3,071	21.45%	-3.09	-1.94
Bay City, MI	359	0	34,100	-1.6%	2.18	-18.0%	0.49	-8.80%	0.067	-1.0%	1.02%	3,251	2,096	18.01%	-1.53	-1.14
Pocatello, ID	360	1	33,260	-1.9%	1.61	-18.5%	0.38	-8.79%	0.082	-0.8%	0.79%	2,818	2,331	18.55%	1.73	-1.12
Gettysburg, PA	361	-1	32,880	-0.8%	3.43	-18.7%	0.76	-9.83%	0.060	-1.6%	1.66%	3,514	2,176	18.70%	-0.13	-1.35
Homosassa Springs, FL	362	1	32,220	-1.5%	3.03	-14.0%	0.69	-6.84%	0.067	-1.1%	1.09%	3,111	3,071	13.97%	3.43	-1.40
Sierra Vista-Douglas, AZ	363	-1	32,000	-0.4%	2.72	-28.5%	0.63	-15.01%	0.061	-2.0%	2.01%	3,995	2,240	28.47%	7.13	-1.66
Pine Bluff, AR	364	0	30,960	-1.7%	1.79	-18.9%	0.42	-9.08%	0.069	-0.9%	0.86%	3,076	2,555	18.90%	-1.27	-1.21
Carson City, NV	365	0	28,100	-1.2%	1.49	-25.2%	0.37	-12.56%	0.047	-1.0%	0.97%	3,545	2,041	25.23%	7.29	-2.12
Lewiston, ID-WA	366	0	26,820	-1.9%	1.49	-18.3%	0.39	-8.63%	0.074	-0.7%	0.72%	2,512	1,878	18.31%	1.33	-1.40
The Villages, FL	367	1	26,170	-1.8%	1.74	-14.8%	0.45	-6.87%	0.060	-0.7%	0.66%	2,804	1,806	14.79%	2.84	-1.70
Danville, IL	368	-1	26,120	-1.6%	1.98	-17.8%	0.51	-8.67%	0.060	-0.9%	0.92%	2,904	1,377	17.84%	-3.21	-1.45
Sebring, FL	369	0	25,610	-1.9%	1.84	-17.5%	0.48	-8.27%	0.082	-0.8%	0.83%	2,090	2,269	17.45%	4.14	-1.47
Enid, OK	370	0	25,560	-1.8%	1.21	-20.9%	0.32	-9.98%	0.055	-0.6%	0.65%	3,078	1,750	20.93%	-1.25	-1.65
Walla Walla, WA	371	0	25,260	-1.6%	1.64	-20.9%	0.43	-10.14%	0.061	-0.9%	0.88%	2,907	2,361	20.86%	1.38	-1.65
Grants Pass, OR	372	0	24,810	-1.6%	2.11	-18.6%	0.57	-9.10%	0.071	-1.0%	1.04%	2,478	2,498	18.62%	4.26	-1.63
Hinesville, GA	373	0	17,340	1.0%	3.80	-18.2%	1.14	-10.42%	0.033	-1.7%	1.76%	3,470	950	18.22%	0.80	-2.95

CHAPTER 3

Spatial heterogeneity in factors misallocation: European evidence

3.1 Introduction

Total Factor Productivity (TFP) growth is deemed to depend, almost equally, on technological adoption and on the efficiency with which production factors are allocated across firms in the same sector, in brief 'misallocation'. In recent years, misallocation has garnered increased attention, in part due to the introduction of novel methodologies. This increased focus is motivated by the recognition of the significant first-order effects that would ensue from mitigating misallocation.

While the magnitude of allocative inefficiencies is generally larger in developing countries, it has been found to have sizeable effects also in richer economies. In the US for example, the sole misallocation due to markups dispersion has been recently found by Baqaee and Farhi (2020) to account for a 10-25% loss of the economy-wide (*i.e.* including all sectors) output, and overall misallocation for a 22-40% loss of manufacturing output by Bils *et al.* (2021).

Despite the large evidence about its economic relevance, misallocation requires further investigation to shed light on the still unexplored dimension of its spatial heterogeneity, on which the present article focuses. The issue of the extent to which regions may contribute to aggregate misallocation, and the drivers behind such differences, have in fact not been properly investigated yet.

This is quite unfortunate, as there are different reasons to expect that firms' allocative efficiencies vary systematically across administrative regions. Among these, the dispersion in subnational taxation (Fajgelbaum *et al.*, 2019), local land market regulation (Hsieh and Moretti, 2019), and differences in the quality of local credit markets (Lenzu and Manaresi, 2019) and of local institutions (Lenzu and Manaresi, 2019; Misch and Saborowski, 2020) represent the most relevant.

As a way to contribute filling this gap, this article provides a novel analysis of factors misallocation in 9 countries in the European Union, at national (NUTS0) and at different subnational levels (NUTS1, NUTS2, and NUTS3), with two main aims. Firstly, with respect to these countries, the paper quantifies the share of aggregate productivity and allocative efficiency that can be explained at regional level; on this basis, it makes the claim that within-countries disparities should be retained by misallocation research, as they may negatively affect aggregate productivity in substantial ways. Secondly, the paper investigates the correlation between established markers of aggregate misallocation and its components within and across-regions, putting a special emphasis on of the quality of local institutions; in so doing, it provides a first attempt of identifying the relevant factors behind allocative efficiency at regional level, providing policy implications and future direction of research. The latter focus is motivated by the fact that local and national institutions can influence firms' productivity and allocative efficiency in different ways. To start with, they provide a reliable formal institutional environment. By establishing and enforcing the so called rule of law (North, 1990) they reduce transaction costs, and this create the necessary incentive structure for a favorable business environment for firms productivity (Lasagni *et al.*, 2015), survival (Iwasaki *et al.*, 2022) and for factors attraction and accumulation (Iwasaki *et al.*, 2022). Moreover, through the provision of local services and public investments, institutions can again influence the attraction and accumulation of productive factors (Hall and Jones, 1999; Rodrik *et al.*, 2004; Nakabashi and Pereira, 2023). This holds true with respect to different kinds of institutions. However, in this work I will mainly focus on formal institutional quality: that is, on the quality of government public spending at the regional level (Fazekas and Czibik, 2021). This study contributes to the extant literature in different respects. First, to the best of the author's knowledge, this is the first comprehensive attempt of analysing the issue of misallocation on a within-country basis, at all levels of territorial boundaries, and of systematizing previous sparse evidence at sub-national level. Furthermore, being based on a large and internally representative panel of European countries, this study provides the first results with a good extent of general validity. Last, but not least, the paper poses a focus on the role of institutions for misallocation. In doing that, it contributes to previous work showing that institutions are a major factor in explaining cross-country variation in the within-country dispersion of marginal revenue products, once firm-characteristics have been accounted for (David *et al.*, 2021; Gorodnichenko *et al.*, 2018; Bonatti and Fracasso, 2018); and to the growing literature on the geography of institutions providing evidence of the large (within and across-country) variation in the quality of local governments (Fazekas and Czibik, 2021; Charron *et al.*, 2022).

By distinctively investigating within- and across-regions misallocation, this paper shows that the between-regions component accounts for large shares of aggregate allocative efficiency. Firms location within specific administrative boundaries is thereby shown to have an explanatory power on the phenomenon at stake, which is a comparable and often larger than that of size and age, two characteristics strongly associated with systematic differences in the ability to efficiently sourcing and allocating inputs. The role of numerous markers, generally associated with aggregate misallocation - such as firms' age, size, patenting activity and ownership structure, is tested and confirmed at the regional level. In particular, firms' age, patenting activity and, above all, the quality of local public spending, are found to significantly correlate with the between-group component of misallocation, accounting for disparities across sub-national areas at all levels of territorial aggregation.

The rest of the article is organised as follows: Section 3.2 will review the relevant literature and position this paper within it; Section 3.3 will present the theoretical model and the methodology and Section 3.4 the data and the cleaning procedure. Section 3.5 will discuss the results of both the misallocation quantification and the econometric analysis, and Section 3.6 will conclude.

3.2 Reference Literature

The definition of factors misallocation assumes the amount of labour and capital in the economy as given, and refers to the most efficient way to allocate said quantities across heterogeneous producers given a specific productive technology, *i.e.* within each sector (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017). The hypothesis behind this definition is that in absence of distortions labour and capital should be allocated by markets to producers up to the equalization of their marginal revenue-products[49]. In the words of Pete Klenow,

49 This condition implies that larger amounts of inputs should be allocated to more productive firms, resulting in a distribution of firms' sizes proportional to their productivity levels.

"misallocation exists if a social planner could implement budget-neutral targeted taxes and subsidies to induce the reallocation of inputs across activities (e.g., across products, firms or occupations) in a way that would increase the welfare of a representative agent"[50].

Heretofore the main goal of the studies on the topic has been to assess the importance of the so defined misallocation in explaining cross-country differences in total factor productivity. This is generally done through the quantification of a counterfactual amount of additional output that could be generated by re-establishing efficiency. In the extant literature, this quantification has been pursued through two different approaches: a direct approach, based on structural models estimation, identifying and evaluating the impact of specific sources of misallocation, and an *indirect* one, that exploits the condition of inputs allocative efficiency; that is, it considers the equalization of marginal revenue products of firms operating in the same sector, to quantify the incidence of misallocation in each economy.

As influentially proposed by Restuccia and Rogerson (2017), the distinct sources of allocative inefficiency can be grouped in three categories. The first, statutory provisions, includes the potential distortions related to taxation (Eeckhout and Guner, 2015; Fajgelbaum *et al.*, 2019) and regulation (Hsieh and Moretti, 2019), such as size-dependant policies (Fakos, 2020). The second category, discretionary provisions, refers to any market or institutional characteristics that may discriminate specific firms, including preferential access to credit of state-owned firms, criminal organisations (Durante *et al.*, 2019; Piemontese, 2019), cronyism (García-Santana *et al.*, 2020; Saffie, 2014) or corruption (Brugués *et al.*, 2022). The third cate-

50 STEG Lecture Series on Macro Development: Misallocation, March 2021. Retrieved from https://cepr.org/sites/default/files/STEG_misallocation.pdf

gory pertains instead to market imperfections such as financial frictions (Marconi and Upper, 2017, Caggese *et al.*, 2019, David and Venkateswaran, 2019), segmentation and market power (Baqaee and Farhi, 2020; Asker *et al.*, 2019; Brooks *et al.*, 2021). While the issue of the quantification of the 'causes and costs of misallocation' (Restuccia and Rogerson, 2017) has mainly been addressed in the macroeconomics and economic development literature, it recently began to draw the attention of urban and regional economists, especially in terms of its relationship with agglomeration (Fontagné and Santoni, 2019) and city-size (Eeckhout and Guner, 2015), as with spatial frictions including commuting (Hu *et al.*, 2020), transport (Fajgelbaum and Schaal, 2020; Asturias *et al.*, 2014) and housing costs (Hsieh and Moretti, 2019). The macroeconomic nature of the field where issues related to misallocation were originally discussed, nevertheless, made popular in its investigation the resort to a cross-country type of setting, with very few exceptions focusing on a regional analysis, to which the present work aims at contributing. Among these regional studies, a first work to retain is Calligaris *et al.* (2018), who originally proposed a decomposition of Hsieh and Klenow (2009) measure of misallocation, namely the within-sector dispersion of firms' revenue-based productivity ($TFPR_i$), by firms' location in order to investigate the relevance of its geographical dimensions. Calligaris *et al.* (2018)'s article focuses on the salience of regional misallocation in Italy in the period 1997-2012 measured at NUTS1 level, and on its correlation with a number of firm-level markers of allocative distortions. These authors found that the degree of between-NUTS1 regions misallocation in the country accounts for a tenth of aggregate distortions, and disregarded it as rather negligible. Exploiting the same methodology of Calligaris *et al.* (2018), Misch and Saborowski (2020) analyse the drivers and distribution of misallocation across Mexican States, finding that a large share of aggregate allocative inefficiency (up to %40 of the total) is explained by the between-states component, and is strongly associated with state-specific levels of labour informality, crime, corruption, market concentration and access to financial and telecommunications services. In another study, Fontagné and Santoni (2019) obtain evidence of a negative correlation between misallocation intensity and population density. With a slightly different perspective[51], Boeri *et al.* (2021) investigate the impact that collective wage bargaining play on the spatial misallocation of labour in Italian and German provinces. They find that in presence of large geographical across-regions disparities in firms' productivity, as that observed in the two countries, aggregate misallocation could be reduced by allowing for lower degrees of rigidity in local wage-adjustments. Finally, Lenzu and Manaresi (2019) produce estimates of factors misallocation in the Italian corporate sector for the years 1997-2013, finding that average yearly reallocation gains measured at NUTS1 level did not differ substantially from the gains measured at national level. Macro-regions (NUTS1) are here suggested as a convenient unit of analysis in order to account for the underlying feasibility constraints of the re-allocation of productive factors across spatially distant firms[52].

Apart from the previous studies, a regional perspective is still uncommon in

51 Boeri *et al.* (2021) analyse the impact on misallocation that the rigidity of national wage-bargaining systems have on the Italian and German provinces, in the context of the comparably large geographical differences in firms' productivity observed in the two countries. The authors find that the adoption of an higher degree of flexibility for wage-adjustments to local labour productivity levels, to the same extent of the one prevailing in Germany, would produce aggregate employment (11.04%) and earnings (+7.45%) gains in the country in Italy.

52 Aspects such as relocation costs, regional human capital availability and/or the unwillingness to relocate of the specialised workers where said human capital is embodied, are rarely considered in the quantification of potential aggregate reallocation gains. One exception is the work of Heise and Porzio (2022) who estimate the effect of spatial frictions on the efficiency of workers distribution and mobility across German macro-regions. They find that removing all spatial barriers to workers mobility, including the home-bias highlighted in their locational preferences, would produce gains in per capita GDP (5%) and in average real wages (9%).

the analysis of misallocation and this is quite unfortunate, given that there are several reasons (such as agglomeration economies, spatial frictions to factor mobility, market segmentation and subnational tax dispersion, among others) to expect firms' location to affect the efficiency of factors allocation. This paper contributes to this narrow strand of literature, framing the analysis of the relevance and of the most adequate unit of analysis for within-country misallocation.

It should be acknowledged that existing studies on factors misallocation in the European context have largely focused on its explaining the productivity gap between Southern countries and their Northern counterparts (Gamberoni *et al.*, 2016; Pellegrino and Zingales, 2017; Calligaris *et al.*, 2018; García-Santana *et al.*, 2020). These studies have emphasized the influence of institutions, familism, and cronyism in exacerbating factors misallocation within these countries. Additionally, as mentioned earlier, some of these studies (Pellegrino and Zingales, 2017; Calligaris *et al.*, 2018; García-Santana *et al.*, 2020) have also shed light on the presence of within-country disparities in both allocative efficiency and productivity. Despite the relevance of this literature, this chapter aims to examine the spatial dimension of factors misallocation within a number of countries and assess its impact on aggregate allocative efficiency at the national level. Because of that, for the sake of conciseness, the chapter abstains from exploring the role that spatial misallocation plays in productivity divergence in general, and in the slowdown of Southern European countries.

Furthermore, this paper also contributes to another scant stream of research, about the existence of distortions, which affect firms asymmetrically across their location. Only few studies in this stream have so far focused on different distortions which can differ across regions and cities of different sizes, such as the quality of local credit markets (Lenzu and Manaresi, 2019); decentralized public service provision and taxation (Fajgelbaum *et al.*, 2019; Martínez-Vázquez and Li, 2020), and institutional quality[53] (García-Santana *et al.*, 2020; Piemontese, 2019). Also here, the role of financial frictions and credit constraints itself has already been established as an important driver of misallocation (León-Ledesma and Christopoulos, 2016; Shikimi, 2017), the relationship between regional variations in access to finance and regional misallocation remains largely unexplored. To the best of this author's knowledge, the only work linking these two factors is Lenzu and Manaresi (2019), who found misallocation to be higher in Italian NUTS1 regions characterized by weaker financial markets and socioeconomic institutions. In turn, the emerging literature on institutional quality has recently begun to provide evidence of a direct effect of it on productivity (Rodriguez-Pose and Ganau, 2022; Lasagni *et al.*, 2015) and on misallocation (León-Ledesma, 2016; Misch and Saborowski, 2020; David *et al.*, 2021), and of some possible indirect effects, due to distortions in lending relationships (Nifo *et al.*, 2018) and in local factor availability (Rodriguez-Pose and Ganau, 2022). In this stream, a recent contribution by David *et al.* (2021) has applied to a large panel of developing and rich countries a novel methodology proposed by David and Venkateswaran (2019) that allows to disentangle the contribution of different aggregate and firm-level sources of misallocation. Once accounted for the most commonly acknowledged sources of cross-sectional variation in within-country and within-sector marginal products across firms (including adjustment costs, mark-ups dispersion, and heterogeneity in firm-level technologies and firm-characteristics), the authors find that almost half of the observed within-country dispersion in $TFPR$

53 The quality of local institutions and of the business environment have been reported to vary extensively within European countries, across different cities and regions (Charron *et al.*, 2022; Charron *et al.*, 2014) despite the existence of common formal institutions at national level.

remains unexplained. However, the same dispersion emerges as correlated with direct measures of the quality of the business environment (highlighting in particular the role of access to credit, the effectiveness of the legal system and bankruptcy laws). Their findings are consistent with those of Gorodnichenko *et al.* (2018) on the greater impact that a country's business, institutional and policy environment have on misallocation, over that of differences in firms' characteristics. Along the same stream, Bonatti and Fracasso (2018) find that bad performances in terms of Regional Competitiveness Index, post-crisis recovery, lack of convergence and high misallocation levels in the Euro Area periphery (Italy, Spain, Greece and Eastern European countries) are linked with structural factors rather than cyclical ones, and especially with institutional quality, consistently with García-Santana *et al.* (2020)'s findings on the role of cronyism for misallocation levels in Spain, and with Misch and Saborowski, 2020's ones on Mexico's across-regions allocative disparities. I both draw and contribute to this line of research by including measures of local institutional quality in the set of possible markers of misallocation. Furthermore, in the second part of the article, I will regress these markers against measures of regional and aggregate distortions, extending previous general and country-specific findings.

3.3 Methodology

3.3.1 Measures of misallocation

To quantify aggregate misallocation using micro-data, I exploit the methodology proposed by Hsieh and Klenow (2009). This method allow to analyze the effects of two types of distortions faced by a firm i in sector s: a capital wedge $\tau_{K_{si}}$ affecting the relative marginal revenue product of one factor with respect to the other, and an output wedge $\tau_{Y_{si}}$ affecting the marginal products of both factors, human and physical capital, by the same proportion. Their model, of which a summarized derivation is presented in the Appendix, shows sectoral (log) TFP to be negatively correlated with the dispersion in revenue total factor productivity in each sector s[54], in turn proportional to wedges (see Eq.3.6.20 in the Appendix). In absence of distortions, and under the model's assumptions, firm level total factor productivity revenue ($TFPR_{si} = P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}$, eq. 3.6.17)[55], *i.e.* the ability to generate revenue from given inputs of firm i in sector s, should be equal to the sectoral mean.

'Relative $TFPR$', measured as the ratio of the firm's $TFPR$ on its sectoral average ($\overline{TFPR_s}$), can thus be exploited as a firm-level measure of misallocation. A value below unity of relative $TFPR$ indeed implies that the firm is inefficiently oversized, as part of the inputs utilised could generate larger revenues if were to be allocated to producers operating in the sector with above-average $TFPR$. The same but opposite argument applies for values of relative TFPR above-unity, indicating an inefficient under-sizing of the firm.

The economy-wide dispersion of $TFPR$, that can be exploited as a measure of aggregate misallocation, is obtained through the country-level aggregation of sectoral dispersion in $TFPR$, weighted for the value-added share of the firm within the sector ($\frac{VA_{si}}{VA_s}$) and for the sectoral share of value added in the economy ($\sum_{s=1}^{S}\frac{VA_s}{VA}$).

54 The negative correlation among the dispersion in revenue-productivity and the sectoral output can be shown, in the context of Hsieh and Klenow (2009) model, to be the following: $lnTFP_s = \frac{1}{\sigma-1}ln(\sum_i A_{si}^{\sigma-1}) - \frac{\sigma}{2}var(lnTFPR_{si})$.

55 Firm-level TFPR is calculated using the cost-shares method, which exploits first-order conditions from the firm's cost-minimization problem. Under the assumptions of constant return to scales and Cobb-Douglas technology, it can be shown that the share of input expenditures in total costs identify factor elasticities even without data on prices and quantities (Foster *et al.*, 2016; Blackwood *et al.*, 2021).

This is done in order to account for the productive relevance of the firm and the sector in the computation of aggregate distortions.

$$Var(TFPR) = \sum_{s=1}^{S} \frac{VA_s}{VA} \sum_{i=1}^{N_s} \frac{VA_{si}}{VA_s} (TFPR_{si} - \overline{TFPR}_s)^2 \tag{3.1}$$

Within Hsieh and Klenow (2009)'s model assumptions for market and production structure[56], this definition of misallocation enables the quantification of a counterfactual efficient level of output in the case distortions were to be cleared. As such, the potential gains from re-allocation can be obtained as a ratio between the observed (Y) and counterfactual efficient (Y^*) yearly output [57]:

$$\frac{Y^*}{Y} = \prod_{s=1}^{S} \left(\frac{A_s^*}{A_s}\right)^{\theta_s} = \prod_{s=1}^{S} \left[\frac{1}{N_s} \sum_{i_1}^{N_s} \left(\frac{A_s^*}{A_{si}} \frac{TFPR_{si}}{\overline{TFPR}_s}\right)^{\sigma-1}\right]^{\frac{\theta}{\sigma-1}} \tag{3.2}$$

where

$$Y = \prod_{s=1}^{S} (A_s)^{\theta_s} = \prod_{s=1}^{S} \left[\frac{1}{N_s} \sum_{i_1}^{N_s} \left(A_{si} \frac{\overline{TFPR}_s}{TFPR_{si}}\right)^{\sigma-1}\right]^{\frac{\theta}{\sigma-1}} \tag{3.3}$$

and

$$Y^* = \prod_{s=1}^{S} (A_s^*)^{\theta_s} = \prod_{s=1}^{S} \left[\frac{1}{N_s} \sum_{i_1}^{N_s} (A_{si})^{\sigma-1}\right]^{\frac{\theta}{\sigma-1}} \tag{3.4}$$

As shown in Eq.3.3 and Eq.3.4, the observed and efficient level of aggregate TFP

56 Hsieh and Klenow (2009)'s contribution is based on a closed-economy version of Melitz (2003)'s model, where the economy is represented by three levels of production. The first level is constituted by N_s monopolistically competitive firms using labour and capital to produce varieties i_s of each M_s sectoral differentiated product, through a constant returns to scale Cobb-Douglas technology, with output Y_{si}. The second level produces the sectoral output Y_s through a constant elasticity of substitution (CES) function using as input the differentiated product M_s. The third level is constituted by a representative firm in a perfectly competitive final output market, producing a final output Y combining the S sectoral outputs Y_s through a Cobb Douglas technology. Under these assumptions, the $TFPR$ can be shown to depend by the output and capital wedges and by no other firm-level characteristic.

57 Time indexes are suppressed to improve clarity and readability.

only differ by the Relative TFPR ($\frac{TFPR_{si}}{TFPR_s}$) being equal to one when the sectoral dispersion in TFPR is null. Note that the ratio between optimal and observed output equals the ratio between optimal and observed aggregate TFP, such that we will refer to output or productivity gains indistinctly in the text. The percentage gains from reallocation in each country are thus obtained as:

$$\%Gain_t = \left(\frac{Y_t^*}{Y_t} - 1 \right) * 100 \tag{3.5}$$

Turning to the measures of regional misallocation, I refer to Calligaris *et al.* (2018), who proposed a within- and between-group decomposition of the dispersion in aggregate $TFPR$:

$$\begin{aligned} Var(TFPR) = \quad & \underbrace{\sum_{g=1}^{G} \frac{VA_g}{VA} \overbrace{\sum_{s=1}^{S} \frac{VA_{gs}}{VA_g} \sum_{i=1}^{N} \frac{VA_{gsi}}{VA_{gs}} (TFPR_{gsi} - \overline{TFPR_{gs}})^2}^{\text{within-group within-sector dispersion}}}_{\text{within-group: weighted av. of the within-group dispersion from the group mean}} \quad + \\ + \quad & \underbrace{\sum_{g=1}^{G} \frac{VA_g}{VA} \sum_{s=1}^{S} \frac{VA_{gs}}{VA_g} (TFPR_{gs} - \overline{TFPR})^2}_{\text{between-group: weighted av. of the group means from the overall mean}} \end{aligned} \tag{3.6}$$

Eq. 3.6 can be calculated for different types of groupings g: the geographical location, the age, or the size of the firm.

This decomposition enables two things: firstly, the quantification of the between-group component as share of the aggregate dispersion in $TFPR$ is used to evaluate the relevance of a specific geographical grouping in explaining misallocation, comparing it to the between-group component evaluated through firms age and size grouping. If the share of aggregate misallocation expressed by the between-group component were to be sufficiently low, performing an aggregate analysis or by-group analysis would be rather indifferent. Instead, if the between-group component is sufficiently large, a within-group analysis will be justified, meaning that disparities across-groups participate in overall misallocation and, in the case of the geographical grouping, may be driven by distinct local factors. Secondly, the within and between group components can be regressed with respect to a number of economic and institutional characteristics of the location, in order to analyse the correlation with a number of potential misallocation drivers.

The quantification of the between-group component as share of the aggregate dispersion in $TFPR$, when measured for firms' location at different levels of territorial aggregation, is used to evaluate the relevance of a specific geographical grouping in explaining misallocation, comparing it to the between-group component evaluated through that obtained for the firms' age and size grouping. If this share were to be sufficiently low, performing an aggregate analysis or by-group analysis would be rather indifferent. Instead, if the between-group component is sufficiently large, a within-group analysis will be justified, meaning that disparities across-regions participate in overall misallocation and may be driven by distinct local factors. If so, the within and between group components will be regressed with respect to a number of economic and institutional characteristics of the location, in order to analyse the correlation with a number of potential misallocation drivers.

Indeed from Eq.3.6 can be derived, for each territorial level, a within-region misallocation measure (Eq.3.7), from which the correspondent regional output gains can be computed[58], together with the share of aggregate misallocation represented by

58 To compute regional output gains, Eq.3.6.24 is exploited, obtaining regional output as the aggregation of sector-region products.

the between-group component (Eq.3.8). The latter measure, exploited to investigate which characteristics drive disparities across regions, but it has a further meaning: when large shares of the aggregate misallocation are represented by the between-group component, it can be deduced that some areas are significantly more inefficient than others within the same country, such that by targeting misallocation (and its sources) in those areas, policy makers could substantially decrease aggregate misallocation.

$$\textit{Within-group} = \sum_{s=1}^{S} \frac{VA_{gs}}{VA_g} \sum_{i=1}^{N_s} \frac{VA_{gsi}}{VA_g s}(TFPR_{gsi} - \overline{TFPR}_{gs})^2 \tag{3.7}$$

$$\textit{Between-group, \%} = \frac{\sum_{g=1}^{G} \frac{VA_g}{VA} \sum_{s=1}^{S} \frac{VA_{gs}}{VA_g}(TFPR_{gs} - \overline{TFPR})^2}{Var(TFPR)} \tag{3.8}$$

3.3.2 Estimation approach and identification strategy

In order to explore the correlation between regional and aggregate measures of misallocation and those firms and regional characteristics usually associated with firms' allocative efficiency, henceforth referred to as 'markers', I apply the following econometric strategy (similarly to Calligaris *et al.*, 2018 and Misch and Saborowski, 2020).

The dependent variables of interest (misallocation, reallocation gains, and, for subnational regions, the between-group component share of total misallocation), measured for each level of territorial aggregation (g=NUTS0-3), are regressed with respect to a specific misallocation marker (W_{tg}), time (ζ_t) and region (ϵ_g) fixed effects[59]:

$$DepVar_g = \alpha + \delta W_{tg} + \zeta_t + \epsilon_g + \eta_{tg} \tag{3.9}$$

The vector of misallocation markers W_{tg} will in turn identify a number of firm's characteristics, such as the average age, size, patenting activity and ownership type, measured at the specific regional level; a number of controls, such as regional GDP, population density, the degree of market concentration, and the type of region (administrative or statistical); and a proxy for institutional quality: the Public Spending Quality Index, and its sub-dimensions[60].

The selection of these regressors followed the following theoretical considerations. Firm size is linked to misallocation in a number of ways. Firstly, micro and small firms tend to lack the necessary collaterals to obtain debt capital (Pollard, 2003), and as such they may suffers from significant factors price wedges. As such, the degree of allocative efficiency should correlate negatively with firms' size. However, older firms tend to have greater degrees of market power (Mertens and Mottironi, 2023), which has been established as a marker of aggregate misallocation (Asker *et al.*, 2019). Similarly, the age of firms should be accounted for when evaluating misallocation since it correlates to the ability to source debt and equity capital (Beck and Demirguc-Kunt, 2006), given the likely lack of collaterals and track history of the youngest ones. Moreover, firms' allocative efficiency can be linked directly to age, which as it has been found to increase together with the accumulation of organisational capital (Bradford *et al.*, 2001). For all these reasons, firms' age is

59 Exploiting a two-way fixed estimator allows to control for all time-invariant unobservables, and for time-specific shocks and yearly trends, including the European debt-crisis of the period 2011-2013, or the beginning of Covid-19 in 2020.

60 The markers are measured, whenever possible, at the same level of aggregation g as the dependent variable. The subscript g in notation identifies each specific region and different levels of territorial aggregation.

deemed to correlate negatively with misallocation. Another potentially relevant misallocation marker is patent ownership. Patents can be used as signals of quality to investors (Hottenrott *et al.*, 2016) and as collaterals (Amable *et al.*, 2010) in order to reduce financial constraints. Nevertheless, intangible capital have a lower collateral value than physical one (Caggese and Perez-Orive, 2022), deters competition by ensuring a competitive advantage to owners (Aghion *et al.*, 2019) and by rising entry costs (Chiavari and Goraya, 2021), such that intangible-capital intensive sectors tend to exhibit higher degrees of misallocation (Gordeev, 2020).

Firms' ownership structure is also considered, for being linked with firms risk propensity, as family-owned firms might be more conservative than others(Michelacci and Schivardi, 2013), and to efficiency and corruption (Pellegrino and Zingales, 2017) while government-led firms might be less efficiency-oriented than others Restuccia and Rogerson (2017).

One of the main variables of interest, *i.e.* the institutional quality, and specifically the quality of government spending, could be considered endogenous with respect to the level of productivity in a region. The quality of government spending is expected to influence firms' productivity through the type and quality of public investments (e.g. on productive infrastructures and business services), since the business environment is deemed to benefit from well-functioning formal an informal institutions. In turn, productivity might influence institutions in a positive way (through the incentive of potentially higher budget availability and remuneration of politicians and public servants) or negative way (for higher capital inflows may impose lower degrees of efficiency on government spending, as argued by Cai and Treisman, 2005)[61]. For

61 Cai and Treisman (2005) show that the quality of government spending allocation will depend on asymmetric initial endowments in presence of competitive capital mobility across countries and regions such that tax rates will diverge. Moreover, they demonstrate how initially-advantaged

this reason, it will be instrumented through climatic data on historical precipitations and temperature variability (as described in the Section 3.4). The rationale behind it lays on the exogenization of institutional quality through an instrument highly correlated with current levels of institutional quality, but not with productivity in the manufacturing sector. Historical climatic risk, measured through the interannual variability in terms of both precipitations and temperatures, has been recently shown to be significantly and positively correlated with current institutional quality in European regions by Buggle and Durante (2021). The authors explain this relationship through the historical building of cooperative behavior (at the basics social trust and institutions) needed to create forms of insurances and mutual help involving neighboring communities in regions with higher climatic risk. In a period such the pre-industrial one, when the economy was almost entirely based on crops, these cooperative strategies would have been crucial to cope with weather fluctuations that could lead to severe famine otherwise. Given that the building of social trust was traced back to the pre-industrial period, we can safely consider that data on climate variability dating back to that era as not directly correlated with productivity in the industrial sector, if not through the channel of institutional quality. In order to increase the explanatory power of our instruments, and given these are not perfectly multicollinear, the period-mean standard deviation of temperatures and precipitations will be jointly included (instead of exploiting their average as a unique regressor as in Rodriguez-Pose and Ganau, 2022).

areas will continue to attract capital inflows disregarding higher levels of non productive public investments. Viceversa, less attractive regions may be unable to revert their fortune through productive public investments, and may be also prone to devote inefficiently high portion of public spending in welfare measures, not being able to secure sufficient returns to investments when competing in business-services provisions. This would produce an asymmetric equilibria in which both type of regions may lack the incentives to efficiently allocate public funds among productive and non productive public goods

The choice of the estimation strategy followed the execution of a number of tests. The inclusion of regional fixed effects was determined by both theoretical and statistical inferential reasons. The theoretical motive behind the inclusion of country or regional fixed effects in cross-country analysis is that, in such contexts, measures (and errors) are deemed to be influenced by location-specific characteristics. As far as the inferential reasons are concerned, the Hausman test, performed for all models, produced results that varied depending on the set of variables included, but with a strong predominance of responses in favor of a regional fixed-effect estimator. The inclusion of time-fixed effects was instead dictated by the the joint significance, uncovered through LM tests, of the year-dummies coefficients in all models. Finally, the presence of heteroskedasticity was ascertained through a modified Wald test, and the presence of serial autocorrelation through a Wooldridge test. In this situation, we are required to compute cluster-robust standard errors, since the 'inclusion of cluster-specific fixed effects may not fully control for cluster correlation (and/or heteroskedasticity), and default standard errors that assume errors to be i.i.d. may be invalid' (Cameron and Miller, 2015, p. 14). Cross-sectional dependence was instead discarded through both the Pesaran parametric test and the Breusch-Pagan LM test. As a result of the above procedures, all the regression models presented in the results Section, and specifically in Tables 3.2-3.4, are estimated as linear models with the inclusion of year and region fixed effects, controlling for cluster-robust standard errors by-region. For the estimations at NUTS0 and NUTS1 level however, I compute and report wild-bootstrapped p-values, clustered respectively at NUTS0 and NUTS1 level, in order to improve accuracy in the face of a small number of clusters (Roodman *et al.*, 2019; Cameron *et al.*, 2008). Finally, the IV models' in Tables 3.2-3.4 report instead jackknife pvalues, which have been recently suggested to be more accurate than wild-bootstrapping methods in presence of weak instruments, small sample sizes and heteroskedastic cluster disturbance (Young, 2022). Regional fixed effects and error clustering are measured at the territorial aggregation of the dependent variables, thus in turn at NUTS0,1,2,3[62] level depending on the model.

3.4 Data

The main analysis is based on firm-level balance-sheet data, extracted from the Orbis Bureau Van Dijk database. This census-like database have been extensively adopted for firm-level cross-country research on misallocation[63]. Nonetheless, it suffers from a number of drawbacks, that have been documented (Bajgar *et al.* (2020); Gopinath *et al.*, 2017) and addressed by the literature in recent years (Kalemli-Ozcan *et al.*, 2022). Among these, the most concerning one is the lack of coverage and representativeness of firms with all necessary data entries for some countries and years. I followed the methodology suggested by Kalemli-Ozcan *et al.* (2022), to make sure to download all the historical data available in Orbis since 1995[64], and to select the countries and period of analysis that could ensure a sufficiently high coverage and representativeness for a good number of years[65]. The final panel sample is

62 It must be noted that at NUTS3 level the only measure of institutional quality available, *i.e.* the Public Spending Quality Index, is provided as a time-invariant mean for the period 2005-2016. As such, the models referring to institutional quality at this regional level could not be estimated with fixed-effects, and are not reported.

63 See Heise and Porzio (2021); David *et al.* (2021); Kochen (2022); Fakos (2020) and Gopinath *et al.* (2017) for some recent applications.

64 As also reported by Bajgar *et al.* (2020), in the Orbis dataset firms that do not report their informations for more than 3 years, or that are inactive for at least 5 years, are removed. Using different vintages of the Orbis dataset, as suggested by Kalemli-Ozcan *et al.* (2022), allows to ensure the best coverage and to reduce the survival bias in the data.

65 I collected data for 16 EU countries for the period 1995-2021. By comparing the collected data with the SBS dataset (merged at 2-digits Nace Rev.2 level), I then selected the countries that

constituted by about 470'000 manufacturing firms, located in 9 EU-member countries (Austria, Czech-Republic, France, Germany, Italy, Poland, Portugal, Slovenia, Spain), analysed for the years 2011-2020. These countries are evenly distributed in three different areas of the EU (Southern, Western and Eastern) and are characterised by different levels of development, with Eastern countries still in the process of convergence. The final sample match quite well the by-size decomposition of Value Added in the full population of manufacturing firms, as shown in Table 3.6 where I report the share of Value Added and of firms' number with respect to those in the official Census data from the Eurostat SBS sectoral statistics. For the selected countries, our sample is able to reproduce around 84% of total manufacturing value added and 72% of the turnover, through a mere 19.46% of the total number of firms. However, as can be inferred through the by-size decomposition, the majority of the missing firms in Orbis are micro-firms (with less than 10 employees) which, notwithstanding their numerosity (these represent the 86.5% of the total number of firms in the SBS dataset and the 46.7% in the final Orbis sample, as can be derived from Table 3.6 in the Appendix), tend to account for a negligible share of the total value added (1.3% in this sample and 5.6% in the SBS one), such that their absence should not distort too heavily our results. Moreover, this is a common problem to all book-value based researches, being due to the lower financial reporting requirements in every country, that make their data largely unavailable (see Kalemli-Ozcan *et al.*,

ensured a coverage of at least the 60% of the Value Added and a sufficiently stable share of covered firms in the final period under analysis. The sole exceptions to this rule are Czech Republic (50.7% of covered Value Added) and Poland (36.6%), that were retained in the final sample to ensure an equal representation of Eastern European countries. Finally, the period 2011-2020 was selected as it did not suffer from notable shifts in representativeness for the selected countries. It worth mentioning that these comparisons have been produced after a data cleaning procedure where all firm-year observations that had either missing or negative values for any of the main variables used to calculate TFPR and TFPQ, being the cost of employees, tangible fixed assets, value added and turnover, where removed.

2022 for a detailed discussion). Finally, since these figures have been produced on the final sample obtained after the complete data cleaning procedure[66], they should be considered as rather conservative.

The main variables needed to quantify *TFPR*[67] and its dispersion (see in eq. 3.6.17 and eq. 3.1), are the cost of labour per worker, the book value of fixed capital net of depreciation (based on the Orbis variable *'Fixed tangible assets'*) and *Value Added* as a measure of total revenues. The labour shares at industry level are computed through the industry mean of the firm level ratios of 'labour expenditure' on 'value added'. Sectors are identified at 3-digits level of the NACE Rev.2 classification codes, for all manufacturing sectors, excluding 'coke and petroleum products' for their peculiar behaviors that relies on international regulations and policies. Nominal variables for Value Added and Cost of Employees are deflated through OECD STAN Isic Rev.2 Value-Added deflators at two-digits sector level (see Bajgar *et al.*, 2020 for a discussion on the most suitable deflator for labour costs). Since the Investment (Gross Fixed Capital Formation) deflator is not available at two-digits sector level for all countries, the Eurostat Producer Price Index were deployed to deflate Fixed Assets. Descriptive statistics by country and for the whole period (2011-2020) for the real values of Value Added, Capital, Cost of labour and number of employees are reported in Table 3.8.

As for the geolocalization of firms, Orbis provide location at either postcode, city,

66 As a usual standard procedure in the field, all negative values for our measures of nominal inputs, costs and revenues have been dropped. The 1st and 99th percentile of the distribution of log total factor productivity (log TFPQ), and of log total factor productivity revenue (TFPR), were trimmed to exclude outliers.

67 Firm-level TFPR is calculated by assuming a Cobb Douglas technology and CRS, as $TFPR_{si} = P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}$. Firms' reported value added is used as measure of output ($P_{si}Y_{si}$). The labor (L_{si}) and capital (K_{si}) inputs are measured through the cost of labour and the book value of tangible fixed assets net of depreciation. Finally, the labor share $(1-\alpha)$ is obtained through the industry mean of firms' labor expenditures on value added.

or NUTS3 level for almost all its entries[68]. In the interest of data availability and sample size, NUTS geographies were treated as stable, *i.e.* disregarding changes in boundaries, adopting NUTS' 2013 boundaries as reference throughout the whole period. This is made possible by the limited number of amendments made to NUTS regions in the countries and period under analysis. However, changes in the type of NUTS regions occurred during the period of analysis, were taken into account through a specific control variable, controlling for if a specific level of NUTS territorial aggregation in each country and year identifies areas with administrative powers, or if the territory is grouped for mere statistical reasons. A summary on the type, number and changes for all levels-NUTS units in our data for the period 2010-2020 is reported in Table 3.11.

The variables for firms' age[69], size[70], ownership[71], patenting activity[72], and market concentration[73] are also created through Orbis data, while regional controls such as GDP and population density are collected through the Eurostat regional and structural databases.

Another set of independent variables are the ones referring to institutional quality. Most of the literature on Institutional Quality make use of the European Quality

68 There are only 35 firms, for a total of 109 firm-year observations, whose location cannot determined at NUTS3 level, 31 of which cannot be located at any other degree of territorial aggregation below the country. These observations were maintained in the country-level analysis, and dropped in the regional one. The country-level misallocation measure did not vary in any tangible way with their exclusion, given they were quite evenly distributed across countries.

69 Firms are divided into five age groups: $<$5 years old, age 5-9, 9-19, 20-29, and $>$30.

70 The categorical variable for the firm's size follows the OECD definition by number of employees (*micro* $<$10; *small* 10-49; *medium* 50-250; *large* $>$250).

71 Ownership is defined through Orbis information on the global shareholder, and distinguishes among: Corporates; Families or Individuals; Private Equity Firms , Venture Capitalist, and Pension funds; Public entities, and Foreign.

72 The patenting dummy takes a value of 1 if the firm has at least one patent, and 0 otherwise.

73 Industrial concentration is measured through the Herfindahl-Hirschman Index (HH Index), calculated as the sum of the squares of each firm's share of operating revenues in a given sector, country (or region) and year.

of Government Index (EQI), developed and released by Charron *et al.* (2014) and Charron *et al.* (2022). The EQI Index is a repeated-cross sectional database, available as a repeated cross-section for the years 2010, 2013, 2016 and 2021, based on a surveys on both the perception and direct-experience of corruption, meritocracy and impartiality in the public services. For this work however, I will instead exploit a new source of data on institutional quality: the Public Spending Quality Index (PSQ). The PSQ dataset was recently developed and made available by Fazekas and Czibik (2021), in order to explicitly address some of the limitations of the EQI data. The PSQ Index is based on Tenders Electronic Daily data on public procurements, and measures transparency, competition, administrative efficiency and the control of corruption risk, and is available at NUTS3 level (as a mean of the period 2006-2015), and yearly at NUTS2 level. The latter will be preferred in the model given that, not being based on surveys and perceptions as the EQI Index, it provides an objective measure of the quality of institutions at subnational level, which is better suited for within- and across-countries comparisons (Fazekas and Czibik, 2021). Moreover, being available at NUTS2 level in a longitudinal form for the years 2006-2015, it can be exploited for the whole period of analysis (2011-2020) through the inclusion of 5-years lag [74].

Since the quality of government spending can be endogenous to the level of productivity in a region, this will be in turn instrumented, similarly to Rodriguez-Pose and Ganau (2022), through data on the climate variability in the pre-industrialization era (1500-1740 A.C.) provided by the European Seasonal Temperature and Precipitation Reconstruction (ESTPR) database. In particular, the interannual standard

74 It is reasonable to assume that it takes some time for the effects of changes in the quality of local government spending to take place and for agents to react to them. Different lags have been tested, with little differences in terms of significance and magnitude of the model coefficients.

deviation of precipitations and temperatures in the crops growing seasons (spring and summer), were measured for each data point (being the centroid of each 0.5° cell of a grid covering all Europe). To ensure that even the smaller or coastal units obtain a measure, and to account for the large variability in territorial sizes existing across NUTS regions, the grid-data points were assigned to each NUTS3 unit through the mean of the 4 nearest neighbors to the regions' centroids. For NUTS0, NUTS1 and NUTS2 regions instead, the instrument is assigned through the mean of all data points within each region's perimeter, given that these regions tend to match more closely historical borders and that their sizes are more comparable and large enough to ensure a positive number of data points within their boundaries. The 240 years of observations between 1500 and 1740 A.C. were then divided into 12 intervals of 20 years each. For each of these two-decades period, the mean of the standard deviation of precipitations (Pauling *et al.*, 2005) and temperatures data (Luterbacher *et al.*, 2004; Xoplaki *et al.*, 2005) exploited to instrument the consecutive lagged values of the yearly (2005-2016) PSQ index at NUTS2 level.

Finally, we control for some major socio-economic characteristics of the region, such as the per capita gross domestic product (at NUTS2 level), the population density (at NUTS3 level), and for the type of region (statistical or administrative).

A summary of the time-span, data sources, type and definitions for all the variables included in the regression models, can be found in Table 3.9 in the Appendix.

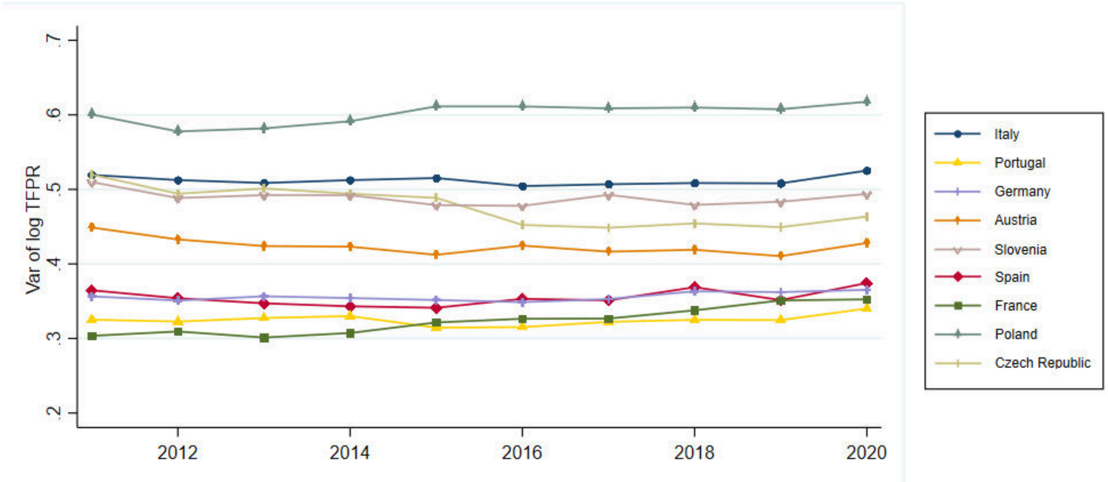


Figure 3.1. Aggregate Misallocation: dispersion of $log\ TFPR$

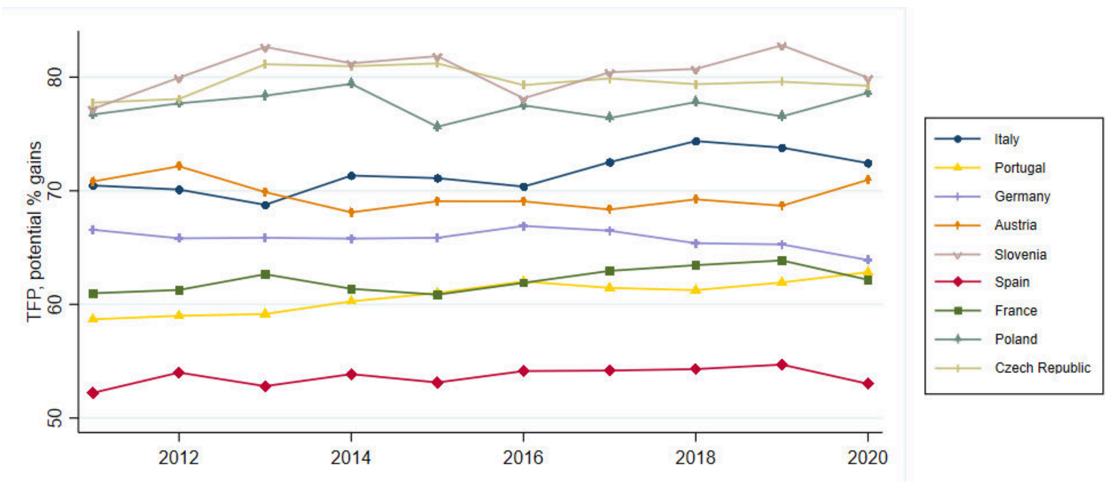


Figure 3.2. Aggregate manufacturing potential TFP gains from reallocation

3.5 Results

3.5.1 Aggregate misallocation

The first results to be presented are the ones of the quantification of misallocation at country level. In Figure 3.1 is reported the evolution of aggregate misallocation in the panel of countries under analysis, for the years 2011-2020. The beginning of this period shows the decrease in misallocation that followed to the Eurozone debt crisis, that have been documented to have had a 'cleansing effect' on markets, resulting on higher degrees of allocative efficiency in all the EU countries (Gamberoni *et al.*, 2016; Calligaris *et al.*, 2018). However, since the recovery around the year 2014, misallocation started increasing again in almost all the countries, with the sole exception of Czech-Republic, with a net decreasing trend, and Austria, with a more stable trend. Southern European countries, Spain and Portugal (with the exception of Italy), and Western European countries, such as Germany, France and Austria, are are the bottom of the misallocation distribution but with a rather stable and slowly increasing trend (with the exception of France, whose evolution is markedly increasing). The three Eastern European countries in the sample (Poland, Czech Republic and Slovenia), are the ones with an higher dispersion of manufacturing TFPR as expected by the lower levels of economic development with respect to their long-course EU members counterparts. The Italian misallocation figures, while perfectly in line with previous publications[75], strikes for its negativity, being second only to Poland in terms of inefficiency.

In addition, it is important to analyse these statistics in comparison with those of the estimated reallocation gains, as the latter derive not only by levels of aggregate misallocation but also by the productivity levels of the sectors under analysis. Indeed our measure of aggregate misallocation, the variance of log. TFPR, is weighted by the Value Added share of firms and sectors, where the output gains from reallocation also depends on the total factor productivity of firms and sectors (as shown in Eq.3.6.24 and 3.6.25). In this sense, while the rank order of countries' misal-

75 Calligaris *et al.* (2018), for the overlapping years of our analysis (2011-2013), provides almost identical estimates (between 0.53 and 0.56) for the Variance of the Italian manufacturing log TFPR.

location does not change substantially in the two figures (with the sole exceptions of Poland and Spain), it is interesting to notice that the re-establishment of allocative efficiency in Spain would entail much lower output gains than it would in Germany, notwithstanding the similar levels of aggregate misallocation measured in the two countries: a signal that TFPR dispersion in Germany concentrates in firms or sectors with better productive potential than in Spain. Similarly, Poland, while exhibiting the largest TFPR variance in the group, would rank only third in terms of reallocation gains. The analysis shows that, if all the variance in TFPR was to be equalized within-sectors, the manufacturing productivity and output of the EU countries could be increased by 53% (in Spain) and 76-84% (in Slovenia, Czech Republic and Poland), with an average for the whole period and panel of countries of 69 percentage points[76].

It should also be noted that, in 2020, the first year of the Covid-19 crisis, all countries suffered from a net increase in misallocation which however, with the exception of Poland, Austria and Portugal, resulted in lower potential reallocation gains attributable to a general drop in average productivity. Notwithstanding the general cleansing effect that crisis are deemed to have in the medium term, the short-term impact of the pandemic on TFP and factors allocation has been that of increasing distortions and decreasing productivity (and potential output gains with it), likely through the wiping-off of both more vulnerable and averagely more productive young and micro firms[77]. However, for the pandemic unleashed in different

76 The estimates of output gains from reallocation could be inflated in presence of measurement error, e.g. in case of multiproduct firms with uncorrelated revenues among products (Bils *et al.*, 2021). As such, these figures should be taken with caution. However, they are again in line with previous publications (Calligaris *et al.*, 2018).

77 When firms that exited the market at any moment (Panel B1-B2 of Figure 3.9), or micro-firms (Panel C1-C2 of Figure 3.9) are excluded from the sample as a robustness check (see Figure 3.9), the drop in potential reallocation gains registered for 2020 is reported as less intense in

moments in each country during 2020, and since we are not able to observe the following years, a study on the effects of such crisis cannot be developed properly with these data.

3.5.2 Regional and by group decomposition

The main aim of this exploratory study is to verify to what extent misallocation can be explained at regional level. The first step in order to do so, is to decompose the country-level measure of misallocation, *i.e.* the weighted aggregation of the within-sector variance of $TFPR$, in its within and between-group components (Eq.3.6).

		Between-group, % of Var. log TFPR					Var. log TFPR	N. of firms
EU Area	Country	Size	Age	NUTS1	NUTS2	NUTS3	Aggregate	
Western	France	8.99%	7.05%	9.92%	11.76%	**17.05%**	0.616	32049
Western	Germany	9.00%	5.88%	9.05%	**14.14%**	**20.01%**	0.657	11286
Western	Austria	**12.57%**	**14.19%**	**14.86%**	**14.86%**	**25.35%**	**0.698**	1327
Southern	Spain	5.95%	7.99%	7.18%	9.05%	9.05%	0.558	58584
Southern	Portugal	9.43%	**11.34%**	7.11%	8.18%	10.41%	0.621	20659
Southern	Italy	**14.07%**	**11.31%**	8.03%	9.34%	13.97%	**0.710**	84720
Eastern	Poland	**25.12%**	**15.89%**	**15.59%**	**18.42%**	**27.97%**	**0.773**	11246
Eastern	Slovenia	**21.29%**	**16.68%**	-	**13.50%**	**18.51%**	**0.804**	2090
Eastern	Czech Republic	6.69%	5.34%	-	9.18%	11.31%	**0.792**	4324
	Total	*12.57%*	*10.63%*	*10.47%*	*12.05%*	*17.07%*	*0.657*	*226285*

Table 3.1. Share of the between-group component on the overall dispersion of $TFPR$: period-averages, 2011-2020

To synthesise the results of the by-group $TFPR$ variance decomposition, the between-group component share on total misallocation, calculated for each grouping

all countries. This indicates that part of the lower reallocation gains shown in Fig. 3.2 should be attributable to the exit of above average revenue-productive (*i.e.* undersized) micro-firms during the first pandemic year.

(firms' age, size, and location at NUTS1, NUTS2, and NUTS3 level), are reported in Table 3.1, where above-average levels of the between-group component share on total misallocation are highlighted in bold. Age and size are mainly exploited in Table 3.1 as reference categories: given that they are related to the firm's extent of information incompleteness and to collaterals ownership, affecting their access and price of debt and equity capital (Midrigan and Xu, 2014), these dimensions are considered as two main predictors of allocative efficiency. Moreover, these two dimensions also concerns the application of a number of size-related policies determining firms' differential exposure to employment protection schemes, incentives and taxation which may directly affect decisions on the capital-labour ratio (Dias *et al.*, 2020).

The between-group component calculated for size and age accounts respectively for an average of a 12.6% (ranging 6%-25% for different countries) and of a 10.6% (5.3%-16.7%) of the country-level variances in our panel. Notably, the between-group components calculated for firms location at NUTS2 (12% on average, ranging 8.2%-18.4%) and NUTS3 (17%, ranging 9%-28%) level, are shown to explain a similar and often larger share of aggregate misallocation than do age and size. NUTS1 location, with a between-group component ranging among 7.1% and 15.6% of the aggregate misallocation, is found to be the weakest predictor of aggregate TFPR dispersion among the locational-groupings (generalizing previous punctual results by Calligaris *et al.* (2018) for Italy), but still with a comparable value to the by-age component. To frame these results, it's worth pointing out that, while one of the criteria for the definition of NUTS regions[78] was the mirroring of the territorial ad-

78 The European Nomenclature of Territorial Units for Statistics (NUTS) was designed by the Regulation (EC) N. 1059/2003 of the European Parliament and of the Council of 26 May 2003, and successively modified by the Commission Regulation (EU) 2016/2066 of 21 November 2016, follows three principles, namely: i) population thresholds (to ensure size-comparability among

ministrative division of each country, the existence of population thresholds for each level of aggregation, together with the great variety in population sizes and in pre-existing territorial and administrative structures across members states, derived into a certain variability in the nature of regions belonging to each classification level. In particular, NUTS1 level units tend to serve a merely statistical purpose, with very few exceptions (in our sample, Germany and, since 2016, France). Instead, NUTS2 and NUTS3 level unit boundaries tend to match those of administrative regions in most countries, as shown in Table 3.11 in the Appendix). Following our hypothesis on the role of local institutions, the lower explanatory power of the between group component measured at NUTS1 level could depend on the equal lack of administrative salience of this type of unit. To investigate this hypothesis, a dummy accounting for the existence of administrative authority in each country-level-year observation is introduced in the econometric analysis of regional misallocation.

To provide an intuition on the time and spatial distribution of misallocation and reallocation gains at sub-national level, these results are mapped in Fig. 3.3 and Fig. 3.4 for the years 2013, 2016 and 2019. Particularly, in Fig. 3.3 it is possible to seize the different trends that misallocation followed in the three broad EU-regions: it remained fairly stable in Southern-countries (at all levels of territorial aggregation), it increased in Western countries, and decreased in two Eastern ones. Furthermore, through the mapping of the regional decomposition of the dispersion in TFPR and of reallocation gains (Fig. 3.4), it is possible to appreciate some sign of spatial convergence in misallocation levels, both within and across countries, that should be specifically investigated in future dedicated contributions. To shed a light on the potential drivers behind these figures, I will now exploit the regression framework

regions); ii) the mirroring of the territorial administrative division of the Member States, and iii) regular amendments should occur not more often than every three years.

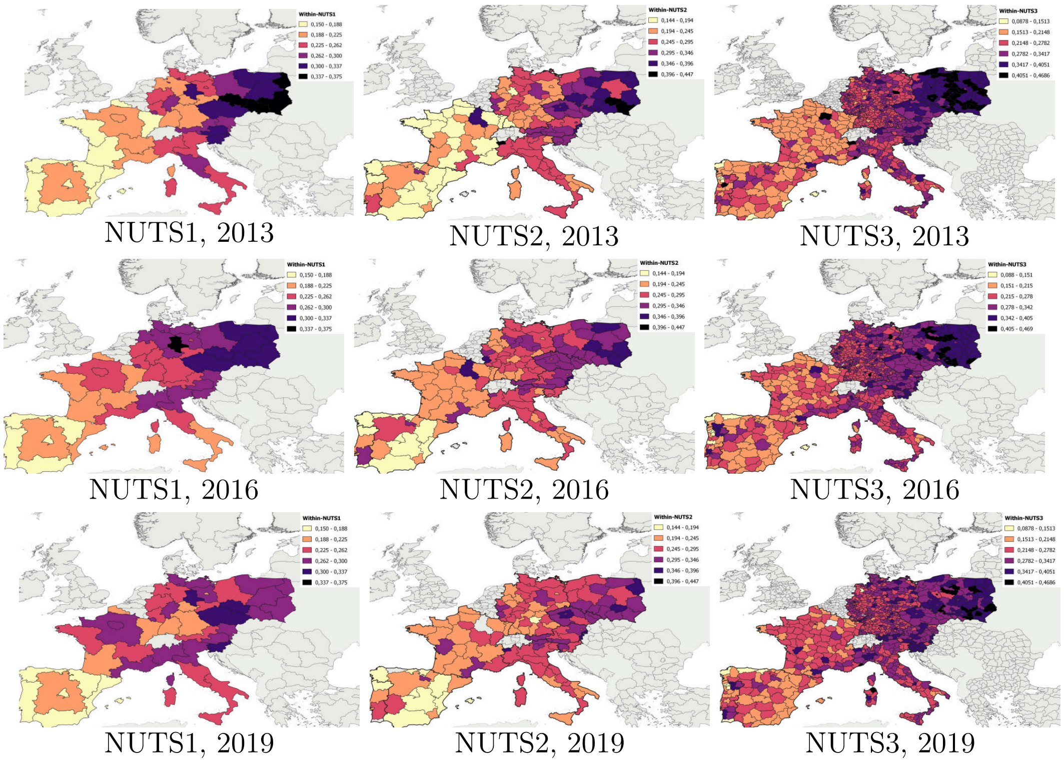


Figure 3.3. Evolution of within-region misallocation.

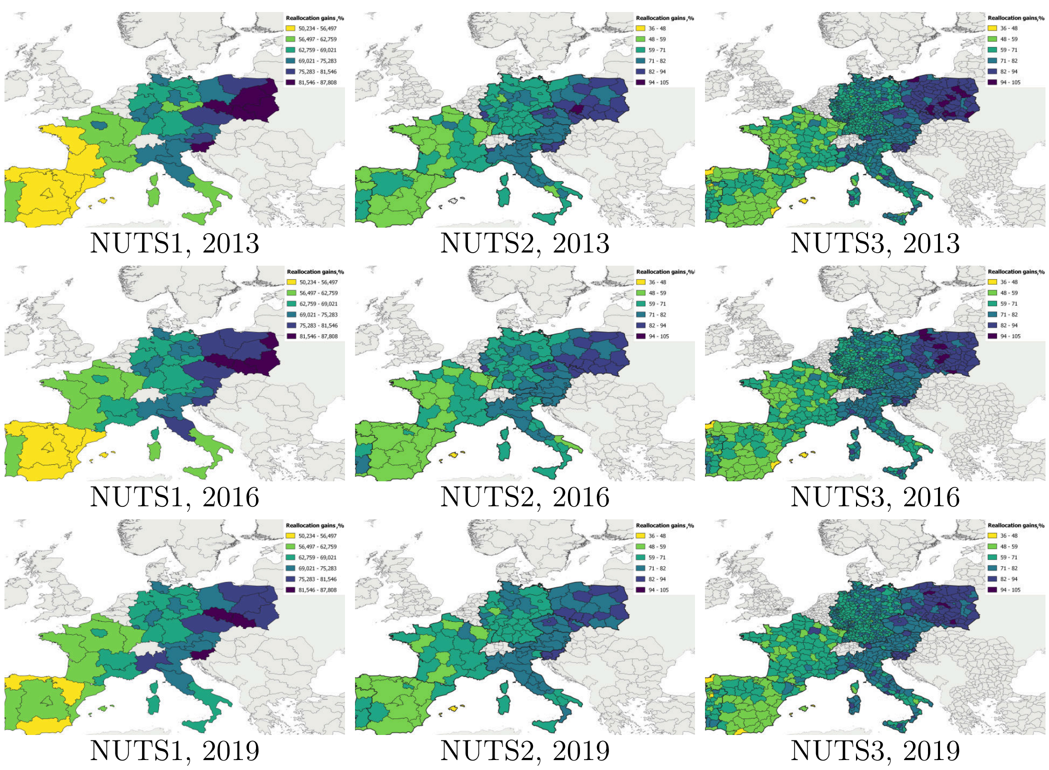


Figure 3.4. Evolution of regional reallocation gains, % of manufacturing output.

detailed in Section 3.3.2.

3.5.3 The impact of markers on within and across countries and regions misallocation

In Table 3.2-3.4 are reported the results of the regression model in Eq.3.9, where the role of various regional characteristics on three distinct dependent variables are evaluated through a two-way fixed effects estimator.

In columns 1 to 4 of Tables 3.2-3.3, the within-group misallocation, measured at country-NUTS0 (column 1), NUTS1 (column 2) NUTS2 (column 3) and NUTS3 level (column 4) is regressed with respect to a set of specific firm-level characteristics measured at regional level. In columns 5-8, the same model is estimated with respect to the reallocation gains in each territory, in order to analyse the role of each specific marker on the productivity and output losses, in addition to that on misallocation. Finally, columns 9-11 will portray the results on the role of the same characteristics on the share of aggregate misallocation represented by the between-group component, in order to evaluate how these variables affect disparities in within-country across-regions allocative efficiency. Please note that this last dependent variable is not available at country (NUTS0) level, as it refers to the subnational across-regions misallocation component, and that the number of observations in each model match the number of regions in our sample[79].

79 In the 9 countries under analysis, there are 50 Nuts1, 156 Nuts2, and 839 Nuts3 regions, as reported in Table 3.11. However, firms located in the Overseas France (which include the territories of Martinique, Guadeloupe, La Réunion, New Caledonia), in Spanish Canary Islands and in the Portuguese Azores Archipelago, were excluded from the sample which is therefore constituted by 47 Nuts1, 139 Nuts2 and 812 Nuts3 regions.

Firm characteristics

The first set of misallocation markers is that of firms characteristics, such as size, age, patenting activity and ownership, all expressed in terms of the share of firms with a specific characteristic located within each territory. Panel A of Table 3.2, dedicated to the role of firms' size, shows that increasing shares of micro firms are significantly correlated with lower degrees of misallocation at all levels of territorial aggregation, while the opposite is true for larger firms. Output gains, proportional to misallocation and value added, are significantly and negatively influenced by the share of small firms. To understand this result, it must account for the fact that large firms, being the excluded category, are the ones producing the highest output (about 70% of the total Value Added, as can be inferred by Table 3.6), while the opposite is true for micro firms. Larger shares of micro firms are also negatively correlated with disparities across regions and this is true also for small ad medium-sized firms, indicating that regional differences are likely to be driven by the location of large firms (representing the omitted group in the regression to avoid perfect collinearity). These results hold at all levels of territorial aggregation, indicating that countries and regions with greater shares of large firms tend to be both more inefficient and more unequal, matching previous country-specific results[80].

Turning to the role of the average age of firms, analysed in Panel B of Table 3.2, greater shares of mature firms are found to be associated with larger degrees of misallocation at all levels of territorial aggregation. The role of age is positive and significant in terms of reallocation gains, indicating that aggregate output could be

80 In particular, Calligaris *et al.* (2018) found misallocation in Italy to be stronger and to have increased the most among big firms. The authors, by analysing the relationship between the by-size sectoral misallocation and the speed of technological change, explained this result with the stronger technological frontier shocks faced by the average large firm.

	Misallocation				Reallocation gains (% output)				Between group component		
	NUTS0	NUTS1	NUTS2	NUTS3	NUTS0	NUTS1	NUTS2	NUTS3	NUTS1	NUTS2	NUTS3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: firms' size											
Micro firms	-0.224**	-0.194***	-0.155***	-0.067***	-15.372	-16.319***	-11.944***	-4.469***	-0.195***	-0.105***	-0.004
share	(0.031)	(0.000)	(0.000)	(0.000)	(0.1741)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.362)
Small firms	-0.091	-0.080	-0.063***	-0.011	-6.936	-8.812	-6.120***	-0.824	-0.161***	-0.098***	0.047***
share	(0.579)	(0.13613)	(0.003)	(0.194)	(0.639)	(0.338)	(0.007)	(0.424)	(0.002)	(0.000)	(0.000)
Medium firms	-0.103*	-0.043***	-0.024	0.006	-20.241	-11.243***	-7.139**	0.894	-0.234***	-0.124***	0.012***
share	(0.096)	(0.000)	(0.438)	(0.539)	(0.564)	(0.000)	(0.014)	(0.464)	(0.000)	(0.000)	(0.000)
Constant	0.376***	0.343***	0.334***	0.298***	80.385	77.742***	76.000***	71.787***	0.284***	0.234***	0.187***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	90	470	1,390	8,111	90	470	1,390	8,111	460	1,390	8,111
R-squared (within)	0.539	0.351	0.178	0.051	0.291	0.165	0.076	0.029	0.268	0.262	0.336
N. of regions	9	47	139	812	9	47	139	812	46	139	812
Panel B: firms' age											
Age, log	0.116**	0.110*	0.092***	0.035***	-0.011***	3.925**	3.803***	2.554***	-0.012*	-0.012***	-0.016***
mean	(0.026)	(0.073)	(0.000)	(0.000)	(0.000)	(0.039)	(0.002)	(0.000)	(0.0530)	(0.005)	(0.000)
Constant	-0.110	-0.102***	-0.032	0.167***	67.969***	54.309***	56.255***	62.365***	0.147***	0.173***	0.254***
	(0.159)	(0.000)	(0.312)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	90	470	1,390	8,111	90	470	1,390	8,111	460	1,390	8,111
R-squared (within)	0.432	0.317	0.178	0.047	0.143	0.107	0.060	0.027	0.125	0.223	0.317
N. of regions	9	47	139	812	9	47	139	812	46	139	812
Panel C: patenting											
Patenting	0.305	0.263***	0.215***	0.042***	-0.884	13.152***	11.224***	2.783**	0.042***	0.042***	-0.022***
share	(0.153)	(0.000)	(0.000)	(0.000)	(0.388)	(0.000)	(0.000)	(0.023)	(0.007)	(0.001)	(0.000)
Constant	0.194***	0.192***	0.215***	0.268***	68.085***	64.073***	66.080***	69.935***	0.099***	0.128***	0.209***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	90	470	1,390	8,111	90	470	1,390	8,111	460	1,390	8,111
R-squared (within)	0.460	0.300	0.144	0.028	0.143	0.138	0.066	0.017	0.130	0.226	0.311
N. of regions	9	47	139	812	9	47	139	812	46	139	812
Panel D: firms' ownership											
Govt.-owned	3.300***	0.840***	0.608	0.029	188.119***	92.182***	93.005	2.047	-0.537***	-0.063	-0.023
share	(0.000)	(0.000)	(0.229)	(0.686)	(0.000)	(0.000)	(0.118)	(0.795)	(0.000)	(0.713)	(0.126)
Equity-owned	1.640	0.822***	0.811***	0.086**	27.759**	27.697***	31.370***	5.785	0.024***	-0.255***	0.028***
share	(0.959)	(0.007)	(0.000)	(0.012)	(0.017)	(0.000)	(0.002)	(0.176)	(0.007)	(0.000)	(0.000)
Private-owned	0.026***	0.136**	0.095**	0.045***	1.134***	13.178***	7.720**	3.590***	0.045	0.075***	-0.035***
share	(0.000)	(0.011)	(0.024)	(0.000)	(0.000)	(0.000)	(0.047)	(0.001)	(0.479)	(0.000)	(0.000)
Foreign-owned	-0.018***	0.021***	0.030	-0.001	9.738***	8.609***	7.217**	-0.017	-0.025***	-0.052***	0.004
share	(0.000)	(0.000)	(0.309)	(0.864)	(0.000)	(0.000)	(0.013)	(0.983)	(0.000)	(0.000)	(0.248)
Constant	0.214***	0.198***	0.215***	0.270***	63.635***	60.836***	63.730***	69.942***	0.110***	0.147***	0.207***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	90	470	1,390	8,111	90	470	1,390	8,111	460	1,390	8,111
R-squared (within)	0.620	0.313	0.187	0.035	0.259	0.196	0.101	0.022	0.139	0.260	0.316
N. regions	9	47	139	812	9	47	139	812	46	139	812

Wild-bootstrapped clustered pvalues (at NUTS0-1 level) and cluster-robust pvalues (NUTS2-3) in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
All estimations include region and year FE.

Table 3.2. Firm-characteristics: Size, age, patenting activity and ownership of firms, on regional and aggregate misallocation, reallocation gains and between-group component.

consistently increased by reallocating inputs from older to younger firms. However, the age of the firm is negatively and significantly associated with across-regions disparities in allocative efficiency, which are thus likely to be be linked with startups concentration.

The analysis of the correlation between the success in patenting activities and the misallocation measures are reported in Panel C of Table 3.2. The coefficient of the share of patenting firms is found, consistently to Gordeev (2020), to be significantly and positively associated with greater misallocation and output gains at all levels of territorial aggregation (with the exception of output gains at country-level, where the coefficient of this regressor is not significant), indicating that patenting firms tend to be more productive than the average firm, and undersized. This variable is also positively and significantly associated with the between-group component measured at NUTS1 and NUTS2 level, and negatively at NUTS3 level.

Another characteristic under analysis is the ownership structure (Panel D of Table 3.2). In our data, countries and regions with greater shares of equity- and family-owned firms tend to be more inefficient. Individual or family-owned firms are also significantly more productive, such that allocating larger amounts of inputs to such firms would produce significant output gains. Potential output gains are also higher in regions with greater shares of foreign-owned firms. However, since this characteristic is not significantly correlated with the dispersion in TFPR at any level of territorial aggregation in our panel, this result says more about their productivity than it does about their relative allocative inefficiency.

Economic and political controls

The adoption of a two-way fixed-effects estimator allows to control for unobserved time- and regional-invariant heterogeneity and omitted variables: it is anyhow interesting to explore the relationship that exists between misallocation and some economic and political characteristics of the countries and regions analysed. While we are not able to control for the form of government (as none of the countries analysed changed their form of government during the period analysed, e.g. from

	Misallocation				Reallocation gains (% output)				Between group component		
	NUTS0	NUTS1	NUTS2	NUTS3	NUTS0	NUTS1	NUTS2	NUTS3	NUTS1	NUTS2	NUTS3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: administrative regions											
Administrative region, dummy		0.031**	-0.033***	0.005		1.313***	-1.333***	0.416	-0.001***	0.014***	0.014***
		(0.011)	(0.000)	(0.490)		(0.000)	(0.000)	(0.321)	(0.000)	(0.000)	(0.000)
Constant		0.230***	0.286***	0.274***		65.700***	69.410***	70.243***	0.108***	0.122***	0.192***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations		450	1,390	8,111		450	1,390	8,111	450	1,390	8,111
R-squared (within)		0.171	0.115	0.017		0.092	0.050	0.013	0.124	0.247	0.309
N. regions		45	139	812		45	139	812	45	139	812
Panel b: agglomeration											
L1. Population density, log	0.000	-0.000	-0.000	0.000	-0.002	-0.001	-0.005**	-0.003	-0.000*	0.000	0.000
	(0.827)	(0.131)	(0.879)	(0.793)	(0.690)	(0.79579)	(0.048)	(0.224)	(0.063)	(0.837)	(0.978)
Constant	0.144*	0.254***	0.257***	0.276***	68.405***	66.861***	69.804***	71.212***	0.116***	0.134***	0.203***
	(0.054)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	81	423	1,245	7,260	81	423	1,245	7,260	414	1,245	7,260
R-squared (within)	0.103	0.045	0.017	0.013	0.155	0.064	0.027	0.009	0.103	0.126	0.283
N. regions	9	47	139	812	9	47	139	812	46	139	812
Panel C: GDP											
L1.GDP, p.c. pps., log.	-0.103***	-0.068***	-0.085***	-0.080***	-12.403***	-9.465***	-9.237**	-7.317***	0.040***	-0.038	-0.044***
	(0.016)	(0.011)	(0.008)	(0.000)	(0.000)	(0.003)	(0.010)	(0.000)	(0.000)	(0.187)	(0.000)
Constant	1.283	0.938	1.116***	1.090***	192.692***	162.943***	162.104***	145.442***	-0.296	0.519*	0.656***
	(0.110)	(0.162)	(0.001)	(0.000)	(0.009)	(0.002)	(0.000)	(0.000)	(0.503)	(0.074)	(0.000)
Observations	76	383	1,141	6,820	76	383	1,141	6,820	374	1,141	6,820
R-squared (within)	0.190	0.050	0.035	0.012	0.187	0.061	0.029	0.009	0.104	0.139	0.387
N. regions	9	47	139	812	9	47	139	812	46	139	812
Panel D: Market concentration											
L1.HH Index	0.176***	0.026***	0.003	-0.000	5.250***	2.800***	0.651*	0.000	0.006***	0.000	0.001***
	(0.000)	(0.000)	(0.529)	(0.642)	(0.000)	(0.000)	(0.091)	(0.996)	(0.000)	(0.960)	(0.003)
Constant	0.212***	0.234***	0.254***	0.278***	66.924***	65.566***	67.827***	70.598***	0.106***	0.135***	0.203***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	90	470	1,390	8,101	90	470	1,390	8,101	460	1,390	8,101
R-squared (within)	0.188	0.074	0.023	0.017	0.156	0.102	0.038	0.012	0.121	0.218	0.308
N. regions	9	47	139	812	9	47	139	812	46	139	812

Wild-bootstrapped clustered pvalues (at NUTS0-1 level) and cluster-robust pvalues (NUTS2-3) in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
All estimations include region and year FE.

Table 3.3. Economic and political controls on regional and aggregate misallocation, reallocation gains and between-group component.

federal to unitary) one related dimension that exhibits some variation both in time and space is the typology of each level of NUTS regions. As argued before, and shown in Table 3.11, NUTS territories may or may not coincide with political regions with some degree of administrative power and autonomy, being merely statistical groupings otherwise. Controlling for this characteristic enables the model to uncover that countries where NUTS2 regions have some degree of administrative power are significantly less misallocated than others. Administrative decentralization at NUTS2 and NUTS3 level is also correlated with larger disparities across regions. However, misallocation and output gains are lower in NUTS2 regions with administrative powers. NUTS2 level regions with administrative powers are thus more efficient but also more unequal than their counterparts in our panel. Both signals points towards the role of institutions, that will be analysed later in the text. Finally, since only two countries, and one of them (France) only since 2018, present non-statistical NUTS1 regions, the positive and significant coefficient for misallocation at this level of aggregation should be taken as less of a general results.

Another relevant economic control (Panel B, Table 3.3) is population density, which can be considered as a raw measure of agglomeration. While agglomeration is expected to increase efficiency, through those sharing, matching and learning mechanisms (Duranton and Puga, 2004) that facilitate input sourcing and productivity, this variable is only significant in terms of lower output gains at NUTS2 level. More densely populated countries are found to also be less unequal in their level of misallocation at NUTS1 level.

Turning to the role of per capita GDP, wealthier countries display significantly higher levels of allocative efficiency at NUTS2 and NUTS3 level, lower potential gains from reallocation, and lower shares of across-regions misallocation at NUTS3 level. While the negative correlation of development levels and efficiency is a standard result, the negative correlation with the between-group component is, if not novel (Misch and Saborowski, 2020), relevant to be produced in the European context.

Finally, in Panel D of Table 3.3 I examine the role of *market concentration*, measured as the territorial yearly mean of the sectoral Herfindahl-Hirschman Index. Market concentration can be used as a proxy for market power, whose role on misallocation has been established (Asker *et al.*, 2019) as to be strongly detrimental of allocative efficiency. In line with the literature, I find market concentration to significantly correlate with larger log TFPR dispersion at country and NUTS1 level and with potential output gains at NUTS1 and NUTS2 level. Moreover, it could lead to higher aggregate misallocation through the additional channel of its positive association with the between-group component at NUTS3 level.

Institutional quality

Lastly, I explore the role of institutional quality as a potential driver of systematic differences in allocative efficiency within and across regions and countries.

Since, as argued in Section 3.3 , this variable could be endogenous[81], in Table 3.4 the results of the two-way fixed-effect estimations are compared to those of the second stage of a 2SLS two-way fixed-effects IV estimation. First stage statistics and tests regarding the validity of the instrumental approach are added to each IV model column[82].

The quality of public spending is deemed to have at least one direct link with mis-

81 This hypothesis seems confirmed by the Durbin-Wu-Hausman endogeneity test, which results are reported for each model in Table 3.4, where, with few exceptions of its evaluation at NUTS0 level, the null-hypothesis of exogeneity should be rejected.

82 Instrumental variables must display certain properties in order to be valid, *i.e.* for their coefficients to be unbiased: exogeneity, and strength. For all IV models in Table 3.4 are reported the Kleiberg-Paap LM statistic and its relative P-value, an underidentification statistic used to measure the significant correlation of the instruments with respect to the endogenous excluded regressor. The null hypothesis of no-correlation is rejected, indicating the relevance of the instruments, in almost all the models, with the sole exceptions of the ones in which the excluded endogenous regressors were the PSQ sub-dimensions of the Efficiency score (Panel D), or Competition (Panel C) when measured at NUTS0 level. Finally, the strength of the instrumental variable is tested through the comparison of the Kleiberg-Paap F statistic (robust to clustered errors and valid with multiple regressors) with the Montiel-Olea & Pflueger (2013) 2SLS thresholds, as these are robust to weak-instruments and heteroskedasticity as in this application. With the exception of the NUTS0 level models, the Kleiberg-Paap F statistic exceeded these thresholds, establishing the cluster and heteroskedasticity robust instruments' strength to be above a bias level at 10% confidence level. This is an important result, given that the bias introduced by weak instruments could otherwise easily exceed that of the inclusion of endogenous regressors.

	Misallocation						Reallocation gains (% output)						Between group component			
	NUTS0		NUTS1		NUTS2		NUTS0		NUTS1		NUTS2		NUTS1		NUTS2	
	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV
Panel A: Public spending quality Index																
L4. PSQ Index	-0.001	-0.004*	-0.001*	-0.005**	-0.000**	-0.007***	-0.119*	-0.222	-0.048	-0.485**	-0.020	-0.677***	-0.000	-0.003	-0.000**	-0.003**
	(0.196)	(0.086)	(0.0530)	(0.010)	(0.045)	(0.003)	(0.072)	(0.194)	(0.201)	(0.027)	(0.416)	(0.006)	(0.630)	(0.274)	(0.032)	(0.011)
Constant	0.285***		0.296***		0.289***		75.806***		70.814***		70.293***		0.117***		0.160***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Observations	80	80	415	415	1,168	1,167	80	80	415	415	1,168	1,167	406	406	1,168	1,167
R-squared (within)	0.071		0.056		0.028		0.121		0.057		0.041		0.130		0.265	
N. regions	9	9	47	47	137	136	9	9	47	47	137	136	46	46	137	136
Durbin-Wu-Hausman P-value		0.175		0.0942		5.80e-05		0.535		0.0137		0.000211		0.0650		0.00643
Kleibergen-Paap F statistic		4.602		9.728		7.353		4.602		9.728		7.353		8.690		7.353
Olea-Pflueger 10% F threshold		9.252		6.144		6.517		7.691		6.262		5.875		5.551		5.836
Kleibergen-Paap LM statistic		8.255		18.83		14.03		8.255		18.83		14.03		16.96		14.03
Kleibergen-Paap LM P-value		0.0161		8.14e-05		0.000899		0.0161		8.14e-05		0.000899		0.000208		0.000899
Panel B: Public spending quality: Transparency																
L4. Transparency score	-0.000	-0.006*	-0.000	-0.004***	-0.000	-0.003***	-0.059	-0.225	-0.008	-0.259**	0.003	-0.349***	-0.000	-0.001	-0.000***	-0.002***
	(0.535)	(0.078)	(0.806)	(0.007)	(0.311)	(0.003)	(0.274)	(0.260)	(0.659)	(0.033)	(0.832)	(0.002)	(0.532)	(0.580)	(0.006)	(0.005)
Constant	0.272***		0.257***		0.272***		72.936***		68.576***		68.946***		0.114***		0.159***	
	(0.001)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Observations	80	80	415	415	1,168	1,167	80	80	415	415	1,168	1,167	406	406	1,168	1,167
R-squared (within)	0.064		0.048		0.024		0.093		0.053		0.040		0.131		0.270	
N. regions	9	9	47	47	137	136	9	9	47	47	137	136	46	46	137	136
Durbin-Wu-Hausman P-value		0.0323		0.00159		0.000148		0.442		0.0155		7.92e-05		0.670		0.00341
Kleibergen-Paap F statistic		2.720		9.087		10.32		2.720		9.087		10.32		8.620		10.32
Olea-Pflueger 10% F threshold		9.942		5.184		6.803		9.966		5.350		6.792		5.836		7.111
Kleibergen-Paap LM statistic		4.817		14.66		19.19		4.817		14.66		19.19		13.94		19.19
Kleibergen-Paap LM P-value		0.0899		0.000655		6.81e-05		0.0899		0.000655		6.81e-05		0.000938		6.81e-05
Panel C: Public spending quality: Competition																
L4. Competition score	-0.000	-0.010	-0.000*	-0.008***	-0.000	-0.009**	-0.003	-0.437	-0.025	-0.578**	-0.015	-0.453*	-0.000	-0.002	0.000	-0.000
	(0.763)	(0.176)	(0.072)	(0.004)	(0.311)	(0.014)	(0.966)	(0.271)	(0.367)	(0.012)	(0.497)	(0.091)	(0.280)	(0.396)	(0.435)	(0.608)
Constant	0.255***		0.287***		0.277***		69.407***		69.810***		70.180***		0.125***		0.136***	
	(0.001)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Observations	80	80	415	415	1,168	1,167	80	80	415	415	1,168	1,167	406	406	1,168	1,167
R-squared (within)	0.056		0.054		0.024		0.075		0.055		0.041		0.133	-0.055	0.261	
N. regions	9	9	47	47	137	136	9	9	47	47	137	136	46	46	137	136
Durbin-Wu-Hausman P-value		0.0273		0.00565		3.88e-06		0.198		0.0151		0.0217		0.365		0.626
Kleibergen-Paap F statistic		1.276		4.902		4.446		1.276		4.902		4.446		5.982		4.446
Olea-Pflueger 10% F threshold		7.385		6.108		7.528		7.691		6.096		6.709		5.868		6.284
Kleibergen-Paap LM statistic		2.884		9.822		8.647		2.884		9.822		8.647		11.85		8.647
Kleibergen-Paap LM P-value		0.236		0.00736		0.0133		0.236		0.00736		0.0133		0.00267		0.0133
Panel D: Public spending quality: Efficiency																
L4. Efficiency score	0.001*	-0.014	0.000	-0.002	0.000	-0.061	0.043	-0.621	-0.008	-0.973	0.008	-1.555	0.000	-0.011	-0.000*	0.009
	(0.097)	(0.418)	(0.714)	(0.619)	(0.395)	(0.759)	(0.573)	(0.448)	(0.819)	(0.273)	(0.654)	(0.767)	(0.403)	(0.386)	(0.074)	(0.766)
Constant	0.212***		0.251***		0.260***		68.055***		68.274***		68.953***		0.102***		0.146***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Observations	80	80	415	415	1,168	1,167	80	80	415	415	1,168	1,167	406	406	1,168	1,167
R-squared (within)	0.098		0.048		0.024		0.081		0.053		0.041	0.332	0.132		0.262	
N. regions	9	9	47	47	137	136	9	9	47	47	137	136	46	46	137	136
Durbin-Wu-Hausman P-value		0.0122		0.781		4.68e-05		0.197		0.0299		0.277		0.000747		0.364
Kleibergen-Paap F statistic		0.418		0.667		0.0473		0.418		0.667		0.0473		0.679		0.0473
Olea-Pflueger 10% F threshold		11.52		7.652		5.345		11.26		8.086		5.539		7.601		5.425
Kleibergen-Paap LM statistic		0.962		1.382		0.0956		0.962		1.382		0.0956		1.392		0.0956
Kleibergen-Paap LM P-value		0.618		0.501		0.953		0.618		0.501		0.953		0.499		0.953
Panel E: Public spending quality: Control of corruption																
L4. Control of corruption score	-0.001**	-0.001*	-0.000***	-0.002*	-0.000**	-0.004**	-0.062***	-0.083	-0.005	-0.274*	-0.004	-0.388*	-0.000	-0.003	-0.000	-0.002*
	(0.035)	(0.083)	(0.006)	(0.060)	(0.036)	(0.049)	(0.003)	(0.192)	(0.745)	(0.070)	(0.657)	(0.054)	(0.911)	(0.100)	(0.951)	(0.063)
Constant	0.279***		0.285***		0.275***		73.286***		68.436***		69.533***		0.109***		0.142***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Observations	80	80	420	420	1,236	1,236	80	80	420	420	1,236	1,236	411	411	1,236	1,236
R-squared (within)	0.100		0.066		0.024		0.143		0.052		0.037		0.130		0.257	
N. regions	9	9	47	47	139	139	9	9	47	47	139	139	46	46	139	139
Durbin-Wu-Hausman P-value		0.546		0.589		7.30e-05		0.763		0.0117		0.000128		0.00302		0.00111
Kleibergen-Paap F statistic		6.153		5.249		2.379		6.153		5.249		2.379		4.340		2.379
Olea-Pflueger 10% F threshold		12.62		7.494		7.775		12.20		7.330		7.520		7.100		7.594
Kleibergen-Paap LM statistic		8.693		10.63		4.737		8.693		10.63		4.737		8.779		4.737
Kleibergen-Paap LM P-value		0.0129		0.00491		0.0936		0.0129		0.00491		0.0936		0.0124		0.0936

Wild-bootstrapped clustered pvalues in parenthesis for NUTS0-1 level models. jackknifed for 2SLS models (Young *et al.*, 2022), and cluster-robust pvalues in NUTS2-3 models.
Pvalues in parentheses (*** p<0.01, ** p<0.05, * p<0.1). All estimations include region and year FE.

Table 3.4. Public spending quality on regional and aggregate misallocation, reallocation gains and between-group component.

allocation, that of reducing distortions in regulations and provisions, and numerous indirect ones (see Lasagni *et al.*, 2015 for a comprehensive review), since through the open and fair public investments procedures it is possible to influence competition, thus firms selection and average productivity with it (Syverson, 2011). A negative and significant coefficient for the measures of public spending quality with respect to misallocation (or output gains), means that higher levels of institutional quality are associated with lower levels of allocative inefficiency (or output losses). When regressed with respect to the between-group component instead, the a negative coefficient for a public spending quality measure means that regions with higher levels of institutional quality have a lower degree of disparity in allocative efficiency.

As it can be appreciated in Panel A of Table 3.4, the coefficient of the Public Spending Quality (PSQ) Index is always negative with respect to all the dependent variables under analysis[83], confirming the hypothesis that local institutions may affect positively the allocative efficiency of firms within and across territories. However, this is mainly significant once instrumented.

Fazekas and Czibik (2021) adverts that the general PSQ Index is composed by a number of dimensions whose effect and information may sometimes overlap (and are positively correlated among each other), such that the above illustrated coefficients might be overestimated. To shed a light on the mechanisms behind the role of Institutional Quality while testing the robustness of the previous results, in Panel B-E of Table 3.4 I analyse the correlation of each sub-dimension of the Public Spending Index with respect to the three dependent variables.

83 Since the PSQ Index is available at NUTS3 level only as a time-invariant mean for the period 2007-2015, we are not able to estimate it through a two-way fixed effect estimator at this level of territorial aggregation. The results for the time-invariant PSQ Index at NUTS3 level obtained with NUTS2-level regional fixed effects, available upon request, showed largely significant and negative coefficients with respect to all dependent variables.

The transparency score (being based on the share of published call for tenders on all awarded contracts, the proportion of open procedures, the voluntary reporting and the completeness of information provided in each call) represents the ability for a larger pool of firms to participate to public tenders, thus favoring a fair and more competitive business environment. Accordingly, this dimension displays negative and, once instrumented, significative coefficients with respect to the dispersion of TFPR, to output gains and to the between-group component at almost all NUTS levels.

The second sub-dimension of the PSQ Index is the Competition score, accounting for the number of bids submitted and the share of awarded contracts to non-local supplier. The competition score, somehow intertwined with the Transparency one, is expected again to have a positive effect on allocative efficiency, confirmed in our data by the negative and, at sub-national level, significant coefficients with respect to misallocation and output gains. Nevertheless, this dimension is not significantly correlated with the across-regions disparities, possibly indicating that the benefit and losses derived from public-led competitive procedures are not geographically bounded.

The third sub-dimension provided in the Fazekas and Czibik (2021)'s dataset is the Administrative Efficiency score, computed through the consideration of the decision-making speed, the share of tenders assigned the most economically advantageous bid, and the price savings with respect to the cost of comparable contracts. While the efficiency of public spending is can lead to welfare enhancements through savings and provision of better or wider public goods and service, it is not deemed to be linked directly with firms' specific allocative efficiency, if not through a general incentive through market selection. Indeed, while a positive effect on allocative efficiency is somehow confirmed in our data by the negative coefficients with respect to all the misallocation measures, these are not significant.

A last PSQ dimension is the Control of corruption, an indicator built to capture if, and to what extent, measures[84] were enacted in order to avoid the favoring of some connected firms with respect to others. This dimension is expected to be critically connected with misallocation and to partially correlate with open and fair competition. As such it should positively affect allocative efficiency through the selection channel and through an healthier business environment. This correlation is again confirmed at all levels of territorial aggregation and in terms of all the dependent variables, even its coefficient is less significant than that of transparency.

Overall, these negative and significant coefficients (and the absence of a significant and positive coefficient in Table 3.4) should be interpreted as evidence of the positive role that institutional quality could play in reducing misallocation within and across-regions.

3.5.4 A NUTS2-level horse-race model

One of the main flaws of the applied indirect methodology is that it does not allow to estimate a structural model. However, in order to provide some insights on the extent at which the magnitude and sign of the reported correlations could hold within a unique regression, this section provides a brief horse-race model exercise. In it, I focus on the NUTS2-level, given that these type of regions are found to display large degrees of allocative disparities, similarly to NUTS3 level ones, while disposing of panel data on the PSQ variable. The first four columns of Table 3.5 explore the im-

84 These measures are measured through the length and place of publication of the call for tenders, or the length and type of assignment, e.g. "without unusually high weights of non-quantitative evaluation" Fazekas and Czibik, 2021, p. 7.

pact of the progressive insertion of different categories of markers on the correlation with the within-region misallocation. Here the main regressors retaining the same sign and significance level with respect to Table 3.2 are the share of patenting firms, government and equity ownership. However, once all the economic controls and the Public Spending Quality index are included, the only significant regressor among them becomes the share of Government-owned firms. This implies that regional misallocation, once other industrial, structural and institutional characteristics are accounted for, tend to mainly correlate with higher degrees of political control over the economy, consistently with the abundant evidence of the role of cronyism on misallocation (García-Santana *et al.*, 2020). Firms' size and age however, do retain the same sign, significance and magnitude of their coefficients with respect to reallocation gains (columns 5-8, Table 3.5) and to between-group component (columns 9-12, Table 3.5). The coefficient of the Public Spending Quality Index, while maintaining the same expected sign, also remains (weakly) significant in terms of disparities and with respects to output gains, but not with respect to within-region misallocation. Economic controls, in turn, are the most unstable regressors, with market concentration gaining significance with respect to its individual inclusion (in Table 3.3), thus pointing to the high degrees of correlation with other variables in the model. All in all, the results concerning disparities between regions are the most novel and stable, and should receive further analysis through structural methodologies in order to quantify the role of each specific marker in them.

Robustness checks

To sustain the validity of the results, I perform a number of robustness checks.

To test for the role of the selection channel in the data, I quantify aggregate and regional misallocation on a balanced version of the dataset, *i.e.* excluding all

	Misallocation NUTS2				Reallocation gains (% output) NUTS2				Between group component NUTS2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV	2WFE	IV
Firms' characteristics												
Micro firms	-0.022	0.000	-0.005	-0.076	-4.125	-8.359	-3.063	-12.827*	-0.114***	-0.212***	0.116***	0.078*
share	(0.616)	(0.998)	(0.927)	(0.302)	(0.277)	(0.145)	(0.480)	(0.083)	(0.000)	(0.000)	(0.000)	(0.069)
Small firms	0.035	0.074	0.092*	0.025	3.726	0.693	9.855	7.173	0.109*	-0.111**	0.144***	0.119
share	(0.507)	(0.313)	(0.090)	(0.820)	(0.496)	(0.938)	(0.123)	(0.610)	(0.066)	(0.045)	(0.002)	(0.138)
Medium firms	-0.002	-0.041	-0.010	-0.107	-9.367**	-17.708***	-9.960**	-23.526***	-0.193***	-0.203***	-0.056***	-0.029
share	(0.964)	(0.432)	(0.845)	(0.116)	(0.014)	(0.000)	(0.023)	(0.000)	(0.000)	(0.000)	(0.000)	(0.336)
Age, log	-0.016	-0.035*	-0.004	-0.002	-7.677***	-10.768***	-4.384	0.299	-0.057***	-0.120***	-0.190***	-0.202***
mean	(0.434)	(0.097)	(0.915)	(0.966)	(0.003)	(0.000)	(0.285)	(0.946)	(0.000)	(0.000)	(0.000)	(0.000)
Patenting	0.195**	0.348**	0.220*	0.228	31.455***	34.684*	33.833**	48.928*	0.212	-0.374***	0.361***	-0.058
share	(0.035)	(0.028)	(0.063)	(0.287)	(0.005)	(0.096)	(0.013)	(0.087)	(0.166)	(0.000)	(0.004)	(0.660)
Govt.-owned	5.615***	6.116***	3.194*	5.534**	837.459***	731.545***	552.813***	468.959*	-1.661	8.016***	-3.971	5.204**
share	(0.000)	(0.000)	(0.052)	(0.012)	(0.000)	(0.001)	(0.001)	(0.070)	(0.338)	(0.000)	(0.105)	(0.015)
Equity-owned	0.819***	0.915	1.188**	1.198	10.959	86.768	22.575	-31.723	-0.690	1.558***	0.929*	2.610***
share	(0.008)	(0.128)	(0.018)	(0.241)	(0.776)	(0.289)	(0.655)	(0.789)	(0.134)	(0.001)	(0.091)	(0.000)
Private-owned	-0.052	-0.063	0.008	-0.100	-2.527	-7.248	3.797	-8.676	0.293***	0.129***	0.474***	0.448***
share	(0.322)	(0.288)	(0.907)	(0.459)	(0.660)	(0.427)	(0.567)	(0.576)	(0.000)	(0.006)	(0.000)	(0.003)
Foreign-owned	0.127***	0.116*	0.069	0.070	15.898***	10.705	13.078**	19.895*	-0.239***	-0.389***	-0.296***	-0.486***
share	(0.006)	(0.072)	(0.193)	(0.363)	(0.007)	(0.360)	(0.030)	(0.054)	(0.000)	(0.000)	(0.000)	(0.000)
Public spending quality												
L4. PSQ Index		-0.000		-0.002		-0.327*		-0.274		-0.003**		0.002
		(0.734)		(0.442)		(0.065)		(0.294)		(0.045)		(0.218)
Controls												
L1. Population density, log			-0.000	-0.000			-0.002	0.001			-0.000***	-0.000***
			(0.478)	(0.667)			(0.833)	(0.927)			(0.001)	(0.000)
L1.GDP, p.c. pps., log.			-0.065**	0.039			-2.497	6.413			-0.266***	-0.226***
			(0.048)	(0.527)			(0.544)	(0.467)			(0.000)	(0.000)
L1.HH Index			0.066**	0.013			8.513***	-3.540			0.047**	-0.005
			(0.020)	(0.803)			(0.001)	(0.503)			(0.014)	(0.895)
Constant	0.263***		0.901**									
	(0.000)		(0.018)									
Observations	1,390	1,168	1,141	949	1,390	1,168	1,141	949	1,390	1,168	1,141	949
R-squared (within)	0.918		0.918		0.965		0.965		0.896		0.934	
N. clusters	137	137	137	137	137	137	137	137	137	137	137	137

Cluster-robust pvalues at NUTS level, jackknifed for IV models.
Pvalues in parentheses (*** p<0.01, ** p<0.05, * p<0.1). All estimations include region and year FE.

Table 3.5. Horse-race model at NUTS2-level including all regressors.

firms that entered or exited the market during the 2011-2020 period. As it can be appreciated in the panels B1 and B2 of Figure 3.9 in the Appendix, these estimates are qualitatively and quantitatively similar to those obtained on the original sample. If anything, the dispersion of revenue productivity is higher in the balanced sample for France, meaning that it benefited from the selection channel through firms entry and exit more than the rest of the countries. To control for the potential bias that could derive from the low micro-firms coverage in gathered data, in an additional test I exclude all firms with less than 10 employees, and repeat the quantification exercise. As shown in Panels C1-C2 of Fig. 3.9, the dispersion in TFPR estimated without micro-firms is lower than that of the full sample (Panel A1-A2) in Italy and

Spain, two countries that reportedly suffers from disproportionate high shares of micro firms (Mas *et al.*, 2008; García-Santana *et al.*, 2020; Vacca, 2013). Since for other countries, with the exceptions of Portugal, very low shares of micro-firms were included in the original full sample, this result could also indicate an underestimation of aggregate misallocation for the rest of the panel's countries. However, since the misallocation trend was mainly unaffected even in countries with large micro firms coverage (Spain, Italy and Portugal), we can conclude that results are dynamically robust to size-related compositional effects.

3.6 Conclusions and future direction of research

This research has addressed the indirect quantification of factors misallocation at national and subnational level for nine European countries (Germany, Italy, France, Spain, Poland, Austria, Portugal, Slovenia, Czech-Republic), selected for their importance in manufacturing value-added, representative data-coverage, and for the external validity that their heterogeneity in terms of levels of development, form of government, and location in distinct areas of the EU (West, East and South) would provide.

The study's main aim was to ascertain, on a wider basis than in the extant literature, the importance of the regional dimensions of misallocation. The study's findings reveal two critical facets of productivity dispersion that require the attention of researchers in the field. Firstly, it is demonstrated that allocative efficiencies vary significantly across regions within all the countries under analysis, and that up to a quarter of the observed aggregate dispersion in revenue-based productivity is dependent on within-country imbalances. Secondly, the study establishes that the by location (NUTS1-3) between-group components of revenue-based productivity dispersion contribute more substantially to aggregate misallocation than their by-size or by-age counterparts. Furthermore, institutional quality, expressed in terms of the quality of public spending, and particularly in its transparency, control of corruption and competition subdimensions, was found to negatively and significantly correlate with all the dimensions of misallocation analysed. Both the hypothesis of this study, that across-regions disparities can significantly affect aggregate misallocation, and that these could be at least partly affected (and thus potentially manoeuvred) through the quality of public spending, have thus found some empirical support.

These findings are expected to stimulate further research on the topic of the spatial and sub-national dimensions of misallocation, and to provide guidance for policymakers seeking to reduce overall inefficiency, suggesting that this objective could be addressed by reducing disparities in allocative efficiency across regions.

Furthermore, the evidence presented can help identify which regional and firm characteristics, associated with misallocation, could yield greater output and productivity gains if targeted effectively by policies aimed at restoring efficiency.

For example, policies targeting reallocation across older firms would not result in significant gains as they tend to be less productive compared to their younger counterparts, even if mature firms are found as significantly inefficient. Instead, promoting the reallocation of factors across firms of different sizes, particularly among larger firms, could lead to considerable output increases. Furthermore, policies that promote the creation of new firms could help to reduce inefficiencies at the local or aggregate level as young and small firms are significantly associated with lower degrees of both the within and between-group components of misallocation[85].

85 While the search for specific potential ways to reduce misallocation was not the focus of this work, the extant literature suggest to a number of policy objectives, such as the reduction of

The work comes with several limitations to be acknowledged, some of which derive by the methodology, others by data availability. With regard to the former, the indirect methodology applied does not allow to claim any causality in the econometric analysis of misallocation markers. The adoption of this method was deemed necessary to produce evidence at the regional level with a good extent of general validity. In future country-specific studies, alternative methodologies[86] could be adopted in order to provide reassurance regarding the robustness and causality of specific misallocation markers included in the analysis. Furthermore, in the reference model of this methodology (Hsieh and Klenow, 2009), capital and output wedges are firm-specific, while the study relies on the hypothesis that regional-specific distortions could be at place. Formally including such variation in a structural model would allow to move a further step towards the quantification of specific sources of regional distortions. It is important to acknowledge another possible limitation related to the lack of distinction between tangible and intangible capital, mainly due to data-availability constraints. While this is a commonly adopted approach in the literature, dedicated studies have found misallocation to be more pronounced in sectors that rely heavily on intangible assets (Caggese and Perez-Orive, 2022; Gordeev, 2020). Therefore, the estimates of misallocation presented in this work could potentially be overestimated, particularly in regions and countries with a

frictions to trade, *i.e.* by decreasing import and export tariffs (Tito and Wang, 2021), the promotion of the access to credit especially to most financially constrained ones such as small, young and innovative firms, e.g. through capital market integration (Bau and Matray, 2023), or the reduction of regulatory-led rigidity to wage adjustments (Boeri *et al.*, 2021; Lashitew, 2016).

86 The recent extension of Hsieh and Klenow, 2009's framework proposed by David and Venkateswaran (2019) for example, would allow to disentangle the role of each of the markers analyzed. However, this would come at the cost of a less reliable quantification of aggregate and regional misallocation measures at the core of this contribution, as, being more data-demanding, it would have required large use of imputation.

higher proportion of firms operating in intangible-intensive sectors. Moreover, since previous studies have established misallocation to be higher in the service sector (see Dias *et al.*, 2020 or García-Santana *et al.*, 2020), future direction for research would be to evaluate the regional level of allocative efficiency of non-manufacturing sectors. The main contribution of this work has been that of establishing the existence of sizeable disparities across regions in allocative efficiency at all degrees of territorial aggregation, and especially at NUTS3 and NUTS2 level. As such, a future direction of research on the topic would be that of analysing the dynamic role of spatial misallocation on productivity growth and divergence path across countries.

References

Aghion, P., Bergeaud, A., Boppart, T. Klenow, P.J., Li, H. (2019). A theory of falling growth and rising rents. *NBER* Working Paper, w26448.

Amable, B., Chatelain, J-B., Ralf, K., (2010). Patents as Collateral. *Journal of Economic Dynamics and Control*, 34 (6): 1092–1104.

Asker, J., Collard-Wexler, A., De Loecker, J., (2019). (Mis)Allocation, Market Power, and Global Oil Extraction. *American Economic Review*, 109 (4): 1568–1615.

Asturias, J., Garcia-Santana, M., Ramos, R. (2014). Misallocation, Internal Trade, and the Role of Transportation Infrastructure. *2014 Meeting Papers from Society for Economic Dynamics*, 1035: 1–35.

Bajgar, M., Berlingieri, G., Calligaris, S., Criscuolo, C., Timmis, J. (2020). Coverage and representativeness of Orbis data. *OECD Science, Technology and Industry Working Papers*, 2020(6): 1-63.

Bartelsman, E. J., Haltiwanger, J., Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review* 103 (1): 305–34.

Baqaee, D. R., Farhi, E. (2020). Productivity and Misallocation in General Equilibrium. *Quarterly Journal of Economics*, 135(1): 105–163.

Bau, N., Matray, A., (2023). Misallocation and Capital Market Integration: Evidence From India. *Econometrica*, 91 (1): 97–106.

Beck, T., Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking and Finance*, 30(11):

2931–2943.

Bils, M., Klenow, P.J., Ruane, C., (2021). Misallocation or Mismeasurement? *Journal of Monetary Economics*, 124: S39–S56.

Blackwood, G. Jacob, Foster, L.S., Grim, C.A., Haltiwanger, J., Wolf, Z., (2021). Macro and Micro Dynamics of Productivity: From Devilish Details to Insights. *American Economic Journal: Macroeconomics*, 13(3): 142–72.

Boeri, T., Ichino, A., Moretti, E., Posch, J., (2021). Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces. *Journal of the European Economic Association*, 1–46.

Bonatti, L., Fracasso, A. (2018). Addressing the core-periphery imbalances in Europe: Resource misallocation and expansionary fiscal policies. In: Antoniolli, L., Bonatti, L., Ruzza, C. (eds) *Highs and Lows of European Integration*. Springer, Cham.: 167–200.

Bradford J. J., McGuckin R. H., Stiroh K. J., (2001). The impact of vintage and survival on productivity: Evidence from cohorts of US manufacturing plants. *The Review of Economics and Statistics*, 83 (2): 323-332

Brooks, Wyatt J., Joseph P. Kaboski, and Yao Amber Li. 2021. Agglomeration, Misallocation, and (the Lack of) Competition. *American Economic Journal: Macroeconomics*, 13 (4): 483-519.

Brugués, F., Brugués J., Giambra S.. (2018). Political Connections and Misallocation of Procurement Contracts: Evidence from Ecuador. *STEG Working Paper*, WP012, 1-58.

Buggle, J. C., Durante, R., (2021). Climate Risk, Cooperation and the Co-Evolution of Culture and Institutions. *The Economic Journal*, 131: 1947–87.

Caggese, A., Cuñat V., Metzger, D., (2019). Firing the Wrong Workers: Financing Constraints and labour Misallocation. *Journal of Financial Economics*, JFE Special Issue on labour and Finance, 133 (3): 589–607.

Caggese, A., Perez-Orive, A., (2017). Capital misallocation and secular stagnation. *European Economic Review*, 142(C): 103987.

Cai, H., Treisman, D.,. (2005). Does Competition for Capital Discipline Governments? Decentralization, Globalization, and Public Policy. *American Economic Review* 95 (3): 817–30.

Calligaris, S., Del Gatto, M., Hassan F., Ottaviano, G., Schivardi, F., (2018). The Productivity Puzzle and Misallocation: An Italian Perspective. *Economic Policy*, 33(96): 635–684.

Charron, N., Dijkstra,L., Lapuente, V., (2014). Regional Governance Matters: Quality of Government within European Union Member States. *Regional Studies* 48 (1): 68–90.

Charron, N., Lapuente, V., Bauhr, M., Annoni, P., (2022). Change and Continuity in Quality of Government: Trends in subnational quality of government in EU member states. *Journal of Regional Research*, 53: 5-23.

Chiavari, A., Goraya, S., (2021). The rise of intangible capital and the macroeconomic implications. JMP, retrived from https://www.upf.edu

Cameron, C. A., and Miller, D. L., (2015). A Practitioners Guide to Cluster-Robust Inference. *Journal of Human Resources* 50(2): 317–372.

Cameron, A.C., Gelbach, J.B., Miller, D.L., (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3): 414–427.

David, J. M., Venkateswaran, V., (2019). The sources of capital misallocation. *American Economic Review*, 109(7): 2531–2567.

David, J. M., Venkateswaran, V., Cusolito, A. P., Didier, T. (2021). Capital Allocation in Developing Countries. *The World Bank Economic Review*, 35(4): 1102–1121.

Dias, D. A., Robalo Marques, C., Richmond, C., (2020). A Tale of Two Sectors: Why is Misallocation Higher in Services than in Manufacturing? *Review of Income and Wealth*, 66(2): 361–393.

Durante, R., Gollin, D., Hassan, T., Rohner, D., Satyanath, S., Waldinger, F. (2019). Mafia and Misallocation in Northern Italy : a new approach of estimating the cost of organised crime. *Working Paper*, 1–52.

Duranton, G., Puga, D., (2004). Micro-foundations of urban agglomeration economies. In *J. V. Henderson J.-F. Thisse (Eds.), Handbook of Urban And Regional Economics, 4:2063–2117. Elsevier B.V.*

Eeckhout, J., Guner, N. (2015). Optimal Spatial Taxation: Are Big Cities Too Small? Discussion Paper 8781. *IZA Discussion Paper Series*, 8781.

Fajgelbaum, P. D., Morales E., Suárez Serrato J.C., Zidar, O., (2019). State Taxes and Spatial Misallocation. *Review of Economic Studies*, 86 (1): 333–76.

Fajgelbaum, P. D., Schaal, E. (2020). Optimal Transport Networks in Spatial Equilibrium. *Econometrica*, 88(4): 1411–1452.

Fakos, A. (2020). Firm Subsidies and Resource Misallocation. *Working Paper*, 1–45.

Fazekas, M., Czibik, Á., (2021). Measuring Regional Quality of Government: The Public Spending Quality Index Based on Government Contracting Data. *Regional Studies*, 55 (8): 1459–72.

Fontagné, L., Santoni, G. (2019). Agglomeration economies and firm-level labour misallocation. *Journal of Economic Geography*, 19(1): 251–272.

Foster, L., Haltiwanger, J. and Syverson C. (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?. *The American Economic Review*, 98(1): 394-425

Foster, L.S., Grim, C.A., Haltiwanger, J., Wolf, Z., (2016). Firm-Level Dispersion in Productivity: Is the Devil in the Details, *The American Economic Review* 106(5): 95–98.

Gamberoni, E, Giordano, C., Lopez-Garcia, P., (2016). Capital and Labour (Mis)Allocation in the Euro Area: Some Stylized Facts and Determinants. *ECB Working Paper Series*, 1981: 1-55.

García-Santana, M., Moral-Benito, E., Pijoan-Mas, J., Ramos, R. (2020). Growing Like Spain: 1995–2007. *International Economic Review*, 61(1): 383–416.

Gopinath, G., Kalemli-Özcan, Şe., Karabarbounis, L., Villegas-Sanchez, C. (2017). Capital allocation and productivity in South Europe. *Quarterly Journal of Economics*, 132(4): 1915–1967.

Gorodnichenko, Y., Revoltella, D., Svejnar, J., Weiss, C.T., (2018). Resource Misallocation in European Firms: The Role of Constraints, Firm Characteristics and Managerial Decisions. *Economics – Working Papers*, 2018/06, European Investment Bank.

Gordeev, S., (2020). Intangible Capital, Tangible Misallocation. Working Paper, accessed 3 March 2023, from https://sgordeev.com/.

Hall, R.E., Jones, C.I., (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others?, *Quarterly Journal of Economics*, 114: 83-116.

Heise, S., Porzio, T. (2021). The Aggregate and Distributional Effects of Spatial Frictions. *SSRN Electronic Journal*.

Heise, S., Porzio, T., (2022). labour Misallocation Across Firms and Regions. *NBER* Working Paper, 30298.

Hottenrott, H., Hall, B. H. Czarnitzki, D. (2016). Patents as quality signals? The implications for financing constraints on R&D. *Economics of innovation and new technology*, 25(3): 197–217.

Hsieh, C.-T., Klenow, P. J., (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, CXXIV(4): 1403–1447.

Hsieh, C.-T., Moretti, E. (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2): 1–39.

Hu, Y., Deng, T., Zhang, J. (2020). Can commuting facilitation relieve spatial misallocation of labour? *Habitat International*. 777:102-136.

Iwasaki, I., Kočenda, E., Shida, Y., (2022). Institutions, Financial Development, and Small Business Survival: Evidence from European Emerging Markets. *Small Business Economics* 58 (3): 1261–83.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., Yesiltas, S. (2022). How to Construct Nationally Representative Firm Level Data from the Orbis Global Database. New Facts on SMEs and Aggregate Implications for Industry Concentration. *Tinbergen Institute Discussion Papers*.

Kochen, F. (2022). Finance Over the Life Cycle of Firms. *Job Market Paper*.

Lasagni, A., Nifo, A., Vecchione, G. (2015). Firm productivity and institutional quality: Evidence from italian industry. *Journal of Regional Science*, 55(5): 774–800.

Lashitew, A. A. 2016. Employment Protection and Misallocation of Resources across

Plants: International Evidence. *CESifo Economic Studies*, 62(3): 453–90.

Lee, N., Luca, D. (2019). The big-city bias in access to finance: evidence from firm perceptions in almost 100 countries. *Journal of Economic Geography*, 19(1): 199–224.

Lenzu, S., Manaresi, F. (2019). Sources and Implications of Resource Misallocation: New Evidence from Firm-Level Marginal Products and User Costs. *SSRN Electronic Journal*.

León-Ledesma, M. (2016). Potential Growth, Misallocation, and Institutional Obstacles: Firm-Level Evidence *ADB Economics*.

León-Ledesma M., and Christopoulos, D., (2016). Misallocation, Access to Finance, and Public Credit: Firm-Level Evidence. *Asian Development Review*, MIT Press, 33(2):119-143.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., Wanner, H. (2004). European seasonal and annual temperature variability, trends and extremes since 1500. *Science*, 303: 1499–1503.

Marconi, D., Upper, C. (2017). Capital Misallocation and Financial Development: A Sector-Level Analysis. *SSRN Electronic Journal*, (671).

Martínez-Vázquez, J., Li, Z. (2020). Fiscal Decentralization and Interregional Capital Misallocation : Evidence from China. *International Center for Public Policy*, Working Paper, 20-14.

Mas, M., Milana, C., Serrano, L. (2008). Spain and Italy: Catching up and falling behind . Two different tales of productivity slowdown. *MPRA* Paper, 15834: 1-39.

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

Mertens, M., Mottironi, B., (2023). Do Larger Firms Have Higher Markups? *IWH-CompNet Discussion Paper.*

Michelacci, C., Schivardi, F. (2013). Does Idiosyncratic Business Risk Matter for Growth? *Journal of the European Economic Association*, 11(2): 343-368.

Midrigan, V., Xu, D.Y., (2014). Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review*, 104 (2): 422–58.

Misch, F., Saborowski, C. (2020). The Drivers and Consequences of Resource Misallocation: Exploiting Variation across Mexican Industries and States. *Economia*, 2: 61–96.

Nakabashi, L., Pereira, A.E., (2023). Factors of Production, Productivity, Institutions, and Development: Evidence from Brazil. *Review of Development Economics*, 27 (2): 1034–55.

Nifo, A., Ruberto, S., Vecchione, G. (2018). Does institutional quality matter for lending relationships? Evidence from Italy. *Journal of Applied Finance Banking*, 8(2): 69–100.

North D.C. (1990). Institutions, institutional change and economic performance. Cambridge University Press.

Pauling, A., Luterbacher, J., Casty, C., Wanner, H., (2005). 500 years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Climate Dynamics*, 26: 387–405.

Pellegrino, B., Zingales, L., (2017). Diagnosing the Italian Disease. *NBER*, Working Paper Series, 23964.

Piemontese, L., (2019). Mafia and Misallocation in Northern Italy: a new approach of estimating the cost of organised crime. *Working Paper.*

Pollard, J. S., (2003). Small firm finance and economic geography. *Journal of Economic Geography* 3(4): 429–452.

Restuccia, D., Rogerson, R., (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics*, 11: 707–720.

Restuccia, D., Rogerson, R., (2017). The Causes and Costs of Misallocation. *The Journal of Economic Perspectives*, 31(3): 151–174.

Rodríguez-Pose, A., Ganau, R., (2022). Institutions and the Productivity Challenge for European Regions. *Journal of Economic Geography*, 22(1): 1–25.

Rodrik D., Subramanian A. and Trebbi F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9: 131-165.

Roodman, D., Nielsen, M.Ø., MacKinnon, J. G., Webb, M.D., (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19:1, 4-60.

Saffie, F. E., (2014). The Impact of Firm Heterogeneity in Aggregated Productivity: Firm Entry, Financial Selection, and Capital Misallocation. *Publicly Accessible Penn Dissertations*, 1431.

Shikimi, M. (2017). Bank Concentration and Misallocation of Credit: Evidence from Firm Entry and Exit. *SSRN Electronic Journal*.

Syverson, C., (2011). What Determines Productivity? *Journal of Economic Literature*, American Economic Association, 49(2):326–365.

Tito, M., Wang, R., (2021). Misallocation in Open Economy. *SSRN Scholarly Paper*.

Vacca, V. (2013). Agglomeration and the Italian North-South divide. *Questioni Di Economia e Finanza*, Banca d'Italia. 2013(4): 1-48.

Xoplaki, E., Luterbacher, J., Paeth, H., Dietrich, D., Steiner, N., Grosjean, M., and

Wanner, H. (2005). European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophysical Research Letters*, 32: L15713.

Young, A. (2022). Consistency without inference: Instrumental variables in practical application. *European Economic Review*, 147: 104112.

Additional tables and figures

Country	Size	N. firms			VA (Mill.€)			Turnover (Mill.€)		
		Orbis	SBS	Orbis/SBS	Orbis	SBS	Orbis/SBS	Orbis	SBS	Orbis/SBS
Austria	micro (<10 empl.)	45	16428	0.27%	186	2782	6.67%	535	7384	7.25%
	small (10-49 empl.)	247	2458	10.05%	1070	5316	20.13%	2553	15704	16.26%
	medium (50-249 empl.)	993	1319	75.28%	10600	16462	64.39%	26200	56911	46.04%
	large (>250empl.)	438	474	92.41%	46000	43682	105.31%	146000	159228	91.69%
	Total	***1723***	***20679***	***8.33%***	***57856***	***68243***	***84.78%***	***175288***	***239227***	***73.27%***
Czech Republic	micro (<10 empl.)	1379	142737	0.97%	201	3873	5.19%	772	12499	6.18%
	small (10-49 empl.)	1338	4178	32.02%	970	4071	23.83%	4282	13937	30.73%
	medium (50-249 empl.)	1461	2895	50.47%	6143	13756	44.66%	25100	51669	48.58%
	large (>250empl.)	654	850	76.94%	22600	37299	60.59%	113000	189192	59.73%
	Total	***4832***	***150660***	***3.21%***	***29914***	***58998***	***50.70%***	***143155***	***267296***	***53.56%***
Germany	micro (<10 empl.)	240	125936	0.19%	501	25406	1.97%	2789	61292	4.55%
	small (10-49 empl.)	1011	33189	3.05%	4661	45029	10.35%	17100	117465	14.56%
	medium (50-249 empl.)	7359	14609	50.37%	65100	136902	47.55%	123000	419014	29.35%
	large (>250empl.)	2611	4426	58.99%	691000	703679	98.20%	2090000	2781836	75.13%
	Total	***11221***	***178160***	***6.30%***	***761262***	***911016***	***83.56%***	***2232889***	***3379608***	***66.07%***
Spain	micro (<10 empl.)	41001	128839	31.82%	6729	12548	53.62%	21600	39785	54.29%
	small (10-49 empl.)	19886	12699	156.60%	20700	20061	103.19%	75800	75574	100.30%
	medium (50-249 empl.)	4289	4530	94.68%	29200	40122	72.78%	122000	162577	75.04%
	large (>250empl.)	892	941	94.79%	58700	86422	67.92%	257000	418844	61.36%
	Total	***66068***	***147009***	***44.94%***	***115329***	***159153***	***72.46%***	***476400***	***696780***	***68.37%***
France	micro (<10 empl.)	6779	162814	4.16%	2552	17732	14.39%	10700	51563	20.75%
	small (10-49 empl.)	8286	11345	73.04%	14500	17820	81.37%	53900	57930	93.04%
	medium (50-249 empl.)	3941	4279	92.10%	31600	42951	73.57%	131000	157163	83.35%
	large (>250empl.)	1279	1383	92.48%	338000	258878	130.56%	1170000	1165063	100.42%
	Total	***20285***	***179821***	***11.28%***	***386652***	***337381***	***114.60%***	***1365600***	***1431719***	***95.38%***
Italy	micro (<10 empl.)	48936	264458	18.50%	10500	33006	31.81%	44000	102290	43.02%
	small (10-49 empl.)	39493	32624	121.06%	49400	48337	102.20%	191000	170774	111.84%
	medium (50-249 empl.)	8076	8141	99.20%	64200	91818	69.92%	260000	362422	71.74%
	large (>250empl.)	1531	1360	112.57%	128000	140290	91.24%	476000	595717	79.90%
	Total	***98036***	***306583***	***31.98%***	***252100***	***313452***	***80.43%***	***971000***	***1231203***	***78.87%***
Poland	micro (<10 empl.)	4269	174140	2.45%	685	6185	11.08%	3712	31181	11.90%
	small (10-49 empl.)	5706	10306	55.37%	3482	8722	39.92%	17600	31058	56.67%
	medium (50-249 empl.)	3896	5755	67.70%	10400	23826	43.65%	48100	89047	54.02%
	large (>250empl.)	1137	1544	73.64%	26400	73096	36.12%	135000	315550	42.78%
	Total	***15008***	***191745***	***7.83%***	***40967***	***111829***	***36.63%***	***204412***	***466838***	***43.79%***
Portugal	micro (<10 empl.)	13586	50081	27.13%	1149	2451	46.89%	3871	8486	45.62%
	small (10-49 empl.)	7785	5302	146.83%	4330	4508	96.06%	14700	15586	94.31%
	medium (50-249 empl.)	2159	2264	95.36%	7274	10508	69.23%	29100	43442	66.99%
	large (>250empl.)	309	309	100.00%	9848	10983	89.67%	43600	51733	84.28%
	Total	***23839***	***57956***	***41.13%***	***22602***	***28450***	***79.44%***	***91271***	***119247***	***76.54%***
Slovenia	micro (<10 empl.)	1174	14153	8.30%	180	1026	17.50%	598	3145	19.02%
	small (10-49 empl.)	407	963	42.26%	469	1090	43.07%	2482	3419	72.61%
	medium (50-249 empl.)	294	481	61.12%	1553	3056	50.80%	5142	9944	51.71%
	large (>250empl.)	91	118	77.12%	4014	4890	82.08%	14000	21598	64.82%
	Total	***1966***	***15715***	***12.51%***	***6215***	***10062***	***61.77%***	***22223***	***38106***	***58.32%***
Total	*micro (<10 empl.)*	*117409*	*1079586*	*10.88%*	*22683*	*105009*	*21.60%*	*88578*	*317625*	*27.89%*
	small (10-49 empl.)	*84159*	*113064*	*74.43%*	*99583*	*154955*	*64.27%*	*379418*	*501447*	*75.66%*
	medium (50-249 empl.)	*32468*	*44273*	*73.34%*	*226070*	*379400*	*59.59%*	*769642*	*1352190*	*56.92%*
	large (>250empl.)	*8942*	*11405*	*78.40%*	*1324562*	*1359220*	*97.45%*	*4444600*	*5698760*	*77.99%*
	Total	***242978***	***1248328***	***19.46%***	***1672898***	***1998584***	***83.70%***	***5682238***	***7870022***	***72.20%***

Table 3.6. Coverage of final sample from Orbis, relative to Eurostat (SBS) data, by country and firm size, year 2019.

Sector (2 digit, Nace Rev.2)	Variable	Austria		Czech Republic		Germany		Spain		France		Italy		Poland		Portugal		Slovenia		*Total*	
		N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.	N.	freq.
Food products	Nfirms	210	0.7%	458	1.5%	950	3.1%	10,103	32.7%	3,816	12.4%	9,842	31.9%	2,087	6.8%	3,230	10.5%	183	0.6%	***30,879***	*100%*
	VA (Mill.€)	3,652	3.6%	1,831	1.8%	17,700	17.5%	17,600	17.4%	31,000	30.6%	21,500	21.2%	5,425	5.4%	2,306	2.3%	368	0.4%	***101,382***	*100%*
Beverages	Nfirms	25	0.4%	94	1.7%	180	3.2%	2,494	44.0%	782	13.8%	1,288	22.7%	166	2.9%	617	10.9%	21	0.4%	***5,667***	*100%*
	VA (Mill.€)	2,390	7.7%	935	3.0%	3,910	12.7%	5,687	18.4%	11,800	38.2%	4,305	13.9%	863	2.8%	894	2.9%	95	0.3%	***30,880***	*100%*
Textiles	Nfirms	36	0.4%	79	0.9%	223	2.7%	2,206	26.4%	470	5.6%	3,561	42.6%	378	4.5%	1,362	16.3%	48	0.6%	***8,363***	*100%*
	VA (Mill.€)	515	3.6%	465	3.2%	3,041	21.0%	1,613	11.1%	1,703	11.7%	5,434	37.5%	535	3.7%	1,106	7.6%	87	0.6%	***14,499***	*100%*
Wearing apparel	Nfirms	16	0.2%	98	0.9%	116	1.1%	1,929	18.5%	278	2.7%	4,827	46.3%	329	3.2%	2,798	26.8%	43	0.4%	***10,434***	*100%*
	VA (Mill.€)	163	0.3%	140	0.2%	2,803	4.5%	1,140	1.8%	5,040	8.0%	6,079	9.7%	958	1.5%	1,190	1.9%	30	0.0%	***62,902***	*100%*
Leather and related products	Nfirms	14	0.2%	14	0.2%	48	0.7%	1,731	23.4%	163	2.2%	3,851	52.2%	94	1.3%	1,458	19.7%	10	0.1%	***7,383***	*100%*
	VA (Mill.€)	179	0.8%	34	0.2%	7,242	33.3%	1,096	5.0%	2,332	10.7%	9,916	45.6%	143	0.7%	758	3.5%	45	0.2%	***21,745***	*100%*
Wood and wood/cork products	Nfirms	105	0.9%	272	2.4%	238	2.1%	3,683	32.5%	930	8.2%	3,481	30.8%	794	7.0%	1,700	15.0%	117	1.0%	***11,320***	*100%*
	VA (Mill.€)	2,022	14.0%	585	4.0%	2,998	20.7%	1,955	13.5%	2,133	14.7%	2,615	18.1%	1,155	8.0%	828	5.7%	173	1.2%	***14,463***	*100%*
Paper and paper products	Nfirms	46	1.0%	87	2.0%	276	6.2%	1,018	22.9%	490	11.0%	1,724	38.8%	474	10.7%	300	6.7%	34	0.8%	***4,449***	*100%*
	VA (Mill.€)	2,428	9.6%	585	2.3%	5,953	23.6%	3,573	14.2%	4,132	16.4%	5,617	22.3%	1,666	6.6%	1,140	4.5%	112	0.4%	***25,205***	*100%*
Chemicals and chemical products	Nfirms	78	1.0%	174	2.2%	692	8.6%	2,418	29.9%	837	10.4%	2,742	34.0%	667	8.3%	405	5.0%	63	0.8%	***8,076***	*100%*
	VA (Mill.€)	2,958	2.1%	1,575	1.1%	72,000	50.4%	9,517	6.7%	39,500	27.6%	13,900	9.7%	2,245	1.6%	907	0.6%	362	0.3%	***142,964***	*100%*
Pharmaceutical products and pharmaceutical preparations	Nfirms	31	2.2%	15	1.1%	246	17.6%	329	23.5%	181	12.9%	411	29.4%	106	7.6%	71	5.1%	9	0.6%	***1,399***	*100%*
	VA (Mill.€)	1,990	1.8%	165	0.1%	45,500	41.1%	6,308	5.7%	29,300	26.5%	14,100	12.8%	373	0.3%	531	0.5%	1,313	1.2%	***110,581***	*100%*
Rubber and plastic products	Nfirms	116	0.9%	396	3.1%	900	7.1%	2,706	21.3%	1,141	9.0%	5,153	40.5%	1,412	11.1%	709	5.6%	196	1.5%	***12,729***	*100%*
	VA (Mill.€)	2,079	2.3%	2,374	2.7%	33,000	37.3%	5,809	6.6%	22,700	25.6%	16,900	19.1%	3,726	4.2%	1,481	1.7%	509	0.6%	***88,577***	*100%*
Other non-metallic mineral products	Nfirms	115	0.9%	211	1.6%	452	3.5%	3,752	28.9%	994	7.7%	4,977	38.3%	846	6.5%	1,563	12.0%	72	0.6%	***12,982***	*100%*
	VA (Mill.€)	3,507	5.5%	1,581	2.5%	15,800	24.8%	6,253	9.8%	21,200	33.3%	9,694	15.2%	2,943	4.6%	2,422	3.8%	275	0.4%	***63,673***	*100%*
Basic metals	Nfirms	75	1.8%	91	2.2%	405	9.6%	1,168	27.7%	356	8.4%	1,617	38.3%	310	7.3%	164	3.9%	33	0.8%	***4,219***	*100%*
	VA (Mill.€)	759	1.3%	1,019	1.7%	26,800	45.1%	4,300	7.2%	6,928	11.7%	10,500	17.7%	1,514	2.5%	456	0.8%	261	0.4%	***59,373***	*100%*
Fabricated metal products, except machinery	Nfirms	269	0.5%	1,459	2.7%	1,687	3.1%	13,966	25.8%	4,209	7.8%	23,908	44.2%	2,970	5.5%	5,122	9.5%	542	1.0%	***54,132***	*100%*
	VA (Mill.€)	4,611	4.6%	4,717	4.7%	28,900	28.9%	10,900	10.9%	13,300	13.3%	29,600	29.5%	4,980	5.0%	2,555	2.6%	607	0.6%	***100,171***	*100%*
Computer, electronic and optical products	Nfirms	103	1.6%	143	2.2%	945	14.6%	1,000	15.4%	686	10.6%	2,904	44.7%	479	7.4%	149	2.3%	84	1.3%	***6,493***	*100%*
	VA (Mill.€)	4,370	4.2%	856	0.8%	51,000	48.8%	1,754	1.7%	34,000	32.5%	10,700	10.2%	1,263	1.2%	400	0.4%	165	0.2%	***104,508***	*100%*
Electrical equipment	Nfirms	72	1.0%	296	4.0%	648	8.8%	1,201	16.3%	570	7.7%	3,614	49.1%	551	7.5%	323	4.4%	83	1.1%	***7,358***	*100%*
	VA (Mill.€)	3,457	5.2%	2,039	3.1%	20,500	30.8%	3,881	5.8%	21,500	32.3%	11,500	17.3%	2,464	3.7%	629	0.9%	504	0.8%	***66,475***	*100%*
Machinery and equipment n.e.c.	Nfirms	252	1.1%	515	2.3%	2,024	9.2%	4,114	18.7%	1,448	6.6%	11,426	52.1%	1,097	5.0%	877	4.0%	197	0.9%	***21,950***	*100%*
	VA (Mill.€)	9,753	4.9%	3,055	1.5%	123,000	61.9%	7,802	3.9%	11,600	5.8%	38,700	19.5%	2,537	1.3%	1,690	0.9%	582	0.3%	***198,720***	*100%*
Motor vehicles, trailers and semi-trailers	Nfirms	64	1.5%	126	2.9%	304	7.1%	1,158	26.9%	583	13.5%	1,259	29.2%	423	9.8%	335	7.8%	56	1.3%	***4,308***	*100%*
	VA (Mill.€)	4,204	1.3%	6,665	2.0%	243,000	73.7%	14,000	4.2%	41,100	12.5%	14,100	4.3%	4,035	1.2%	1,911	0.6%	576	0.2%	***329,591***	*100%*
Other transport equipment	Nfirms	16	0.7%	41	1.7%	144	5.9%	471	19.2%	229	9.3%	1,210	49.2%	214	8.7%	119	4.8%	15	0.6%	***2,459***	*100%*
	VA (Mill.€)	922	1.3%	340	0.5%	12,700	18.3%	5,847	8.4%	35,800	51.5%	12,700	18.3%	892	1.3%	221	0.3%	28	0.0%	***69,451***	*100%*
Furnitures	Nfirms	40	0.4%	110	1.0%	163	1.5%	3,651	33.1%	448	4.1%	4,119	37.4%	766	6.9%	1,634	14.8%	96	0.9%	***11,027***	*100%*
	VA (Mill.€)	411	3.2%	172	1.4%	1,773	14.0%	1,939	15.3%	1,178	9.3%	4,865	38.4%	1,681	13.3%	625	4.9%	31	0.2%	***12,675***	*100%*
Other manufacturing	Nfirms	39	0.6%	152	2.4%	0	0.0%	1,819	28.5%	667	10.5%	2,735	42.9%	0	0.0%	900	14.1%	64	1.0%	***6,376***	*100%*
	VA (Mill.€)	586	5.2%	484	4.3%	0	0.0%	1,099	9.8%	3,582	31.8%	4,931	43.8%	0	0.0%	486	4.3%	90	0.8%	***11,257***	*100%*
Total	Nfirms	***1,722***	***0.7%***	***4,831***	***2.1%***	***10,643***	***4.6%***	***60,929***	***26.3%***	***19,283***	***8.3%***	***94,658***	***40.8%***	***14,166***	***6.1%***	***23,845***	***10.3%***	***1,969***	***0.8%***	***232,003***	
	VA (Mill.€)	***50,956***	***3.1%***	***29,615***	***1.8%***	***717,620***	***44.1%***	***112,071***	***6.9%***	***339,828***	***20.9%***	***247,657***	***15.2%***	***39,398***	***2.4%***	***22,536***	***1.4%***	***6,215***	***0.4%***	***1,629,091***	

Table 3.7. Sectoral distribution of firms and Value Added in the final sample, by country (2019 data).

	Austria		Czech Republic		Spain		Germany		France		Italy		Poland		Portugal		Slovenia		Total	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
log TFPR (firm-level)	0.46	0.36	0.37	0.46	0.48	0.33	0.24	0.42	0.46	0.31	0.52	0.52	0.51	0.61	0.29	0.47	0.45	0.44	0.41	0.47
Relative TFPR (firm-level)	1.60	0.63	1.75	0.90	1.58	0.65	1.27	0.55	1.54	0.51	1.67	1.17	1.83	1.47	1.34	0.66	1.89	1.05	1.52	0.91
VarlnTFPR (NUTS0)	0.42	0.01	0.48	0.03	0.35	0.01	0.35	0.02	0.31	0.02	0.51	0.01	0.61	0.01	0.32	0.02	0.49	0.01	0.42	0.10
Turnover (MEUR, firm-level)	107.85	321.66	18.79	179.27	187.02	3,044.39	4.60	61.50	17.70	492.23	7.02	77.44	13.37	62.46	2.52	23.98	7.45	46.62	14.78	560.81
Value Added (MEUR, firm-level)	26.32	84.06	3.89	34.20	33.82	648.93	1.07	9.75	4.02	89.15	1.70	13.65	2.56	9.16	0.61	5.72	2.26	18.41	3.83	149.30
Cost of labour (MEUR, firm-level)	17.17	53.60	2.25	13.10	21.78	361.05	0.73	6.19	2.85	52.61	1.12	9.38	1.32	4.36	0.39	2.35	1.37	9.75	2.49	82.95
N. of employess (firm-level)	341.41	1,252.64	147.67	480.62	445.37	6,358.61	20.62	137.53	100.50	1,574.69	27.49	180.69	120.25	322.69	21.63	72.07	50.97	293.49	58.02	1,482.08
Age (firm-level)	37.25	35.25	19.32	7.99	42.13	38.31	22.23	12.71	26.19	17.17	22.50	15.41	19.89	16.79	21.52	15.62	20.01	10.85	23.88	17.65
Patenting firms, share	44.1%	49.7%	10.2%	30.2%	42.6%	49.4%	4.7%	21.3%	10.1%	30.1%	10.8%	31.1%	14.2%	34.9%	1.7%	12.9%	8.7%	28.1%	10.2%	30.2%
Private-owned, share	29.0%	45.4%	44.0%	49.6%	37.5%	48.4%	23.7%	42.5%	32.4%	46.8%	42.4%	49.4%	36.3%	48.1%	41.0%	49.2%	64.9%	47.7%	34.7%	47.6%
Equity-owned, share	1.3%	11.2%	2.0%	14.0%	2.4%	15.2%	0.9%	9.3%	3.5%	18.5%	0.9%	9.2%	1.5%	12.3%	0.8%	9.0%	0.9%	9.5%	1.5%	12.1%
Govt.-owned, share	0.4%	6.6%	0.3%	5.2%	0.4%	6.3%	0.1%	2.3%	0.1%	2.8%	0.1%	2.3%	0.6%	7.5%	0.0%	1.6%	0.3%	5.3%	0.1%	3.2%
Foreign-owned firms	27.2%	44.5%	20.9%	40.6%	21.0%	40.7%	11.1%	31.4%	17.9%	38.3%	3.2%	17.5%	19.8%	39.9%	31.4%	46.4%	36.4%	48.1%	12.3%	32.9%
HH Index (NUTS0)	0.18	0.14	0.11	0.13	0.08	0.11	0.02	0.04	0.03	0.06	0.02	0.04	0.06	0.09	0.03	0.06	0.14	0.14	0.03	0.07
PSQ Index (NUTS2)	58.31	3.66	52.43	3.31	59.64	2.79	61.53	4.93	59.47	2.78	50.41	3.40	55.32	3.03	53.82	6.16	53.84	1.88	55.57	6.29
Empl% in HT manuf. (NUTS2)	1.21%	0.55%	1.75%	0.60%	1.76%	0.80%	0.70%	0.47%	1.15%	0.62%	1.11%	0.43%	0.98%	0.54%	0.50%	0.12%	2.13%	0.41%	1.00%	0.58%
Empl% in financial sector	3.16%	0.70%	2.31%	1.22%	3.16%	1.08%	2.36%	0.89%	3.13%	1.14%	2.96%	0.77%	2.54%	1.43%	1.82%	0.88%	2.69%	0.67%	2.69%	1.02%
R&D MIO pps (NUTS2)	885.45	484.71	312.73	182.16	1924.34	2,492.74	825.60	728.29	2180.77	3,062.35	1321.86	1,026.97	409.45	538.21	373.32	173.70	370.26	67.70	1098.56	1,367.45
GDP, p.c. (th.EUR NUTS2)	35.63	5.57	23.78	9.82	33.33	7.69	26.06	5.52	30.43	9.14	30.11	6.47	20.90	8.32	19.45	3.61	24.03	4.78	27.59	7.45
Population per square km. (NUTS2)	577.06	1,382.42	403.35	772.75	425.36	583.88	222.17	232.80	234.51	301.85	274.65	117.91	187.86	132.09	222.12	254.07	106.18	17.67	258.19	319.27

Table 3.8. Summary statistics, means by country

Variable name	Proxy for	Years	Definition	Data Source
Dependent variables				
Misallocation (NUTS0-3)	*Within-region misallocation*	2011-2020	Within country (NUTS0) or region (NUTS1-3) variance of log. TFPR.	Own calculation, on Orbis BvD data
Reallocation gains (% output) (NUTS0-3)	*Reallocation gains*	2011-2020	Potential manufacturing output gains from reallocation in percentage points, measured at country (NUTS0) and regional level (NUTS1-3)	Own calculation, on Orbis BvD data
Between-group component (NUTS1-3)	*Between-group component, % of aggregate*	2011-2020	Share of the between-group component over total TFPR dispersion.	Own calculation, on Orbis BvD data
Firm-level markers				
Micro; Small; Medium; Large	*Size*	2011-2020	Yearly territorial share of firms in each OECD(2011) size category by n. of employees: micro(0-10), small(10-50), medium(50-250), large(>250).	Own calculation, on Orbis BvD data
Age, log	*Age*	2011-2020	Age of the firm, in logs. (yearly territorial means)	Own calculation, on Orbis BvD data
Equity-owned; Foreign-owned; Govt-Owned.; Private-owned	*Ownership*	2011-2020	Yearly territorial share of firms owned by financial corporations or equity; by a foreign individual or group; by the government; individually or family owned.	Own calculation, on Orbis BvD data
Patenting	*Innovativeness*	2011-2020	Yearly territorial share of firms that have at least one patent at the end the observed period.	Own calculation, on Orbis BvD data
Controls				
Administrative region, dummy (NUTS1-3)	*Decentralization*	2011-2020	If the territorial units at the NUTS level under analysis in a specific year identifies an administrative region.	Eurostat
GDP p.c. pps., log (NUTS0-3)	*Economy*	2011-2020	Regional GDP, pps. per inhabitant.	Eurostat
Population density, log (NUTS0-3)	*Agglomeration*	2011-2020	Population density, persons per square kilometre.	Eurostat
HH Index (NUTS0-3)	*Market Concentration at regional level*	2011-2020	Herfindal-Hirschman Index: sum of the squares of the market share (operating revenue) of each firm in the sector ((3digits Nace Rev.2) at NUTS2 level	Own calculation, on Orbis BvD data
Institutional quality				
PSQ Index (NUTS0-2)	*Quality of public spending*	2006-2015	Public Spending Quality Index based on measures of transparency, open and fair competition, administrative efficiency and control of corruption in public procurements.	Fazekas & Czibik (2021)
PSQ: Competition score (NUTS0-2)	*Competition in public spending*	2006-2015	"The beneficial effects of multiple bidders competing against each other are harnessed to achieve low prices, high quality and ontime delivery of procured goods, works and services".	Fazekas & Czibik (2021)
PSQ: Efficiency score (NUTS0-2)	*Efficiency of public spending*	2006-2015	"Timely and balanced public decisions underpinning impartiality"[...] "minimizing the total cost of achieving the predetermined successful completion of the contract."	Fazekas & Czibik (2021)
PSQ: Transparency score (NUTS0-2)	*Transparency of public spending*	2006-2015	"Compliance with the information disclosure requirements in EU Public Procurement Directives". "Information about public procurement should be readily available in a precise, reliable and structured form for the public as a whole or its representatives."	Fazekas & Czibik (2021)
PSQ: Control of corruption score (NUTS0-2)	*Control of corruption in public spending*	2006-2015	"Corruption control […] captures the lack of favouring connected bidders" (e.g. "by bending prior explicit rules and principles of open and fair public procurement")	Fazekas & Czibik (2021)
St. Dev. of precipitations, (NUTS0-3)	*IV for Instiutional Quality: Historical climate risk*	1500-1740	Average for 20-years intervals of interannual standard deviation in precipitations within current NUTS3-0 (2013) regions.	Own calculation on Pauling *et al.* (2005)'s data
St. Dev. of temperatures, (NUTS0-3)	*IV for Instiutional Quality: Historical climate risk*	1500-1740	Average for 20-years intervals of interannual standard deviation in temperatures within current NUTS3-0 (2013) regions.	Own calculation on Luterbacher *et al.* (2004) and Xoplaki *et al.*(2005)'s data

Table 3.9. Main variables: type, description and source.

Number of firms											
	Age	<10 y.o.		10-19 y.o.		20-29 y.o.		>30 y.o.		Total	
Size		*N*	*Freq.*	*N*	*Freq.*	*N*	*Freq.*	*N*	*Freq.*	*N*	*Freq.*
Micro	*N*	36791	38.19%	24732	25.67%	20285	21.06%	14521	15.07%	***96329***	***100%***
	Freq.	65.41%		50.90%		43.82%		26.35%		***46.71%***	
Small	*N*	15754	21.12%	17586	23.58%	17488	23.45%	23762	31.86%	***74590***	***100%***
	Freq.	28.01%		36.19%		37.77%		43.13%		***36.17%***	
Medium	*N*	3150	11.09%	5134	18.08%	6943	24.45%	13175	46.39%	***28402***	***100%***
	Freq.	5.60%		10.57%		15.00%		23.91%		***13.77%***	
Large	*N*	552	7.98%	1140	16.49%	1580	22.86%	3641	52.67%	***6913***	***100%***
	Freq.	0.98%		2.35%		3.41%		6.61%		***3.35%***	
Total	*N*	***56247***	**27.27%**	***48592***	**23.56%**	***46296***	**22.45%**	***55099***	**26.72%**	***206234***	***100%***
	Freq.	**100%**		**100%**		**100%**		**100%**		***100%***	

Table 3.10. Age and size distribution of firms in the final sample by country, year 2019.

EU Area	Country	NUTS1 Name	Type	N.	NUTS2 Name	Type	N.	NUTS3 Name	Type	N.
Western	*Germany*	States	administrative	16	States + government regions	administrative	38	Districts	administrative	401
Western	*France: until 2015*	Macro-regions	statistical	8	Regions & DOM	administrative	22	Departments + DOM	administrative	101
	France: since 2016	Metro areas	administrative	14	Former regions	statistical	27	Departments + DOM	administrative	101
Western	*Austria*	Groups of states	statistical	3	States	administrative	9	Groups of districts	statistical	35
Eastern	*Slovenia*	Whole country	NA	1	Macro-regions	statistical	2	Regions	administrative	12
Eastern	*Czech Republic*	Whole country	NA	1	Macro-regions	statistical	8	Regions	administrative	14
Eastern	*Poland: until 2017*	Macro-regions	statistical	6	Regions	administrative	17	Sub-regions	statistical	73
	Poland: since 2018	Macro-regions	statistical	7	Regions	administrative	16	Sub-regions	statistical	72
Southern	*Italy: 2010-2015*	Groups of regions	statistical	5	Regions	administrative	21	Provinces +	administrative	110
	Italy: since 2016							metropolitan cities	administrative	107
Southern	*Portugal: until 2014*	Continental + Islands	statistical	3	Regions + Islands	administrative	7	Groups of municipalitie	statistical	25
	Portugal: since 2015			3			7	Metro areas + other	administrative	25
Southern	*Spain*	Groups of autonomous communities	statistical	7	Autonomous communities + 2cities	administrative	19	Provinces + Islands + Ceuta & Melilla	administrative	59

Table 3.11. Type, changes and number of regions, by NUTS level. Territorial units are defined as statistical when these do not coincide with administrative regions.

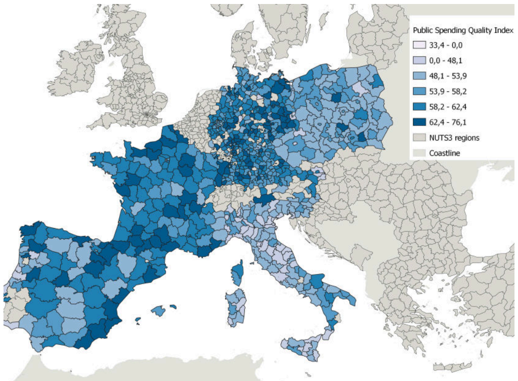


Figure 3.5. Public Spending Quality Index at NUTS3 level, period average (2006-2015). **Source:** own calculations on Fazekas and Czibik (2021)'s data.

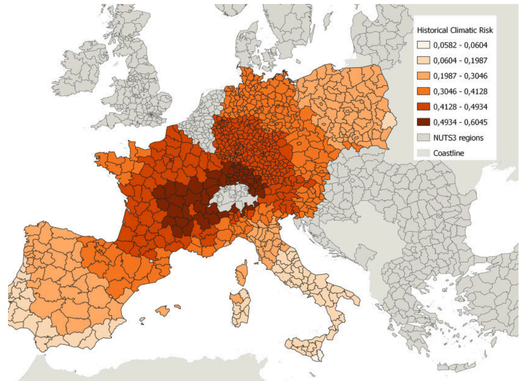


Figure 3.6. Historical climatic risk: mean of the period-averages (1500-1740) of the min-max-normalised interannual standard deviation of precipitations (fig. 3.7) and of temperatures (fig. 3.8) in the growing seasons.

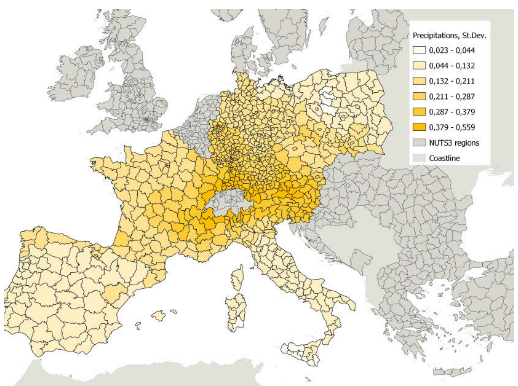


Figure 3.7. Normalised St.Dev. of precipitations, period average (1500-1740). **Source**: own calculations on Pauling *et al.* (2005)'s data.

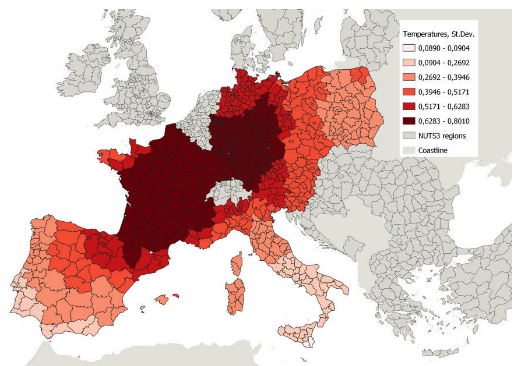
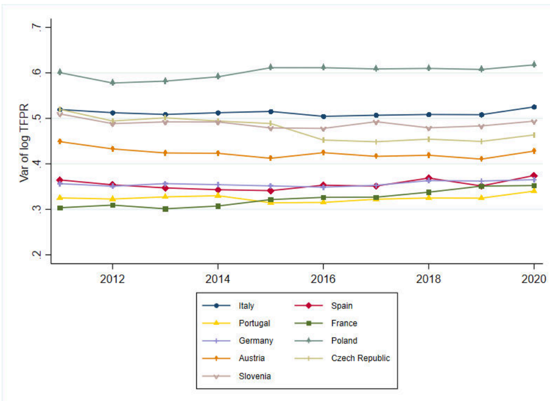
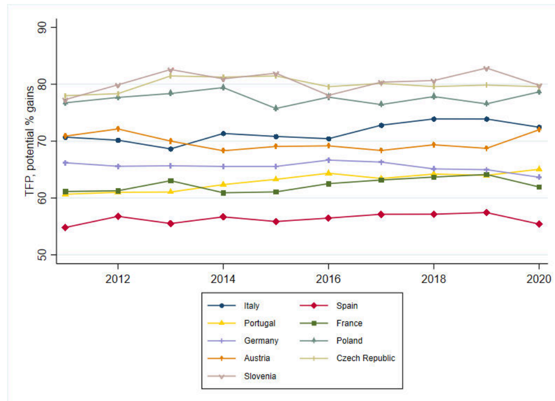


Figure 3.8. Min-Max-Normalised St.Dev. of temperatures, period average (1500-1740). . **Source**: own calculations on Luterbacher *et al.* (2004) and Xoplaki *et al.* (2005)'s data.

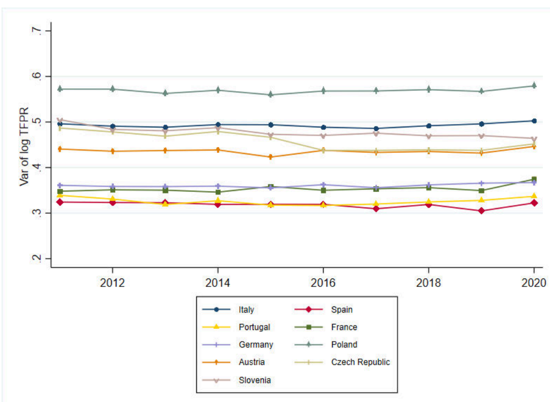
A1: Full Sample, misallocation



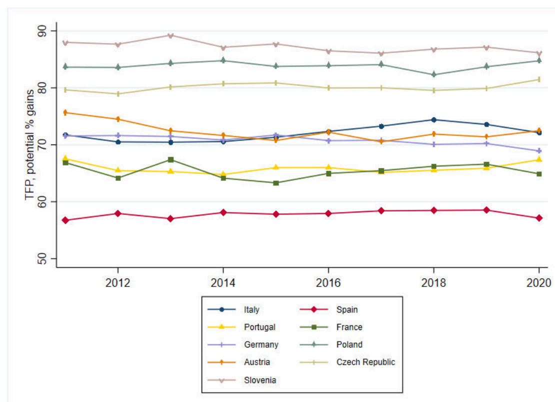
A2: Full Sample, reallocation gains



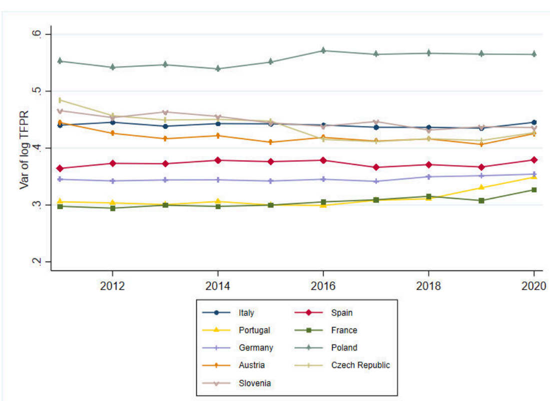
B1: Balanced sample, misallocation



B2: Balanced sample, reallocation gains



C1: Without micro-firms, misallocation



C2: Without micro-firms, reallocation gains

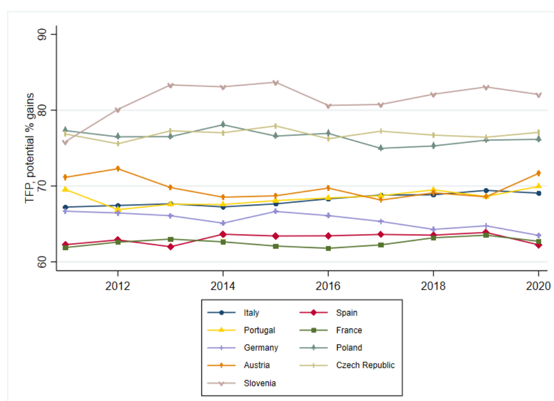


Figure 3.9. Robustness checks: aggregate misallocation and reallocation gains, calculated on different samples

Appendix: Hsieh and Klenow (2009)'s model

where heterogeneous firms, differing in their physical TFP (A_i), face the same marginal cost of inputs and distinct firm-specific input constraints.

Hsieh and Klenow (2009) developed a model à la Melitz (2003), but in a closed economy, where heterogeneous firms, differing in their physical TFP (A_i), face the same marginal cost of inputs and distinct firm-specific input constraints. In this economy, production happens with a three-levels structure, where a final homogeneous good is produced through a Cobb-Douglas technology combining the output of S manufacturing sectors:

$$Y = \prod_{s=1}^{S} Y_s^{\theta_s} \tag{3.6.10}$$

with input shares θ_s based on sectoral value-added shares. In turn, the sectoral level's output Y_s is the CES aggregation of the N_s firm-level varieties, with σ[87] constant elasticity of substitution among them:

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{3.6.11}$$

Finally, each firm i in sector s produces one single variety i_s under a constant returns to scale Cobb-Douglas technology:

$$Y_{si} = A_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s} \tag{3.6.12}$$

Firms are heterogeneous in productivity (A_{si}) and face two types of idiosyncratic distortions, a capital wedge $\tau_{k_{si}}$ affecting the relative marginal revenue product of

87 We set $\sigma = 3$ as in the previous works from Hsieh and Klenow (2009) and Calligaris *et al.* (2018).

one factor with respect to the other, and an output wedge $\tau_{Y_{si}}$ affecting the marginal products of both factors by the same proportion. Distortions appear in the firm's profit function:

$$\Pi_{si} = (1 - \tau_{Y_{si}}) P_{si} Y_{si} - w L_{si} (1 + \tau_{K_{si}}) R K_{si} \tag{3.6.13}$$

where P_{si} is the price, $P_{si}Y_{si}$ the value-added, w is the wage, and $R = (r - \delta)$ is the rental rates of capital, being the difference between the real interest rate (r) and the depreciation rate (δ).

From the profit maximization's FOC of firm's i in sector s, we obtain the following marginal revenue products of capital and labour:

$$MRPK_{si} = P_{si} \frac{\partial (1 - \tau_{si}^{Y}) Y}{\partial K_{si}} = \alpha_s \frac{P_{si} Y_{si}}{K_{si}} = (1 - \tau_{si}^{K})(r - \delta) = (1 - \tau_{si}^{K}) R \tag{3.6.14}$$

$$MRPL_{si} = P_{si} \frac{\partial (1 - \tau_{si}^{Y}) Y}{\partial L_{si}} = (1 - \alpha_s) \frac{P_{si} Y_{si}}{L_{si}} = w = W_L \tag{3.6.15}$$

Assuming each sector to be monopolistically competitive, prices can be obtained according to the mark-up rule:

$$P_{si} = \frac{\sigma}{\sigma - 1} \beta_s \left((1 - \tau_{si}^{K}) R \right)^{\alpha_s} w^{1 - \alpha_s} \frac{(1 + \tau_{si}^{K})^{\alpha_s}}{(1 + \tau_{si}^{Y})^{1 - \alpha_s}} \frac{1}{A_{si}} \tag{3.6.16}$$

with $\beta_s = \alpha_s^{\alpha_s} (1 - \alpha_s)^{1 - \alpha_s}$ being a scaling constant for all firms in the sector. Firms' prices are thus affected by distortions, such that sectoral price-indexes does not constitute valid deflators.

Physical Total factor productivity ($TFP_{si} = A_{si}$) is normally difficult to retrieve in absence of data on firm-specific prices and quantities, and as a common practice it used to be obtained by deflating firm-level revenue based measures (TFPR) with sectoral prices. Foster *et al.* (2008) demonstrated the importance of distinguishing between physical (TFP, or TFPQ) and revenue-based measures of productivity by showing that the former are inversely, and the second ones positively, correlated with prices. The practice of deflating revenues with sectoral average prices instead of firm-specific prices was thus shown to lead to confounding and understating the role of technical efficiency and demand effects. Hsieh and Klenow (2009), following Foster *et al.* (2008)'s findings, exploit the distinction among revenue productivity (TFPR) and physical productivity (TFP, also called TFPQ to distinguish it from the industry-price deflated measure of TFP), and show they can both be obtained as a function of value-added, as follows:

$$TFPR_{si} = P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}} \tag{3.6.17}$$

$$TFPQ_{si} = TFP_{si} = A_{si} = \frac{Y_{si}}{K_{si}^{\alpha} L_{si}^{1-\alpha}} = \kappa_s \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha} L_{si}^{1-\alpha}} \tag{3.6.18}$$

where $\kappa_s = (P_s Y_s)^{-\frac{1}{\sigma-1}}/P_s$ is a sectoral scaling constant that, not depending on wedges nor on firms production, does not affect reallocation gains and can be set equal to 1.

Combining Eq.3.6.16 with Eq.3.6.17, we have:

$$TFPR_{si} = \frac{\sigma}{\sigma-1}\beta_s \left((1-\tau_{si}^K)R\right)^{\alpha_s} w^{1-\alpha_s} \frac{(1+\tau_{si}^K)^{\alpha_s}}{(1+\tau_{si}^Y)} \tag{3.6.19}$$

where the only firm-specific variables are the distortions, and which can be shown to imply the following:

$$\overline{TFPR}_s \propto \overline{MRPL}_s^{\alpha_s} \overline{MRPK}_s^{1-\alpha_s} \propto \frac{(1+\tau_{K_{si}})^{\alpha_s}}{1-\tau_{Y_{si}}} \tag{3.6.20}$$

where $\overline{TFPR_s}$, *i.e.* the sectoral mean of $TFPR$, is calculated as the geometric mean of the sectoral average marginal revenue product of capital (MRPK) and labour (MRPL). Finally, if the distributions of TFP and $TFPR$ are assumed to be jointly lognormally distributed, the negative relationship between sectoral TFP and the variance of TFPR can be expressed explicitly as:

$$lnTFP_s = \frac{1}{\sigma - 1} ln(\sum_i A_{si}^{\sigma-1}) - \frac{\sigma}{2} var(lnTFPR_{si}) \tag{3.6.21}$$

In this setting, sectoral TFP can be calculated as:

$$TFP_s = A_s = \left[\sum_{i=1}^{N_s} \left(A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}} \tag{3.6.22}$$

We refer to the term $TFPR_{si}/\overline{TFPR_s}$, the inverse of that shown in the above equation, as *relative* $TFPR$: according to the Hsieh and Klenow (2009)'s framework, this measure should take a value of 1 in absence of distortions, since firms' $TFPR$ should equate the sectoral mean ($\overline{TFPR_s}$). *Relative* $TFPR$ can be seen as a raw measure of misallocation, as it indicates that the firm is too large (small) when it takes values below (above) unity.

This allows to compute a measure of misallocation through the dispersion in sectoral $TFPR$, and to calculate the counterfactual efficient level of sectoral total factor productivity (TFP_s^*) as:

$$TFP_s^* = A_s^* == \left(\sum_{i=1}^{N_s} A_{si}^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \tag{3.6.23}$$

and the ratio between observed and efficient output (Y) as:

$$\frac{Y}{Y^*} = \prod_{s=1}^{S} \left(\frac{A_s}{A_s^*}\right)^{\theta_s} = \prod_{s=1}^{S} \left[\frac{1}{N_s} \sum_{i_1}^{N_s} \left(\frac{A_{si}}{A_s^*} \frac{\overline{TFPR_s}}{TFPR_{si}}\right)^{\sigma-1}\right]^{\frac{\theta}{\sigma-1}} \tag{3.6.24}$$

Such that gains from reallocation can be obtained as:

$$\%Gain_{t/\text{within}} = \left(\frac{Y_t^*}{Y_t} - 1\right) * 100 \tag{3.6.25}$$